1993 IEEE International Symposium on Information Theory
Hilton Palacio del Rio Hotel, San Antonio, Texas, USA
January 17-22, 1993

Co-Chairmen
R. M. Gray
J. D. Gibson

Program Committee
R. E. Blahut (Chairman)
A. R. Barron
I. F. Blake
P. Chevillat
J. M. Cioffi
D. Coppersmith
D. J. Costello, Jr.
T. Fischer
L. E. Franks
B. Hajek
B. Hughes
J. L. Massey
N. Mehravari
J. M. F. Moura
J. A. O'Sullivan
H. V. Poor
P. H. Siegel
S. Verdu
S. G. Wilson
A. Wyner

International Advisory Committee

I. Ingemarsson (Sweden, Chairman)
I. F. Blake (Canada)
G. D. Cohen (France)
B. G. Dorsch (Germany)
P. G. Farrell (England)
T. Helleseth (Norway)
A. Kuznetsov (USSR)
D. Lazic (Yugoslavia)
M. Longo (Italy)
K. Nakagawa (Japan)
T. Nemetz (Hungary)
E. Paaske (Denmark)
J. P. M. Schalkwijk (Netherlands)
A. Tietäväinen (Finland)
E. C. van der Meulen (Belgium)

Finance
J. G. Dunham (Treasurer)

Local Arrangements
C. N. Georghiades (Chairman)
J. Livingston
V. Vaishampayan

Publications
B. Aazhang

Publicity
G. H. Sasaki

Registration
G. C. Orsak

July 9, 1993

Defense Technical Information Center
Building 5, Cameron Station
Alexandria, VA 22314

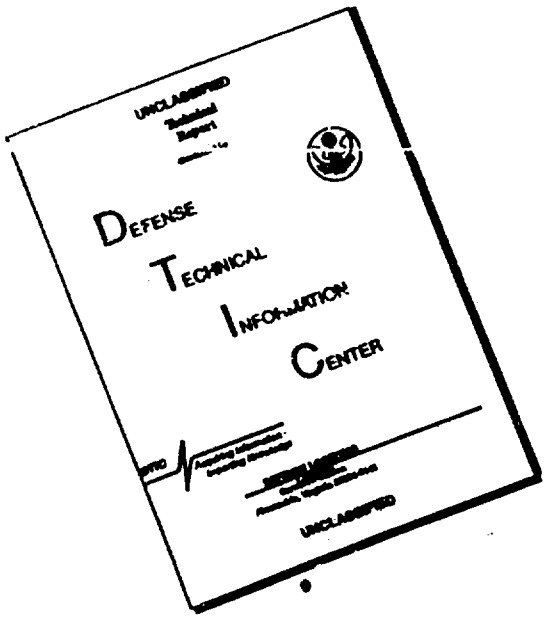
Enclosed is one (1) copy of the Proceedings of the 1993 IEEE International Symposium on Information Theory in fulfillment of the conditions of ONR Contract No. N00014-93-1-0021.

Sincerely,

Jerry Gibson
General Co-Chair

cc: Robert M. Gray
David E. Galicki

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

A268 968

PROCEEDINGS
1993 IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY

Hilton Palacio del Rio
San Antonio, Texas
U.S.A.

January 17-22, 1993

Sponsored by the
Information Theory Society
of the
Institute of Electrical
and Electronics Engineers

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☒
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

DTIC QUALITY INSPECTED 1

1993 IEEE International Symposium on Information Theory

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are premitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 27 Congress Street, Salem, MA 01970. Instructors are premitted to photocopy isolated articles for non-commercial classroom use without fee. For other coping, reprint, or republication permission , write to IEEE Copyright Manager, IEEE Service Center,445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright © 1992 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 92CH3230-0
ISBN Casebound: 0-7803-0878-6
Softbound: 0-7803-0877-8
Microfiche: 0-7803-0879-4
Library of Congress Number: 72-179437

Additional copies of this publication are available from

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
1-800-678-IEEE

Co-Chairmen

Robert M. Gray Jerry D. Gibson

Treasurer:	J.G. Dunham
Publications:	B. Aazhang
Registration:	G.C. Orsak
Publicity:	G.H. Sasaki
Local Arrangements:	C.N. Georghiades (Chairman) J. Livingston V. Vaishampayan

ACKNOWLEDGMENTS

We gratefully acknowledge the following organizations for their financial support for the 1993 IEEE ISIT:

- The National Science Foundation
- The Office of Naval Research
- Motorola Corporation

PROGRAM COMMITTEE

Richard E. Blahut, Chairman

A.R. Barron
I.F. Blake
P. Chevillat
J.M. Cioffi
D. Coppersmith
D.J. Costello, Jr.
T. Fischer
L.E. Franks
B. Hajek
B. Hughes
J.L. Massey
N. Mehravari
J.M.F. Moura
J.A. O'Sullivan
H.V. Poor
P.H. Siegel
S. Verdu
S.G. Wilson
A. Wyner

INTERNATIONAL ADVISORY COMMITTEE

I. Ingemarsson, Chairman (Sweden)
I.F. Blake (Canada)
G.D. Cohen (France)
B.G. Dorsch (Germany)
P.G. Farrell (England)
T. Helleseth (Norway)
D. Lazic (Yugoslavia)
M. Longo (Italy)
K. Nakagawa (Japan)
T. Nemetz (Hungary)
E. Paaske (Denmark)
J.P.M. Schalkwijk (the Netherlands)
A. Tietäväinen (Finland)
E.C. van der Meulen (Belgium)

MONDAY

	Plenary Session: J. Cohen						
	A	B	C	D	E	F	G
AM	Constrained Codes for Recording (Recording I)	Detection Theory	Decoding of Convolutional and Trellis Codes	Soft Decision Decoding	Coding for Special Channels I	Multiuser Detection for CDMA Channels	Universal Source Coding
PM	Coded Modulation I	Shannon Theory I	Coding for the Adder Channel	Detection and Estimation	Block Decoding Techniques	Communication on Fading Channels	Source Coding I

TUESDAY

	Plenary Session: D. Snyder						
AM	Error-correcting Modulation Codes for Recording (Record.II)	Model-Based Imaging (Invited Session)	Convolutional Codes	Structure of Block Codes I	Shannon Theory II	Communications Systems	Vector Quantization I
PM	Coded Modulation II	Estimation Theory	Structure of Block Codes II	Communication Systems	Multiuser and AVC Channel Capacity	Source Coding II	Data Communication Networks

WEDNESDAY

	Plenary Session: R. Olshen						
AM	Cryptography	Trellis Codes for Storage Channels (Recording III)	Algebraic Geometry Codes (Invited Session)	Multiaccess Communication I	Channel Capacity and Error Exponents	Suboptimum Decoding of Convolutional Codes	Image Coding

THURSDAY

	Shannon Lecture: E. Berlekamp						
AM	Coded Modulation III	Coding for Special Channels II	Structure of Block Codes III	Random Processes	Multiaccess Communication II	Signal Processing	Source Coding III
PM	Coded Modulation IV	Neural Networks, Classification, and Regression	Algebraic Techniques and Coding	Combinatorics and Coding	CDMA Signature Waveform Design	Applications of Coding	Vector Quantization II

FRIDAY

	Plenary Session: R. Ahlswede						
AM	Concatenated and Multidimensional Coding	Sequences and Arrays	Coded Modulation V	Detection and Communication	Neural Net Capacity and Complexity	Quantization	

TABLE OF CONTENTS

MONDAY SESSIONS

Network paradoxes and the inefficiency of noncooperative games — *Joel E. Cohen* 1
Block-decodable runlength-limited codes via look-ahead technique — *K. A. Schouhamer Immink* 2
On the optimization of constrained channel codes — *P. A. Franaszek and J. A. Thomas* 3
Construction of polynomial-size encoders with small decoding look-ahead for input-constrained channels — *Jonathan J. Ashley, Brian H. Marcus, and Ron M. Roth* 4
Enumerable multi-track (d, k) block codes — *Edward K. Orcutt and Michael W. Marcellin* 5
A universal algorithm for generating optimal and nearly optimal run-length-limited, charge-constrained binary sequences — *Paul E. Bender and Jack K. Wolf* 6
Design of some new balanced codes — *L. Tallini, R. M. Capocelli, and B. Bose* 7
Irreducible components of canonical diagrams for spectral nulls — *Hiroshi Kamabe* 8
Conservative codes — *S. Al-Bassam and B. Bose* 9
Adaptive OS local detection for data fusion — *M. Longo and M. Lops* 10
Asymptotic refinements in Bayesian distributed detection — *Adrian Papamarcou and Po-Ning Chen* 11
Distributed cell-averaging CFAR detection of dependent signal returns — *Rick S. Blum and Saleem A. Kassam* 12
Integration of complementary detection-localization systems — *T. T. Kadota* 13
On the relationship between suboptimal detectors and measures of discrimination — *Geoffrey C. Orsak and Bernd-Peter Paris* 14
Asymptotic expansions for sample size in signal detection — *Marat V. Burnashev and H. Vincent Poor* 15
Robust detection of weak, known signals using higher-order moments — *Kevin R. Kolodziejski, John W. Betz, and John G. Proakis* 16
Robust continuous-time detection of linear processes — *P. Srinivasa Rao and Don H. Johnson* 17
On the detection of Gaussian cyclostationary random processes — *John W. Betz* 18
Reduced-state sequence detectors are not simpler than the Viterbi algorithm with good convolutional codes — *J. B. Anderson and E. Offer* 19
Mapping the boundaries established by state diagram connectivity — *Oliver Collins* 20
On the maximum difference between path metrics in a Viterbi decoder — *Andries P. Hekstra* 21
List output and soft symbol output Viterbi algorithms: Extensions and connections — *Christiane Nill and Carl-Erik W. Sundberg* 22
New low complexity soft maximum likelihood decoding of partial unit memory codes — *V. V. Zyablov, B. Honary, and G. Markarian* 23
On the evaluation of the error performance of trellis codes — *Christian Schlegel* 24
Soft syndrome decoding of trellis coded modulation codes — *Meir Ariel and Jakov Snyders* 25
Error probability analysis in reduced state TCM — *Carlos Valdez, Hiroyuki Fujiwara, and Ikuo Oka* 26
Efficient maximum-likelihood soft-decision decoding of linear block codes using algorithm A^* — *Yunghsiang S. Han, Carlos R. P. Hartmann, and Chih-Chieh Chen* 27
On the performance evaluation of approximate APP decoding — *S. A. Raghavan* 28
Maximum-likelihood soft decision decoding of BCH codes — *Alexander Vardy and Yair Be'ery* 29
Fast generalized-minimum-distance decoding — *Ulrich K. Sorger* 30
Suboptimal soft decision decoding of linear codes — *Ilya I. Dumer* 31
An efficient soft decision decoding algorithm for array codes — *Xiao-Hong Peng and P. G. Farrell* 32
A new efficient error-erasure location scheme in GMD decoding — *Ralf Kötter* 33
The generalized syndrome polynomial and its application to the efficient decoding of Reed–Solomon codes based on GMD criterion — *Kiyomichi Araki, Masayuki Takada, and Masakatu Morii* 34
A family of BCH codes for the Lee metric — *Ron M. Roth and Paul H. Siegel* 35
On minimum Lee distances of generalized Reed–Muller codes — *Tomoharu Shibuya, Hajime Jinushi, and Kohichi Sakaniwa* .. 36
A class of error magnitude subset correcting codes over GF(q) — *A. Di Porto, F. Guida, E. Montolivo, and G. M. Poscetti* .. 37
A class of single error correcting codes for channels with localized errors — *Per Larsson* 38
On perfectness of binary block codes for correcting asymmetric errors — *G. Fang and Iiro S. Honkala* 39
Single byte unidirectional error locating codes — *Eiji Fujiwara and Shuxin Jiang* 40
Efficient maximum likelihood decoding algorithms for linear codes over Z-channel — *Tomohiko Uyematsu* 41
Reduced state sequence detection for asynchronous Gaussian multiple-access channels — *Mahesh K. Varanasi* 42
Error probabilities for fiber-optic code division multiple access systems — *Narayan B. Mandayam and Behnaam Aazhang* .. 43
Performance analysis of optimum demodulation in optical CDMA — *Laurie B. Nelson and H. Vincent Poor* 44
A linear adaptive fractionally spaced single user receiver for asynchronous CDMA systems — *Predrag B. Rapajic and Branka S. Vucetic* 45

Fading resistant multiuser detection for CDMA communications — *Subramanian Vasudevan and Mahesh K. Varanasi* 46
Equalization techniques for direct sequence code-division multiple access systems in multipath channels — *Sarah Kate Wilson and John M. Cioffi* 47
A comparison of differentially coherent and coherent multiuser detection with imperfect phase estimates in a Rayleigh fading channel — *Zoran Zvonar and David Brady* 48
MMSE detection of CDMA signals: Analysis for random signature sequences — *Upamanyu Madhow and Michael L. Honig* .. 49
Asymptotic multiuser efficiency for 2-stage detectors in AWGN channels — *David Brady* 50
Universal coding of non-discrete sources based on distribution estimation consistent in expected information divergence — *Andrew R. Barron, László Györfi, and Edward C. van der Meulen* 51
Sequential model estimation for universal coding and the predictive stochastic complexity of finite-state sources — *Marcelo J. Weinberger, Meir Feder, and Jorma Rissanen* 52
Rate and distortion redundancies for universal source coding with respect to a fidelity criterion — *Philip A. Chou and Michelle Effros* 53
Information bounds for the risk of Bayesian predictions and the redundancy of universal codes — *Andrew Barron, Bertrand Clarke, and David Haussler* 54
There is no universal source code for infinite alphabet — *László Györfi, István Páli, and Edward C. van der Meulen* 55
Noiseless universal encoding of non-unifilar sources — *Yuri M. Shtarkov* 56
Fast coding of sources with unknown statistics — *B. Y. Ryabko* 57
Minimax redundancy for sources with an unknown model — *Joe Suzuki* 58
Context tree weighting: A sequential universal source coding procedure for FSMX sources — *F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens* 59
New spherical 4-designs — *R. H. Hardin and N. J. A. Sloane* 60
Reduced complexity bounded-distance decoding of the Leech lattice — *Ofer Amrani, Yair Be'ery, and Alexander Vardy* 61
An upper bound on the probability of decoding error for M-ary PSK block coded modulation structures — *Hanan Herzberg and Gregory Poltyrev* 62
A bounded-distance decoding algorithm for lattices obtained from a generalized code formula — *Mauro A. O. da Costa e Silva and Reginaldo Palazzo, Jr.* 63
A new block coded modulation scheme and its soft decision decoding — *Kazuhiko Yamaguchi and Hideki Imai* 64
Correction and interpretation of de Buda's theorem — *Tamás Linder, Christian Schlegel, and Kenneth Zeger* 65
Decoding lattice partitions with application to decoding coset codes — *F.-W. Sun and Henk C. A. van Tilborg* 66
Code optimisation for finite error rate — *A. G. Burr and T. J. Lunn* 67
Evaluation of the block error probability of block modulation codes by the maximum-likelihood decoding for an AWGN channel — *Tadao Kasami, Toru Fujiwara, Toyoo Takata, and Shu Lin* 68
Common information of two correlated random variables — *Hirosuke Yamamoto* 69
Entropy as a function of alphabet size — *Christoph G. Günther and Walter R. Schneider* 70
Generalizing FANO's inequality — *Te Sun Han and Sergio Verdú* 71
Relations between entropy and error probability — *Meir Feder and Neri Merhav* 72
Generalized cutoff rates and Rényi's information measures — *I. Csiszár* 73
A generalizaiton of the entropy power inequality with applications to linear transformation of a white-noise — *Ram Zamir and Meir Feder* 74
Rate-distortion computation and statistical physics — *Kenneth Rose* 75
Zipf's law and information complexity in an evolutionary system — *L. B. Levitin and B. Schapiro* 76
On noiseless diagnosis — *Raymond W. Yeung* 77
Upper bound for uniquely decodable codes in a binary input N-user adder channel — *Shraga Bross and Ian F. Blake* 78
Coding for the synchronized multiple-access binary adder channel with idle sources — *Y. W. Wu and S. C. Chang* 79
On cyclic codes for the T-user Q-ary adder channel — *Valdemar C. da Rocha, Jr.* 80
Coding for the Gaussian multiple-access channel: An algebraic approach — *Bixio Rimoldi* 81
Optimal multiuser codes for the real adder channel — *A. Brinton Cooper, III and Brian Hughes* 82
Two-decodable coding of the two-user binary adder channel — *Jian-Jun Shi and Yoichiro Watanabe* 83
Linear codes for an AWGN multiple access channel with partial access — *Gregory Poltyrev and Jakov Snyders* 84
Coding for the F-adder channel: Two applications of Reed Solomon codes — *Rüdiger Urbanke and Bixio Rimoldi* 85
Joint signal detection (D) and estimation (E) under prior uncertainty: New results — *David Middleton* 86
Optimum incoherent detection of fading signals in non-Gaussian noise — *E. Conte, M. Di Bisceglie, and M. Lops* 87
Detection of time-frequency concentrated transient signals — *Thomas P. Krauss, Thomas W. Parks, and Ram G. Shenoy* 88
Quickest detection of an abrupt change in a random sequence with finite change-time — *Yong Liu and Steven D. Blostein* ... 89

Decentralized encoding for linear estimation of a remote source — *M. Di Bisceglie and M. Longo* 90
Model based motion field estimation — *Christoph Stiller and Frank Müller* 91
Multi-grid methods for mean field theory in EM procedures for Markov random fields — *Jun Zhang and Binglai Chen* 92
Maximum likelihood parameter estimation of the harmonic, evanescent, and purely indeterministic components of homogeneous random fields — *Joseph M. Francos, Anand Narasimhan, and John W. Woods* 93
Quantized receiver R_0 bounds and the asymptotic relative efficiency of quantized detectors — *Marcos O. Cimadevilla and Jerry D. Gibson* 94
A new procedure for decoding cyclic and BCH codes up to actual minimum distance — *G. L. Feng and K. K. Tzeng* 95
A new remainder based decoding algorithm for Reed–Solomon codes — *Tomik Yaghoobian and Ian F. Blake* 96
Inverterless Cauchy cells for a systolic Reed–Solomon encoder — *M. A. Hasan and V. K. Bhargava* 97
A new look at the key equation — *Patrick Fitzpatrick* 98
On the minimum code length of S-step (T, U) permutation decodable cyclic codes — *Anader Benyamin-Seeyar, Tho Le-Ngoc, and Ming Jia* 99
Efficient coding/decoding strategies for channels with memory — *Cuong Hon Lai and Samir Kallel* 100
Comparison of erasure-and-error decoding schemes — *Takeshi Hashimoto* 101
Fault-tolerant distributed decoding of cyclic block codes — *Ahsun H. Murad and Thomas E. Fuja* 102
On the fast decoding of binary BCH codes — *W. T. Penzhorn* 103
Diversity systems for Rayleigh fading channels: An application of multiple description source codes — *Shih-Ming Yang and Vinay Vaishampayan* 104
Error performance over the uninterleaved correlated Rician channel — *Gideon Kaplan and Shlomo Shamai* 105
Error probabilty bounds for M-ary DPSK signaling over doubly selective fading diversity channels — *Daniel L. Noneaker and Michael B. Pursley* 106
Limiting cutoff rate for phase-only modulation on a slow-fading Rician channel — *J. W. Modestino* 107
Bidirectional decoding of convolutional codes over Rayleigh fading channels — *Jean Belzile, David Haccoun, and Serge Forest* 108
A Bayesian method for dependent erasures in frequency-hop communication systems with Rayleigh fading — *Carl W. Baum and Michael B. Pursley* 109
Performance analysis of frequency-hopped digital FM diversity systems — *Leonard E. Miller and Jhong S. Lee* 110
Algorithms for parallel decoding — *Wayne E. Stark and Amer A. Hassan* 111
Exact analysis of the Lempel–Ziv algorithm for i.i.d. source — *Tsutomu Kawabata* 112
On asymptotic optimality of a sliding window variation of Lempel–Ziv codes — *Hiroyoshi Morita and Kingo Kobayashi* 113
Adaptive multi-dictionary model for data compression — *Chia-Lun Yu and Ja-Ling Wu* 114
Universal redundancy rates don't exist — *Paul C. Shields* 115
A novel source coding technique with high convergence speed based on the LZW algorithm — *Junichi Kubo, Takaya Yamazato, Iwao Sasase, and Shinsaku Mori* 116
Finite storage discriminators for ergodic processes — *A. D. Wyner and J. Ziv* 117
Block arithmetic coding for Markov sources — *Charles G. Boncelet, Jr.* 118
A positional representation for noiseless compression — *George H. Freeman* 119

TUESDAY SESSIONS

Likelihood methods in imaging — *Donald L. Snyder* 120
On the principal state method for runlength limited sequences — *Tjalling Tjalkens* 121
Joint runlength/error-control codes based on set-concatenatable collections — *Jian Gu and Tom Fuja* 122
Systematic runlength-limited codes for single error detection in the magnetic recording channel — *Patrick Perry* 123
Reduced complexity encoding and decoding algorithms for a class of runlength limited error control codes — *A. Popplewell and J. J. O'Reilly* 124
A scheme for combined modulation and error correction — *Khaled A. S. Abdel-Ghaffar, Mario Blaum, and Jos H. Weber* 125
Resynchronizing (d,k)-constrained sequences in the presence of insertions and deletions — *Mario Blaum, Jehoshua Bruck, C. Michael Melas, and Henk C. A. van Tilborg* 126
Construction of insertion/deletion correcting RLL codes — *Patrick A. H. Bours* 127
The application of q-ary codes for the correction of single peak-shifts, deletions and insertions of zeros — *A. V. Kuznetsov and A. J. Han Vinck* 128
A construction of codes with special properties — *Alexander Barg* 129
The role of information theory in emission tomography — *Larry Shepp* 130
Recursive CR-type bounds and the EM algorithm: Applications to ECT image reconstruction — *A. O. Hero and J. A. Fessler* 131
Simultaneous recovery of the object and aberrations from a sequence of images degraded by atmospheric turbulence — *Timothy J. Schulz* 132

Searching for circumstellar disks with space telescope observations — *Donald German and Joseph Horowitz* 133
A model-based approach to magnetic resonance image estimation — *Timothy J. Schaewe and Michael I. Miller* 134
Model-based multiresolution restoration of speckle images: Application to radar imaging — *P. Moulin* 135
A Markov random field product model for complex-valued radar imagery — *John D. Gorman and Brian J. Thelen* 136
The normalized second moment of the binary lattice determined by a convolutional code — *A. R. Calderbank and P. C. Fishburn* 137
New constructions of $k/(k+1)$ rate-variable punctured convolutional codes — *Pisit CharnKeitKong, Kazuhiko Yamaguchi, and Hideki Imai* 138
A new bound on the row distance of rate $1/n$ convolutional codes — *Y. Levy and D. J. Costello, Jr.* 139
Block code based analysis of convolutional codes — *Øyvind Ytrehus* 140
Covering properties of convolutional codes and associated lattices — *A. R. Calderbank, P. C. Fishburn, and A. Rabinovich* . 141
The extended invariant factor algorithm with application to the Forney analysis of convolutional codes — *Robert J. McEliece and Ivan Onyszchuk* 142
The performance of convolutional codes on the block erasure channel with various finite interleavers — *Amos Lapidoth* ... 143
Using a modified transfer function to calculate unequal error protection capabilities of convolutional codes — *D. G. Mills and D. J. Costello, Jr.* 144
The MacWilliams–Sloane conjecture on the tightness of the Carlitz–Uchiyama bound and the weights of duals of BCH codes — *Oscar Moreno and Carlos J. Moreno* 145
Coset weight enumerators of three Conway–Pless extremal self-dual binary codes of length 32 — *Paul Camion, Bernard Courteau, and André Montpetit* 146
Weight hierarchies of binary linear codes of dimension 4 — *Torleiv Kløve* 147
MacWilliams identities and coordinate partitions — *Juriaan Simonis* 148
On the weight distribution of certain primitive binary cyclic codes — *Jacques Wolfmann* 149
A threshold property of linear codes — *Gilles Zémor and Gérard D. Cohen* 150
The automorphism group of double-error-correcting BCH codes — *T. Berger* 151
A bound on the zero-error list coding capacity — *Erdal Arikan* 152
Approximation theory of output statistics — *Te Sun Han and Sergio Verdú* 153
The Sperner capacity of linear and nonlinear codes for the cyclic triangle — *A. R. Calderbank, R. L. Graham, L. A. Shepp, P. Frankl, and W.-C. W. Li* 154
Secrecy enhancement via public discussion — *Alon Orlitsky and Avi Wigderson* 155
Towards combining Shannon's theory on secrecy systems and the theory of authentication in the case of multiple channel use — *Ben Smeets* 156
Positioning and communication systems — *C. R. Drane* 157
Communicating over a channel constrained to a fixed code — *Aaron B. Kiely and John T. Coffey* 158
About coding without restrictions for the AWGN channel — *Grego*ry *Poltyrev* 159
Noncoherently demodulated convolutional codes — *Y. Kofman, E. Zehavi, and S. Shamai* 160
Selection and square-law combining for NCFSK with correlated branch diversity — *P. J. McLane and C. S. Chang* 161
Undetected error probability of linear block codes on channels with memory — *Francis Swarts, A. J. Han Vinck, and Hendrik C. Ferreira* 162
Simplified reception of convolutionally encoded CPM signals — *Ryszard Bobrowski and Witold Holubowicz* 163
A Markov analysis of digital PLL based MPSK demodulators — *Michael P. Fitz* 164
On the applicability of the Fokker–Planck method in telecommunications — *L. Popken* 165
The synchronization game—PN code acquisition in presence of a white noise, average power constrained, random, symmetric two-state jammer — *Jorge M. N. Pereira* 166
Performance analysis of MPPM in noisy photon counting channel — *Tomoaki Ohtsuki, Iwao Sasase, and Shinsaku Mori* ... 167
Power moments of exponential functionals of Brownian motion — *Yeheskel E. Dallal and Shlomo Shamai* 168
A new structured quantizer for sources with memory — *Rajiv Laroia, Cheng-Chieh Lee, and Nariman Farvardin* 169
Optimal vector quantized nonlinear estimation — *A. Gersho* 170
Robust vector quantization by linear mappings of block-codes — *Roar Hagen and Per Hedelin* 171
An improved tree-structured vector quantizer — *David Miller and Kenneth Rose* 172
Index assignment for progressive transmission of full search vector quantization — *Eve A. Riskin, Les E. Atlas, Ren-Yuh Wang, and Richard Ladner* 173
Generalised theta functions, for lattice vector quantization — *Patrick Solé* 174
Vector quantization codebooks from the Nordstrom–Robinson code and Berlekamp's negacyclic codes — *Peter F. Swaszek* . 175
Vector quantizers trained on small training sets — *David Cohn, Eve A. Riskin, and Richard Ladner* 176
The dynamics of group codes: Syndromes, normal codes, and canonical observers — *G. David Forney, Jr. and Mitchell D. Trott* 177

Realizing trellis codes as isometry codes — *Mitchel D. Trott* 178
On geometrically uniform signal sets and signal sets matched to groups — *Zhe-xian Wan* 179
Euclidean-space coding theorems for linear codes and mod-p lattices — *Hans-Andrea Loeliger* 180
Analysis of block codes designed over the real-field — *Peter Massey and Peter Mathys* 181
On minimality conditions for linear systems and convolutional codes — *Hans-Andrea Loeliger and Thomas Mittelholzer* .. 182
Multilevel codes for unequal error protection — *A. R. Calderbank and N. Seshadri* 183
QPSK modulation codes for unequal error protection — *Robert H. Morelos-Zaragoza and S. Lin* 184
Jump-diffusion processes for unknown model order estimation problems — *Michael I. Miller, Yali Amit, and Ulf Grenander* . 185
The optimal error exponent for Markov order estimation — *Lorenzo Finesso, Chuang-Chun Liu, and Prakash Narayan* 186
On the convergence of the EM algorithm — *A. O. Hero* 187
Necessary and sufficient conditions of channel identifiability based on second-order cyclostationary statistics — *Lang Tong, Guanghan Xu, and Thomas Kailath* 188
Entropy and the consistent estimation of joint distributions — *Katalin Marton and Paul C. Shields* 189
A new bound on the estimation of the probability density function using spectral analysis — *Marcelo S. Alencar* 190
A Cramer–Rao type lower bound for estimators satisfying a bias constraint — *Alfred Hero* 191
Non-linear, non-binary cyclic group codes — *G. Solomon* 192
Classification of cosets of the Reed–Muller code $R(m-3, m)$ — *Xiang-dong Hou* 193
Constructing Reed–Muller codes from Reed–Solomon codes over $GF(q)$ — *Frank R. Kschischang* 195
On the apparent duality of the Kerdock and Preparata codes — *Roger Hammons and P. Vijay Kumar* 196
Normal and abnormal codes — *Tuvi Etzion, Gadi Greenberg, and Iiro S. Honkala* 197
TCH: A new family of cyclic codes length $2m$ — *F. A. B. Cercas, M. Tomlinson, and A. A. Albuquerque* 198
Digital signature schemes based on error-correcting codes — *Mohsen Alabbadi and Stephen B. Wicker* 199
The trellis complexity of equivalent binary [17,9] quadratic residue codes is five — *Yan-Yih Wang and Chung-Chin Lu* 200
Channel equalization for block transmission systems — *Ghassan Kawas Kaleh* 201
Upper bounding the performance of ISI channels — *Sreenivasa A. Raghavan, Jack K. Wolf, and Laurence B. Milstein* 202
Performance of M-algorithm receivers with imperfect channel estimates — *F. Gozzo and J. B. Anderson* 203
New results in signal design for the AWGN channel — *M. Steiner* 204
Practical use of importance sampling in digital communication system simulations — *Kung Yao and Dongrin Kim* 205
A new class of optimum importance sampling strategies derived from statistical distance measures — *Geoffrey C. Orsak and Behnaam Aazhang* 206
Importance sampling using geometry — *A. Dabak and D. H. Johnson* 207
Interference channels with correlated sources — *Masoud Salehi and Erozan Kurtas* 208
Multiuser water-filling — *Roger S. Cheng and Sergio Verdú* 209
The smallest list for the arbitrarily varying channel — *Brian Hughes* 210
Coding strategies for the permuting jammer channel — *Wah Keung Chan* 211
Source divisibility and related problems — *V. N. Koshelev* 212
Data compression with side information and graph entropy — *Alon Orlitsky* 213
Multiple-user distributed information storage — *James R. Roche* 214
Huffman algebras for independent random variables — *Cheng-Shang Chang and Joy A. Thomas* 215
Minimum average cost testing for partially ordered components — *Marc J. Lipman and Julia Abrahams* 216
Extended synchronizing codewords for binary prefix codes — *Wai-Man Lam and Sanjeev Kulkarni* 217
Arithmetic code-like variable-to-variable length data compression code with a fidelity criterion for binary IID sources — *Hisashi Suzuki and Suguru Arimoto* 218
Failure detection for communication networks using finite-state models and Viterbi decoding — *Ender Ayanoglu* 219
Asymptotic non-stationary behavior of statistical multiplexing with multiple types of traffic — *Qiang Ren and Hisashi Kobayashi* 220
The throughput region of networks with time-varying topology — *Leandros Tassiulas* 221
Loss probability approximation for general stationary input traffic — *Kenji Nakagawa* 222
Performance analysis for two Manhattan street network routing algorithms — *Zheng Chen and Toby Berger* 223
Scheduling transmissions in a multicast packet switch when call splitting is allowed — *Charutosh Dixit and Galen Sasaki* .. 224
Minimal standard-path switching networks — *Chris J. Smyth and Liam Halpenny* 225

WEDNESDAY SESSIONS

Binary trees for classification, regression, and clustering, with applications to lossy data compression — *Richard A. Olshen* . 226
Private-key burst correcting code encryption — *F. M. R. Alencar, A. M. P. Léo, and R. M. Campello de Souza* 227
Universal hashing and unconditional authentication codes — *Tran van Trung* 228
Threshold schemes with disenrollment — *Bob Blakley, G. R. Blakley, Agnes Hui Chan, and James L. Massey* 229
An adaptive homofonic algorithm — *Christian Gehrman* 230
Lower bounds on the probability of deception in authentication with arbitration — *Thomas Johansson* 231
On the specification of permutations for block ciphers — *Peter Mathys* 232
A source of cryptographically strong permutations for use in block ciphers — *Lothrop Mittenthal* 233
A new bound for substitution attack — *L. Tombak and R. Safavi-Naini* 234
On RSA signatures — *Thijs Veugen* 235
An attack on Xinmei's digital signature scheme — *Yuan-Xing Li* 236
Asymptotic bounds on the rate of runlength-limited codes — *Shih-Hsuan Yang and Kim A. Winick* 237
Runlength limited trellis codes for partial response recording channels — *Mignon Belongie and Chris Heegard* 238
On trellis codes for peak shift magnetic recording channel — *Ephriam Zehavi and Aaron Biniashvili* 239
A class of DC-free subcodes of convolutional codes — *M. Nasiri-Kenari and C. K. Rushforth* 240
Concatenated coding for binary partial-response channels — *Giovanni Cherubini and Sedat Ölçer* 241
New zero-run length limited codes for partial-response channels — *Kjell Jørgen Hole and Øyvind Ytrehus* 242
Reed–Muller coding for partial response channels — *Sedat Ölçer and Gottfried Ungerboeck* 243
A class of byte error control codes for memory systems —S*b*EC-(S*b*+S)ED codes — *Mitsuru Hamada and Eiji Fujiwara* .. 244
Correcting single-peak shifts with perfect (d, k)-codes — *V. I. Levenshtein and A. J. Han Vinck* 245
Codes on curves and their geometry — *J. W. P. Hirschfeld* 246
Algorithms analogous to algebraic geometric codes — *Gilles Lachaud* 247
Algebraic geometry tools in coding theory — *S. G. Vladut* 248
Decoding of algebraic–geometric codes — *Michael A. Tsfasman* 249
Decoding algebraic–geometric codes up to $(D - 1)/2$ errors — *Dirk Ehrhard* 250
Channel coding strategies for cellular radio — *Gregory J. Pottie and A. Robert Calderbank* 251
A comparison of CDMA and frequency hopping in a cellular environment — *Michael I. Mandell and Robert J. McEliece* .. 252
Capacity of coherent frequency-hop spread-spectrum communications — *Giovanni Cherubini and Wayne Stark* 253
Erlang capacity of a power CDMA system — *Audrey M. Viterbi and Andrew J. Viterbi* 254
Coding decreases delay of messages in networks — *Grigorii A. Kabatianskii and Eugeni A. Krouk* 255
The performance of frequency comb multiple access (FCMA) in interference limited and AWGN environments — *T. J. Stevenson and K. W. Yates* 256
Analysis of a hybrid random-access system with multi-user coding (throughput) — *Rumaih M. Al-Rumaih and Peter Mathys* . 257
Slow frequency hopping patterns derived from polynomial residue class rings — *P. Udaya and M. U. Siddiqi* 258
Bounds on the capacity of an AWGN channel with intertransition constrained bipolar inputs — *Shlomo Shamai and Naftali Chayat* 259
On capacity of frequency non-selective slowly time-varying fading channel — *Roger S. Cheng* 260
The channel capacity in the presence of impulse noise — *Kenneth J. Kerpez* 261
Worst-case power-constrained noise for binary-input channels — *Shlomo Shamai and Sergio Verdú* 262
Error exponents for the ideal Poisson channel with noiseless feedback — *Amos Lapidoth* 263
A direct geometrical method for bounding the error exponent for specific families of channel codes—II: The confining region lower bound for block codes — *Dejan Lazic and Vojin Šenk* 264
Universal decoding for memoryless Gaussian channels with a deterministic interference — *Neri Merhav* 265
On information rates for mismatched decoders — *Neri Merhav, Gideon Kaplan, Amos Lapidoth, and Shlomo Shamai* 266
A Markovian evaluation of the frame error probability for the *M* algorithm — *Jean Belzile and David Haccoun* 267
Systematic feed-forward convolutional encoders are as good as other encoders with an *M*-algorithm decoder — *Harro Osthoff, Rolf Johannesson, and John Anderson* 268
Analysis of list decoding for convolutional codes — *Kamil Zigangirov and Harro Osthoff* 269
Table-driven decoding of binary one-half rate nonsystematic convolutional codes — *Ajay Dholakia, Mladen A. Vouk, and Donald L. Bitzer* 270
Sequential decoding on memoryless soft decision channels under the Pe-criterion — *Ivonete Markman and John B. Anderson* . 271
Optimal trellis decoding at given complexity — *Tor Aulin* 272
Bidirectional sequential decoding algorithms — *Kaiping Li and Samir Kallel* 273
On the computation problem of the stack and Fano decoders for specific time-invariant convolutional codes — *K. Muhammad and K. Ben Letaief* 274

Sequential decoding of linear block codes — *D. J. Tempel and E. Shwedyk* 275
A tree-structure polytopal vector quantizer for real-time image coding — *Shih-Chi Huang and Yih-Fang Huang* 276
Introduction to template coding: An alternative to subpicture coding in black–white image compression — *John C. Kieffer and Greg Nelson* 277
Subpixel accuracy for digitized straight lines — *Jack Koplowitz* 278
Entropy-constrained subband coding of images using a perceptual distortion criteria — *C. F. Harris and J. W. Modestino* .. 279
Multivariate modeling of subband image statistics using spherically symmetric distributions — *Frank Müller and Christoph Stiller* 280
Optimal predictive coding of 2-D fields — *José M. F. Moura and Nikhil Balram* 281
An optimally bit allocated wavelet pyramid image coding system — *Jie Chen and Shuichi Itoh* 282

THURSDAY SESSIONS

Rotationally invariant trellis codes for QAM — *Eric J. Rossin and Chris Heegard* 283
On 90° rotationally invariant lattice codes — *J. A. Sheppard and A. G. Burr* 284
A semi-algebraic construction to achieve rotationally invariant coded QAM on the basis of multilevel convolutional codes — *Werner Henkel and Michael Koch* 285
Rotationally invariant multilevel codes — *J. N. Livingston* 286
Eight-dimensional modulation for bandlimited channels — *Spase L. Drakul and Ezio Biglieri* 287
Efficient splitting of multidimensional alphabets for modulation codes — *Rolf Johannesson, Joakim Persson, and Kamil S. Zigangirov* 288
High performance and low complexity coded modulation schemes for reliable data communications — *Sandeep Rajpal, Do Jun Rhee, and Shu Lin* 289
Multilevel ternary line codes with trellis structure — *Ümit Aygölü and Erdal Panayirci* 290
On the construction and dimensionality of linear block code trellises — *Alan D. Kot and C. Leung* 291
A new construction method for t-EC/AUED codes based on t-EC codes — *Kenji Yoshida, Hajime Jinushi, and Kohichi Sakaniwa* 292
Some new lower and upper bounds on systematic tEC/AUED codes — *Zhen Zhang, Xiang-Gen Xia, and Chungming Tu* ... 293
Coding for simultaneous correction and detection of skew in parallel asynchronous communications — *Mario Blaum and Jehoshua Bruck* 294
Constructions of skew-tolerant and skew-detecting codes — *Mario Blaum, Jehoshua Bruck, and Levon H. Khachatrian* 295
Superimposed codes in Hamming space — *Thomas Ericson and Vladimir Levenshtein* 296
High-dimensional symmetric compacted code—Error-correcting of high bit error rate of $10^{-1} \sim 10^{-2}$ — *Masayasu Hata and Ichi Takumi* 297
Some families of asymptotically optimal optical orthogonal codes — *O. Moreno, Z. Zhang, and P. V. Kumar* 298
On binary synchronization error correcting codes — *A. S. J. Helberg, H. C. Ferreira, W. A. Clarke, and A. J. Vinck* 299
The two-way channel as a computer game — *Alphons H. A. Bloemen, Hendrik B. Meeuwissen, and J. Pieter M. Schalkwijk* . 300
Ten good rate $(m - r)/pm$ binary quasi-cyclic codes — *T. Aaron Gulliver and Vijay K. Bhargava* 301
On certain subcodes of the binary extended quadratic residue codes — *Xuemin Chen, I. S. Reed, and T. K. Truong* 302
On covering polynomials for binary cyclic codes — *Wonjin Sung and John T. Coffey* 303
A novel approach for construction of algebraic geometric codes from affine plane curves — *G. L. Feng and T. R. N. Rao* ... 304
The automorphism groups of the Delsarte–Goethals codes — *Claude Carlet* 305
Algebraic decoding of Zetterberg and Dumer–Zinoviev codes — *S. M. Dodunekov and J. E. M. Nilsson* 306
On a fast decoding algorithm for Goppa codes defined on certain algebraic curves with at most one higher order cusp — *Norifumi Kamiya and Shinji Miura* 307
The expansion factor of error-control codes — *Ali S. Khayrallah* 308
A method for computing shot-noise cumulative distributions and densities — *John A. Gubner* 309
Distributions and expectations of random variables of interference type — *L. L. Campbell, P. H. Wittke, and A. L. McKellips* . 310
The asymptotic equivalence of investing with and without replacement — *Thomas M. Cover* 311
State prices and Gibbs states — *Michael J. Stutzer* 312
On the optimality and stability of exponential twisting in Monte Carlo estimation — *John S. Sadowsky* 313
Convergence of probability measures for continuous sample paths of multidimensional random field simulations using trigonometric series — *Robert Patton Leland* 314
Wavelet approximation of deterministic and random signals: Convergence properties and rates — *Stamatis Cambanis and Elias Masry* 315
On the minimum expected duration of a coin tossing game — *Inchi Hu and Santosh S. Venkatesh* 316

A simulation study of forward error correction for lost packet recovery in B-ISDN/ATM — *Nihat Cem Oguz and Ender Ayanoglu* 317
A robust error control system for broadcast channels — *S. Ram Chandran and Shu Lin* 318
The capacity per channel of a broad class of noisefree CDMA is for very diverse tasks close to $1/e$ under very loose constraints — *Sándor Csibi* 319
On the delay in a multiple access system with large propagation delay — *Bruce Hajek, N. B. Likhanov, and B. S. Tsybakov* . 320
Constructions of protocol sequences for multiple access collision channel — *László Györfi and István Vajda* 321
Certain generalizations on the collision channel without feedback — *Thomas J. Ketseoglou* 322
Capacity and coding for T active users out of M on the collision channel — *Brian Hughes* 323
A model for the approximation of interacting queues that arise in multiple access schemes — *Eytan Modiano and Anthony Ephremides* 324
Analysis of the exhaustive cycle-gated service scheme — *Irfan Ali and Kenneth S. Vastola* 325
Blind Wiener filtering: Estimation of a random signal in noise using little prior knowledge — *Abhijit A. Shah and Donald W. Tufts* 326
A method of multi-dimensional blind equalization — *Hiroyoshi Oda and Yoichi Sato* 327
Sampling designs for estimation of a random process — *Yingcai Su and Stamatis Cambanis* 328
On sampling theorem, wavelets and wavelet transforms — *Xiang-Gen Xia and Zhen Zhang* 329
Construction of discrete orthogonal wavelet bases — *Chao Wei and Douglas Cochran* 330
Time-warped bandlimited signals: Sampling, bandlimitedness, and uniqueness of representation — *James J. Clark and Douglas Cochran* 331
A unified partitioning and folding procedure for systolic algorithms — *Flavio Lorenzelli and Kung Yao* 332
A coding theorem for low-rate transform codes — *Daniel F. Lyons and David L. Neuhoff* 333
Boundedness and consistency of greedy growing for tree-structured vector quantizers — *Andrew B. Nobel and Richard A. Olshen* 334
Source clustering for codebook compression — *Wai-Yip Chan and Allen Gersho* 335
Self synchronising T-codes to replace Huffman codes — *Gavin R. Higgie* 336
Kieffer's sample converses for source coding — *En-hui Yang* 337
Channel error control of variable-length transform coding — *Ning Guo* 338
Integrated index assignment, source and channel coding — *Petter Knagenhjelm* 339
A new deterministic codebook structure for CELP speech coding — *Yu-Hung Kao and John S. Baras* 340
Lossless compression algorithms for high fidelity audio compression — *Talal Shamoon and Chris Heegard* 341
Constellations designed for the Rayleigh fading channel — *X. Giraud, K. Boullé, and J. C. Belfiore* 342
Multilevel trellis MPSK modulation codes for the Rayleigh fading channel — *Jiantian Wu and Shu Lin* 343
Permuted modulation and coded modulation for interleaved slow fading channels — *François Gagnon* 344
Unified analysis on performance limits of coded multilevel DPSK in Rayleigh fading channels — *Tadashi Matsumoto and Fumiyuki Adachi* 345
Performance of joint equalization and trellis-coded modulation of multipath fading channels — *Mao-Ching Chiu and Chi-Chao Chao* 346
Error rate analysis of trellis-coded modulation and optimum code search for impulsive noise channel — *Haruo Ogiwara and Hiroki Irie* 347
Trellis coded modulation for digital microwave radio — *Tomoko Kodama Matsushima and Hirokazu Tanaka* 348
Optimal multi-h phase codes for partial response continuous phase modulation — *Rongqiang Mao and John P. Fonseka* ... 349
A demonstration of a robust Occam-based learner — *Timothy D. Ross, Michael J. Noviskey, Mark L. Axtell, and Michael A. Breen* 350
Nonparametric regression-based method for neural network training — *Terrence L. Fine and Jen-Lun Yuan* 351
A measure of relative entropy between individual sequences with application to universal classification — *Jacob Ziv and Neri Merhav* 352
On radial basis function net and kernel regression: Approximation ability, convergence rate and receptive field size — *Lei Xu, Adam Krzyzak, and Allan Yuille* 353
On the finite sample performance of the nearest neighbor classifier — *Demetri Psaltis, Robert R. Snapp, and Santosh S. Venkatesh* 354
The relative value of labeled and unlabeled samples in pattern recognition — *Vittorio Castelli and Thomas M. Cover* 355
On the posterior probability estimate of the error rate of nonparametric classification rules — *Gábor Lugosi and Miroslaw Pawlak* 356
A Bayesian approach for classification of continuous-time Markov sources — *Erdal Panayirci* 357
A computer algebra algorithm for the adjoint divisor — *D. Polemi, M. Hassner, O. Moreno, and C. J. Williamson* 358

Codes over Gaussian integers — *Klaus Huber* 359
Construction of linear block codes over groups — *Ezio Biglieri and Michele Elia* 360
Debruijn sequences, irreducible codes and cyclotomy — *E. R. Hauge and T. Helleseth* 361
M-sequences and dual bases over GF(*qm*) — *John J. Komo and William J. Reid, III* 362
Linear recurrences on 2D convex lattices and decoding of some codes from algebraic curves — *Shojiro Sakata* 363
Bounds for linear block codes over rings — *Magnus Nilsson* 364
On designs and formally self-dual codes — *George T. Kennedy and Vera Pless* 365
Greedy codes — *Richard A. Brualdi and Vera Pless* 366
On the upper bound of the size of the *r*-cover-free families — *Miklós Ruszinkó* 367
Packing radius vs covering radius — *Patrick Solé and Philip Stokes* 368
Constructive non-existence proofs for linear covering codes — *René Struik* 369
Perfect tilings of binary spaces — *Gerard Cohen, Simon Litsyn, Alexander Vardy, and Gilles Zémor* 370
The uniqueness of the (9, 18, 4) constant-weight-4 code — *H. F. Mattson, Jr.* 371
Optimization of transmitter pulses for two-user data communications — *Michael L. Honig and Upamanyu Madhow* 372
Optimum sequence multisets for symbol-synchronous code-division multiple-access channels — *Marcel Rupf and James L. Massey* 373
Optimally orthogonal time-limited signals under RMS bandwidth constraints — *Dara Parsavand and Mahesh K. Varanasi* . 374
On achievable inter-user orthogonality for multi-user communication systems in multipath fading environments — *Jürg Ruprecht* 375
Channel coding for asynchronous fiberoptic CDMA communications — *M. Dale and R. Gagliardi* 376
A signaling technique for multiple access laser communications — *John E. Hershey, Nabeel A. Riza, and A. A. Hassan* 377
Variable weight optical orthogonal codes for CDMA networks with multiple performance requirements — *Guu-chang Yang* . 378
Optical spectral amplitude code division multiple access system — *Maïté Brandt-Pearce and Behnaam Aazhang* 379
Improved concatenated coding/decoding for deep space probes — *Dale C. Linne von Berg and Stephen G. Wilson* 380
Changing the coding system on a spacecraft in flight — *Kar-Ming Cheung, Dariush Divsalar, Sam Dolinar, Ivan Onyszchuk, Fabrizio Pollara, and Laif Swanson* 381
A modification of generalized concatenated codes and its applications — *Yan Gao, Uwe Dettmar, and Ulrich Sorger* 382
Performance of concatenated coding systems for channels with memory — *G. Ferland* 383
A performance analysis for adaptive rate, trellis coded hybrid-ARQ protocols — *Lars K. Rasmussen and Stephen B. Wicker* . 384
Further results on the convolutionary coded ARQ with GVA decoding — *Takeshi Hashimoto* 385
An adaptive transmission scheme for meteor-burst communication — *Guy Bégin* 386
Real convolutional codes embedded in multichannel demultiplexers for fault tolerance — *Robert Redinbo* 387
Performance evaluation of trellis-coded vector quantization — *René J. van der Vleuten and Jos H. Weber* 388
An efficient algorithm for optimal tree pruning with application to VQ — *Xiaolin Wu and Yonggang Fang* 389
Asymptotic entropy constrained performance of tesselating and universal randomized lattice quantization — *Tamás Linder and Kenneth Zeger* 390
Joint source and channel coding applied to the pyramid vector quantizer — *Michael J. Ruf and Pavel Filip* 391
Finite-state vector quantization over noisy channels — *Yunus Hussain and Nariman Farvardin* 392
Average number of facets per cell in tree-structured vector quantizer partitions — *Kenneth Zeger and Miriam R. Kantorovitz* . 393
A method for examining vector quantizer structures — *Erik Agrell* 394
An optimal data compression code for memoryless Gaussian source — *Hiroki Koga and Suguru Arimoto* 395

FRIDAY SESSIONS

A general theory of information transfer — *Rudolf Ahlswede* 396
On the probability of undetected error for iterated codes — *Toshihisa Nishijima, Osamu Nagata, and Shigeichi Hirasawa* .. 397
On the key equation for *n*-dimensional cyclic codes — *Hervé Chabanne and Graham H. Norton* 398
On the equivalence of some generalized concatenated codes and extended cyclic codes — *B. Liesenfeld and B. G. Dorsch* .. 399
On cyclic product codes — *B. S. Rajan, H. S. Madhusudhana, and M. U. Siddiqi* 400
Algebraic structure and decoding of two-dimensional cascade codes — *Keith Saints and Chris Heegard* 401
The polynomial of correctable patterns of concatenated codes — *Nicolas Sendrier* 402
Constructive codes for arbitrary DMC and the AGNC — *Michael Steiner* 403
Polyphase pseudo-noise sequences with equivalent even and odd correlation properties — *Ryuji Kohno, Hidenobu Fukumasa, and Hideki Imai* 404
New enumeration results for Costas arrays — *Curtis P. Brown, Michal Cenkl, Richard A. Games, Joseph J. Rushanan, Oscar Moreno, and Pei Pei* 405
A partition of the set of permutations by the monotone subsequence structure — *Kingo Kobayashi and Hiroyoshi Morita* ... 406

Optimal and suboptimal biphase sequence of period $2(2^r-1)$ and linear complexity $r(r+3)/2$ — *P. Udaya and M. U. Siddiqi* . 407
Perfect maps — *Kenneth G. Paterson* . 408
New bounds for the size of radar arrays — *Zhen Zhang and Chungming Tu* . 409
Families of four-phase quasi-orthogonal code arrays — *Serdar Boztas* . 410
Crosscorrelation of GMW sequences — *Markus Antweiler* . 411
Non-binary sequences with the perfect periodic auto-correlation and with optimal periodic cross-correlation — *Ernst M. Gabidulin* . 412
Geometrically uniform multidimensional PSK constellations — *S. Benedetto, R. Garello, M. Mondin, and G. Montorsi* . 413
High-rate punctured convolutional codes for trellis-coded modulation — *François Chan and David Haccoun* . 414
On the design criteria for trellis codes with sequential decoding — *Fu-Quan Wang and Daniel J. Costello, Jr.* . 415
Practical trellis coded modulation with punctured rate-2/3 convolutional codes — *Stephen K. How* . 416
Design of optimal filters for use as bandwidth-efficient coded modulation — *Amir Said and John B. Anderson* . 417
On a class of constant envelope continuous phase modulation schemes, obtained by imposing continuous phase transitions on trellis coded asymmetric PSK — *Johan Udén and Göran Lindell* . 418
A trellis coded modulation scheme constructed from block coded modulation with interblock memory — *Shang-Chih Ma and Mao-Chao Lin* . 419
Universal schemes for sequential decision from individual data sequences — *Neri Merhav and Meir Feder* . 420
Some results on sequential detection of weak signals — *V. N. S. Samarasooriya and P. K. Varshney* . 421
Reduced-complexity iterative maximum-likelihood sequence estimation on channels with memory — *J. W. Modestino* . 422
On sequential delay estimation in wideband digital communication systems — *Yossef Steinberg and H. Vincent Poor* . 423
Performance study of maximum-likelihood receivers and transversal filters for the detection of direct-sequence spread-spectrum signal in narrowband interference — *Arif Ansari and R. Viswanathan* . 424
Optimal detection of discrete Markov sources over discrete memoryless channels—Applications to combined source-channel coding — *Nam Phamdo and Nariman Farvardin* . 425
A communication channel modeled by the spread of disease — *Fady Alajaji and Tom Fuja* . 426
Demodulation of AM–FM signals in noise using multiband energy operators — *Alan C. Bovik, Petros Maragos, and Thomas F. Quatieri* . 427
Information theory and radar waveform design — *Mark R. Bell* . 428
An upper bound of the capacity of Hopfield net with perceptron algorithm — *Shiyi Shen and Zhongxing Ye* . 429
Corrective memory by a symmetric sparsely encoded network — *Y. Baram* . 430
Strong universal consistency of neural network classifiers — *András Faragó and Gábor Lugosi* . 431
Sample size requirements of feedforward neural network pattern classifiers — *Terrence L. Fine and Michael J. Turmon* . 432
Constructions of depth-2 majority circuits for comparison and addition using linear block codes — *Noga Alon and Jehoshua Bruck* . 433
Capacity of two-layer networks with binary weights — *Chuanyi Ji and Demetri Psaltis* . 434
A new fixed-rate quantization scheme based on arithmetic coding — *Ahmed S. Balamesh and David L. Neuhoff* . 435
Statistics of the binary quantizer error in sigma–delta modulation with i.i.d. Gaussian input — *Timo Koski* . 436
Design of entropy constrained multiple-description scalar quantizers — *J. Domaszewicz and V. Vaishampayan* . 437
Enumeration encoding and decoding algorithms for pyramid trellis codes — *Thomas R. Fischer and Jianping Pan* . 438
Information rates of pre/post filtered dithered quantizers — *Ram Zamir and Meir Feder* . 439
A frequency domain approach to the optimization of scalar quantizers — *Marcelo S. Alencar* . 440
Performance comparisons of TCQ with difference distortion measures and short coding delays — *Min Wang and Thomas R. Fischer* . 441
Asymptotic quantization for noisy channels — *Steven W. McLaughlin and David L. Neuhoff* . 442
The design of finite-state machines for quantization using simulated annealing — *Ercan Engin Kuruoglu and Ender Ayanoglu* . 443

PAPERS THAT WERE NOT SUBMITTED IN TIME FOR PRINTING

Asymptotically Optimal Rate of Binary Codes with Constant Weight Which Correct Localized Errors.
L. A. Bassalygo and M. S. Pinsker

Formulation of Perfect Uniquely Decodable Code Pair.
X. Shang, F. Guo, and Y. Watanabe

Detecting and Localizing Discontinuities in Nonparametric Regression.
M. Pawlak

Strengthening Viterbi's Upper Bond on the Bit Error Probability for Maximum Likelihood Decoding of Fixed Binary Convolutional Codes.
R. Johannesson and K. S. Zigangirov

On the Solution of the Minimal Rational Interpolation Problem and the Incomplete Iteration Decoding of Reed-Solomon Codes.
X. Dingjia

Nonparametric Identification of linear Continuous-Time Systems.
A. A. Georgiev

Nonparametric Regression Function Estimation with Order Statistics and Application to Nonlinear System Identification.
W. Greblicki and M. Pawlak

On the Zero-Error Capacity of Broadcast Channels.
J. Korner and G. Simonyi

The Capacity of the Arbitrary Varying Channel in List Decoding Case.
V. M. Blinovsky, P. Narayan, and M. S. Pinsker

Exact Convergence of a Parallel Textured Algorithm for Data Network Optimal Routing Problems.
C. M. Huang, Sr. and W. L. Hsieh

On the Maximal Number for DS-CDMA Cellular System Over Multipath Fading Channels.
E. Zehavi and D. Keschet

Signal Detection and Channel Capacity for a Class of Nongaussian Noise Processes.
C. R. Baker

Blowing-up Properties of Stationary Processes and their Connection to Ergodic Properties.
K. Marton

Accuracy and Reliability of 2-D Objects localization in Pictures.
L. P. Yaroslavsky

Galois Theory and the Fast Gelfand Transform.
U. Oberst

Bounds for Codes in Hamming Space with Given Minimum and/or Dual Distance.
V. I. Levenshtein

A New Class of Error-Control Codes for high Speed Adaptive Feedback Coding Systems.
J. Fan, C. Zhi, and K. K. Tzend

On the Rate-Distortion Performance and Complexity of a multistage VQ.
Z. Zhang and V. K. Wei

Convergence Time and Memory Capacity of Complete Fully Interconnected Higher Order Hopfield Networks.
D. Y. Chao and D. T. Wang

Lower Bounds on Threshold and Related Circuits via Communication Complexity.
V. P. Roychowdhury, K. Y. Siu, and A. Orlitsky

NETWORK PARADOXES AND THE INEFFICIENCY OF NONCOOPERATIVE GAMES

Joel E. Cohen
Rockefeller University
1230 York Avenue, Box 20
New York, NY 10021

Summary

One might think that adding an additional road to a traffic network would improve, or at least not worsen, the time travelers take to go from a given origin to a given destination in the network. In 1968, D. Braess showed that adding a road to a congested traffic network can sometimes worsen the travel time from origin to destination for all travelers. Analogous surprises can occur in networks of queues: adding servers may slow the average time through a network for all travelers. (Strangely, in queuing networks, giving travelers more information about queue lengths may make them worse off than giving them less information.) These results are special cases of a general theorem, due to P. Dubey in 1986: in n-person noncooperative games with smooth payoff functions, Nash equilibria are generically Pareto-inefficient. This tutorial talk will assume no prior background in the theory of traffic networks, queues or games.

Block-decodable Runlength-limited Codes Via Look-ahead Technique

K.A. Schouhamer Immink

Since the early 1970s, coding methods based on (d,k)-constrained sequences have been widely used in such high-capacity storage systems as magnetic and optical disks or tapes. Properties and applications of (d,k)-constrained sequences, or runlength-limited (RLL) sequences as they are often called, are surveyed in [1]. The number of sequential like symbols in a (binary) sequence is known as *runlength*. A (d,k) RLL sequence is a sequence of binary symbols characterized by two parameters, $(d+1)$ and $(k+1)$, which stipulate the minimum and maximum runlength, respectively, that may occur in the sequence. Closely related to RLL sequences are (d,k) sequences. A binary sequence is said to be (d,k) constrained if the number of 'zeros' between any pair of consecutive 'ones' is at least d and at most k, $k > d$. A (d, k)-constrained sequence is converted into a (d,k) RLL sequence by a simple coding step which is known as *precoding*. The 'ones' in the (d,k)-constrained sequence indicate the positions of a transition $1 \rightarrow 0$ or $0 \rightarrow 1$ of the corresponding runlength-limited sequence.

Codes are used to translate source data into the constrained sequence. Commonly, the source data is partitioned into words of length m, and under the coding rules, these m-tuples are translated into n-tuples, called codewords. Popular (d, k) codes incorporated in disk file systems are the (2,7) and the (1,7) codes of rate $m/n = 1/2$ and rate 2/3, respectively [1]. The codes are designed using the bounded delay method [2] or the ACH sliding-block code algorithm [3]. The principal feature of a (d,k) (or other finite-type constraints) code constructed with the sliding-block code algorithm is that coded sequences can be decoded by examining a limited number of consecutive symbols without relying on external state information. As an immediate consequence, these codes have a limited amount of error propagation. For example, a single bit error in a received sequence encoded by the (2,7) sliding-block code propagates at most over four decoded bits. Blaum [4] showed that the error propagation of sliding-block codes presents a problem as it entails an extra load to the error correction circuitry usually used in conjunction with the (d, k) code.

An alternative to the above sliding-block coding scheme was proposed by Tang and Bahl [5]. There, the authors use codes compiled from codewords of fixed length which can be decoded without the knowledge of preceding or succeeding codewords. Codes with this property, that is, codes that can be decoded by observing single codewords (the encoding operation is allowed to be state dependent), will be called *block(-decodable) codes*. Evidently, block-decodable codes offer an advantageous solution relative to sliding-block codes since they make it easier to preserve a particular mapping between the source and the code symbols, and, obviously, error propagation is localized to one decoded m-block. Block-decodable codes are highly suitable in conjunction with Reed-Solomon error control codes. In the preferred embodiment of the coding system, the codewords have a 1-1 correspondence with the elements of the finite field $GF(2^m)$, thus enabling the construction of, for instance, a Reed-Solomon code directly over the (d,k)-constrained codewords. A notable drawback of state-of-the-art block-decodable code constructions, however, is the fact that at code rates $R = m/n$ approaching the Shannon capacity of the (d,k)-constrained channel, the implementations can be fairly complex, involving long codewords. For example, the minimum codeword lengths allowing a rate $R = 2/3$, (1,7) block-decodable code and a rate $R = 1/2$, (2,7) block-decodable code are 33 and 34, respectively. Our design approach is based on the observation that good codes must be constructed on RLL sequences rather than (d,k) sequences. In the literature, the terms (d,k) sequence and RLL sequence are usually used as synonyms, and the design of encoders that generate RLL sequences is almost always conducted by designing encoders that generate (d,k) sequences followed by a precoder. It is generally believed that this strategy does not entail a loss of performance in terms of coder complexity and error propagation. It will be shown, however, that it is surprisingly profitable in terms of error propagation to design RLL encoders directly, i.e. without the intermediate step of a (d,k)-constrained sequence. The new RLL codes to be discussed are block decodable, while at the same time they are simpler to implement than (d,k) block-decodable codes currently being used [6].

References

[1] K.A.S. Immink, *Coding Techniques for Digital Recorders*, Prentice-Hall International (UK) Ltd., Englewood Cliffs, New Jersey, 1991.

[2] P.A. Franaszek, 'On Future-dependent Block Coding for Input-restricted Channels', *IBM J. Res. Develop.*, vol. 23, pp. 75-81, 1979.

[3] R.L. Adler, D. Coppersmith, and M. Hassner, 'Algorithms for Sliding Block Codes. An Application of Symbolic Dynamics to Information Theory', *IEEE Trans. Inform. Theory*, vol. IT-29, no. 1, pp. 5-22, Jan. 1983.

[4] M. Blaum, 'Combining ECC with Modulation: Performance Comparisons', *IEEE Trans. Inform. Theory*, vol. IT-37, no. 3, pp. 945-949, May 1991.

[5] D.T. Tang and L.R. Bahl, 'Block Codes for a Class of Constrained Noiseless Channels', *Information and Control*, vol. 17, pp. 436-461, 1970.

[6] K.A.S. Immink, 'Block-decodable Runlength-limited Codes Via Look-ahead Technique', *Philips J. Res.*, vol. 46, no. 6, pp. 293-310, 1991.

The author is with Philips Research Laboratories, P.O. Box 80.000, 5600 JA Eindhoven, The Netherlands. E-mail: immink@prl.philips.nl

On the Optimization of Constrained Channel Codes

P. A. Franaszek J. A. Thomas
IBM T.J.Watson Research Center, P.O.Box 704, Yorktown Heights, NY 10598.

Abstract

A variety of techniques have been proposed for the construction of codes for input restricted channels, including variable length codewords, codes based on partitioning the state successor trees, and methods based on state splitting. In this paper, new methods that avoid exhaustive search are proposed for partitioning state successor trees into subtrees termed independent paths that are used for the coding. First, the approximating eigenvector algorithm is used to determine the weights of the states. These weights are then partitioned into integer parts, which are used to form the independent paths (IP's). Consistency of the weights alone is checked in the first phase of the algorithm, and in the second phase, the partitions are used to form the IP's and to allot information symbols. These methods can also be used to determine the sequence of splits needed for the state splitting technique.

The problem of encoding information to fit constraints has many applications in magnetic recording and optical communication. Given a finite-state desciption of the channel constraints, and a desired coding rate, one way to obtain a code [1, 2] involves the partition of a state successor tree (the tree of paths starting at a particular state) into subtrees with specified properties. For one class of codes, the subtrees correspond[3] to ones associated with a set of split states such as obtained via the method of Adler, Coppersmith and Hassner[4] It is usually desirable to find a code corresponding to a tree partition at minimal depth, or equivalently (for stationary codes) requiring a minimum number of splits. Moreover, practical considerations generally dictate that the partition depth D be bounded by a small integer.

One way to proceed is via an exhaustive search over all possible state splits. We here describe an algorithm which organises the search somewhat differently, first considering only integer partitions of state weights in the tree (no matter how obtained), then, once a complete partition has been obtained, reconstructing potential candidate tree partitions. We begin with the definition of an independent path, a central concept of the method.

Definition: An independent path (IP) of length N is a set of paths, starting at some state σ , with the property that they can represent one sequence of N input blocks, followed by anything else.

Consider the state successor tree associated with state σ. The encoder mapping defines equivalence classes of paths: those that correspond to the same sequence of input bits of length N are in the same class. To distinguish between equivalence classes given only the set of output states then requires that each class correspond to distinct states at some level of the state successor tree. The equivalence classes correspond to IPs, and the decodability requirement is that IPs be distinguishable at some depth D. The partition of the state successor tree into IP's is based on necessary conditions for the existence of a code rather than sufficient conditions as in [4].

Given a partition (i.e. a set of IPs), one may then construct a code[1, 2]. The number of IPs that start at each state (the weight of the state) corresponds to a component of an approximating eigenvector. At each level of the tree, the weight of the node is split among the different IP's that go through the node. At leaves of the tree, each state lies entirely within a single IP, and the entire weight is associated with that IP (a *natural* partition). At higher levels, the partition of the weight of a state σ_i must be a combination of the subweights in the partitions of the successor states to σ_i, since each IP going through state σ_i corresponds to a set of subweights of the successor states of σ_i. Thus a search for a partition of the tree can be reduced to a search for a set of compatible partitions of the weights of the states of the tree. A tree partition can be obtained as follows:

1. Choose a rate for the code p/q. Calcalate the weights of the states using the $(A^q, 2^p)$ approximate eigenvector algorithm[1, 2, 4, 5].
2. **Algorithm for partitioning successor trees into IP's**
   The UP phase:
   - For a specific depth of tree $R \geq 1$, starting with a natural partition at depth $i+1$, find potential partition of weights at depth i that will be compatible with the partitions calculated at depth $i+1$.
     Stop when a complete partition (1,1,...,1) is obtained at the root of the tree.

   The DOWN phase: a depth first search for a compatible partition.
   - Starting with a complete partition (1,1,...,1) at the root, for $i = 0, 1, \ldots, R-1$, choose a partition at depth $i+1$ that is compatible with the chosen partition at depth i.
     Stop when a natural partition is obtained at the leaves of the tree.
3. The partition of the states yields IP's, to which input sequences may be alloted using the methods given in [2].

By varying the conditions on the successor tree partitions, for example by permitting a dependence on the path used to reach a state, there is a potential for more general code structures. The up phase, using only state weights, provides a means for bounding the minimum required value of the depth D. For stationary codes, the algorithms provides an alternative to searching over all possible state splits, and provide a lower bound on the number of rounds of state splitting required.

References

[1] P. Franaszek. A general method for channel coding. *IBM J. Res. Develop.*, 24:638–691, 1980.

[2] P. Franaszek. Construction of bounded delay codes for discrete noiseless channels. *IBM J. Res. Develop.*, 26:506–514, 1982.

[3] P. Franaszek. Coding for constrained channels: a comparison of two approaches. *IBM J. Res. Develop.*, 33(6):602–608, November 1989.

[4] R.L. Adler, D. Coppersmith, and M. Hassner. Algorithms for sliding block codes — an application of symbolic dynamics to information theory. *IEEE Trans. Inform. Theory*, IT-29(1):5–22, 1983.

[5] B.H. Marcus, P.H. Siegel, and J.K. Wolf. Finite state modulation codes for data storage. *IEEE Journal on Selected Areas in Communication*, 10(1):5–37, January 1992.

Construction of Polynomial-Size Encoders with Small Decoding Look-ahead for Input-Constrained Channels

JONATHAN J. ASHLEY* BRIAN H. MARCUS* RON M. ROTH†

Input-constrained channels, also known as *constrained systems*, are widely-used models for describing the read-write requirements of secondary storage systems, such as magnetic disks or optical memory devices. A constrained system S is defined as the set of *constrained sequences* obtained by reading the labels of paths of a finite labeled directed graph G.

One goal in the study of constrained systems is designing encoders that map unconstrained binary sequences, referred to as *source sequences*, into constrained sequences of a given constrained system S. A *rate $p:q$ finite-state encoder* encodes a p-block of source symbols to a q-block in S in a state-dependent manner.

An encoder is *lossless of finite order if there is an* integer N such that the encoder state at each time slot r, together with the q-block generated at times $r, r+1, \ldots, r+N-1$, determine uniquely the source p-block that was input at time slot r. The smallest number for which this is possible is called the *order* of the encoder. An encoder is *sliding-block decodable* if the source sequence which was input to the encoder can be reconstructed by applying a decoding function on a 'window' of symbols in the constrained sequence.

Several schemes have been suggested for constructing finite-state encoders, most notable of which is the Adler-Coppersmith-Hassner (state splitting) algorithm [1]. The latter provides encoders which are lossless of finite-order. Furthermore, for the important subclass of constrained systems of finite memory (such as (d,k)-run-length-limited systems), the resulting encoders are sliding-block decodable. However, there are no known polynomial upper bounds, in terms of the number of states in a deterministic graph presentation G, on the window length or implementation size.

*IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120.

†Computer Science Department, Technion — Israel Institute of Technology, Haifa 32000, Israel.

In [3], Ashley constructed encoders with order which is linear in the number k of states in a deterministic graph G that presents the constrained system S. The resulting encoders have rate $p\ell : q\ell$ for some $\ell = O(k)$. When these encoders are translated into rate $p:q$ encoders, the latter have order $O(k)$, but typically they are not sliding-block decodable.

In this work, we present a class of encoders, called *stethering encoders*, based on a construction of Adler, Goodwyn, and Weiss in [2]. Using complexity results, we show that the number of gates and memory-cells required for a hardware implementation of these encoders is at most polynomial in k. We show that this also holds for the construction in [3].

Then we show that for any constrained system S and any positive integers p and q such that $p/q < c(S)$, stethering encoders have order which is at most linear in k and is slightly smaller than the one guaranteed in [3]. We show that the stethering encoders and those in [3] have polynomial-size decoders.

For constrained systems S of finite memory and rate $p/q \leq c(S) - ((\log_2 e)/(2^p q))$, stethering encoders are sliding-block decodable with window size at most quadratic in k.

References

[1] R.L. ADLER, D. COPPERSMITH, M. HASSNER, *Algorithms for sliding block codes — an application of symbolic dynamics to information theory*, *IEEE Trans. Inform. Theory*, IT-29 (1983), 5–22.

[2] R.L. ADLER, L.W. GOODWYN, B. WEISS, *Equivalence of topological Markov shifts*, *Israel J. Math.*, 27 (1977), 49–63.

[3] J.J. ASHLEY, *A linear bound for sliding block decoder window size*, *IEEE Trans. Inform. Theory*, IT-34 (1988), 389–399.

ENUMERABLE MULTI-TRACK (d, k) BLOCK CODES

EDWARD K. ORCUTT AND MICHAEL W. MARCELLIN
Department of Electrical and Computer Engineering
The University of Arizona
Tucson, AZ 85721

Recently, a new class of run-length-limited codes, referred to as two-dimensional or multi-track modulation codes, has been developed [1]. These codes are useful in applications such as digital magnetic recording in which constraints are placed on both the minimum and maximum number of 0's which occur between 1's. These constraints are parameterized by d and k, respectively.

In NRZI encoding, a '1' is represented by a transition change in polarity states while a '0' is symbolized by no transition. In order to prevent adjacent transitions from interfering with each other (intersymbol interference (ISI)), it is necessary to ensure that some number of '0's occur between '1's. This necessitates the d constraint. Also, because timing information is extracted from the data itself, transitions must not be too far apart. Hence, the k constraint, which limits the number of consecutive '0's which occur between '1's, is enforced.

In the past, (d, k) codes have been devised such that each track individually satisfies both constraints. This results in the average capacity for all tracks to be equal to the capacity of single-track (d, k) constraints; where *capacity* is the theoretical maximum code rate (ratio of source bits to code bits). However, by having each track satisfy the d constraint (ISI must be controlled in each track) but using multiple tracks to satisfy the k constraint, increased capacity can be realized, relative to the conventional single-track (d, k) code [1]. This increase in capacity is realized because, in effect, the k constraint has been relaxed. For multi-track codes, it is assumed that all n tracks are read in parallel and used (jointly) to derive clocking as opposed to single-track codes in which each track individually is required to meet the k constraint.

For example, consider the following sequences which satisfy a two-track $(1, 2)$ constraint.

track 1 00001010001010010
track 2 01000001000000100

It is observed that although both tracks individually have runs of 0's longer than $k = 2$, both tracks are never 0 simultaneously more than twice consecutively.

In this paper, we propose a method to construct multi-track (d, k) block codes which can be implemented using an enumeration scheme based on a trellis. While block codes can be implemented via look-up tables, the amount of memory required for such an implementation increases exponentially with the block length. Our method is a computational algorithm which requires only a linear increase of memory (and computations) with block length [2].

Enumeration is a process in which the elements of a given set are assigned an index according to their lexicographical order. For m-tuples of numbers, a lexicographic ordering can be defined as follows. For $\mathbf{x} = (x_0, \ldots, x_{m-1})$ and $\mathbf{y} = (y_0, \ldots, y_{m-1})$, then $\mathbf{x} < \mathbf{y}$ if there exists some index i such that $x_i < y_i$ and $x_j = y_j \ \forall \ j < i$. We incorporate a modified trellis description of the multi-track (d, k) channel constraints to devise easily enumerable block codes with good code rates.

This work was supported by International Business Machines Corporation and by the National Science Foundation under Grant No. NCR-9258374.

Consider an output symbol $c(l)$ at time l derived from the outputs of each of the n tracks $z_i(l)$ as

$$c(l) = \sum_{i=0}^{n-1} z_i(l) 2^i. \tag{1}$$

Hence, a new symbol set $\mathcal{A} = \{0, 1, \ldots, 2^n - 1\}$ is formed and a codeword of length m is described by $\mathbf{c} = (c_0, c_1, \ldots, c_{m-1}) \in \mathcal{C}$ where $\mathcal{C} \subset \mathcal{A}^m$ is the set of codewords (chosen to satisfy the multi-track (d, k) constraints). The goal of our enumeration scheme is to lexicographically order the set of codewords.

Consider the one-step state transition matrix $\mathbf{B}_n(d, k)$ in which a value of 1 for the $(ij)^{th}$ element indicates that state S_j can immediately follow state S_i. This information can also be summarized in the form of a trellis, indicating allowable state transitions. In our method, we augment the trellis description by associating 2^n values $N_p^{i(l)}(l)$ $(p = 1, 2, \ldots, 2^n)$ with each node. The argument l represents the time index of the trellis and the superscript $i(l)$ denotes the index of a state at time l in the trellis. For a given terminal set, these $N_p^{i(l)}(l)$ describe the number of sequences of length $m - l$ which begin in that state with a symbol less than p and end in any terminal state.

The determination of the lexicographic number of a given codeword is based upon knowing the number of allowable sequences which begin with each of the elements of $\mathcal{A}$ at each node of the trellis; i.e., the $N_p^{i(l)}(l)$. We associate a unique integer value with each allowable codeword with the resulting set of integer values forming a contiguous set. This set takes the form $\{0, 1, \ldots, \mid \mathcal{C} \mid -1\}$ where $\mid \mathcal{C} \mid$ is the number of codewords in the code.

Encoding of a K-bit binary source word begins by first converting it to its decimal equivalent $\tilde{v}$. Then a path corresponding to both $\tilde{v}$ and the $N_p^{i(l)}(l)$ is traversed through the modified trellis. At the path's completion, the codeword which results is that whose value in the lexicographic ordered set of all codewords is equal to $\tilde{v}$.

Decoding of a given codeword $\mathbf{c}$ entails following a path (dictated by $\mathbf{c}$) through the modified trellis and keeping a running sum of the $N_p^{i(l)}(l)$ encountered at each node of the path. At the path's completion, the value of the running sum is equal to the lexicographic number of $\mathbf{c}$.

An interesting facet of this scheme is that the modified trellis contains all the relevant information concerning both decoding and encoding of codewords. Hence, the actual codewords under this scheme need not be known. They are generated automatically by the algorithm.

REFERENCES

[1] M. Marcellin and H. Weber, "Two-dimensional modulation codes," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 254–266, Jan. 1992.

[2] E. K. Orcutt, "Encoding of multi-track (d, k) modulation codes," Ph.D. dissertation, The University of Arizona, Aug. 1992.

A UNIVERSAL ALGORITHM FOR GENERATING OPTIMAL AND NEARLY OPTIMAL RUN-LENGTH-LIMITED, CHARGE-CONSTRAINED BINARY SEQUENCES

Paul E. Bender
QUALCOMM Incorporated
10555 Sorento Valley Road
San Diego, CA 92121

Jack K. Wolf
Center for Magnetic Recording Research
University of California, San Diego
La Jolla, CA 92093

Abstract: This paper presents an algorithm for run-length-limiting and charge-constraining binary data. These constraints are specified by the three parameters (d,k,c). The first two constraints, d and k, put a lower and an upper bound on the run-lengths. The third parameter, c, puts an upper bound on the absolute accumulated charge. An algorithm is optimal if its maximum average rate equals the capacity of the constraint. The algorithm that this paper presents, known as the bit stuff algorithm, is a variable rate algorithm that is both simple and universal. It is optimal for the (d,∞,∞), the $(d,d+1,\infty)$, and the $(2c-2,\infty,c)$ constraints. It is nearly optimal for all other constraints.

Introduction

In [1], Lee presented an algorithm for sequentially satisfying the k and c constraints of a $(0,k,c)$ constrained sequence by inserting bits into an arbitrary data sequence. We present the bit stuff algorithm for simultaneously satisfying the k and c constraints of a (d,k,c) constrained sequence by inserting bits into an arbitrary data sequence.

Bit Stuff Algorithm

The encoder for the (d,k,c) bit stuff algorithm uses two variables to keep track of the information need to correctly insert the extra bits. The first variable, k', keeps track of the current run-length, where a run is a string of consecutive 0's. If k' is ever equal to k, then the encoder inserts a 1 to avoid a possible violation of the k constraint. The second variable, c', keeps track of the accumulated charge in the opposite direction of the current run's charge. If c' is ever equal to $-c+1$, then the encoder inserts a 1 to avoid a possible violation of the c constraint. Then, after every 1 the encoder inserts d 0's to avoid a possible violation of the d constraint.

The decoder also keeps track of the variables k' and c', using the values to delete the extra bits inserted by the encoder.

Performance

We model the encoder by a variable length constraint graph [2, 3]. The states in our graph represent c'. The edges in our graph represent the allowable runs. When $0<d+1\le k\le 2c-1$, we can represent the graph for such a (d,k,c) constrained sequence by the $(2c-d)\times(2c-d)$ matrix

$$A(D_0,D_1)=\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & D_0^dD_1^1 \\ 0 & 0 & \cdots & 0 & D_0^dD_1^1 & \vdots \\ \vdots & \vdots & & & \vdots & D_0^{k-1}D_1^1 \\ 0 & 0 & D_0^dD_1^1 & \cdots & D_0^{k-1}D_1^1 & D_0^kD_1^1 \\ 0 & D_0^dD_1^1 & \cdots & D_0^{k-1}D_1^1 & D_0^kD_1^1 & \vdots \\ D_0^dD_1^1 & \cdots & D_0^{k-1}D_1^1 & D_0^kD_1^1 & \cdots & 0 \end{bmatrix}$$

where the superscripts of D_0 and D_1 represent the number of 0's and 1's respectively.

The capacity of such a graph is, $C(d,k,c)=-\log_2\lambda$, where λ is the smallest positive root of the characteristic equation $\det(\mathbf{I}-\mathbf{A}(z,z))=0$.

In order to get a description of the corresponding data sequences for the bit stuff encoder, we remove the 0's and 1's inserted by the bit stuff algorithm from the variable length graph described by $\mathbf{A}(D_0,D_1)$, getting the $(2c-d)\times(2c-d)$ matrix

$$\hat{\mathbf{A}}(D_0,D_1)=\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & D_0^0D_1^1 \\ 0 & 0 & \cdots & 0 & D_0^0D_1^1 & \vdots \\ \vdots & \vdots & & & \vdots & D_0^{k-d-1}D_1^1 \\ 0 & 0 & D_0^0D_1^1 & \cdots & D_0^{k-d-1}D_1^1 & D_0^{k-d}D_1^0 \\ 0 & D_0^0D_1^1 & \cdots & D_0^{k-d-1}D_1^1 & D_0^{k-d}D_1^0 & \vdots \\ D_0^0D_1^0 & \cdots & D_0^{k-d-1}D_1^0 & D_0^{k-d}D_1^0 & \cdots & 0 \end{bmatrix}$$

Then, assuming that our data are independent, identically distributed, binary digits with the probability of a 0 given by p, the state transition probability matrix for the (d,k,c) bit stuff code is $\hat{\mathbf{A}}(p,1-p)$.

From the state transition probabiliy matrix, we can find p_i, the probability of being in a charge state i, where $-c+d+1\le i\le c$. In terms of these known values, the average input length is

$$L_{in}(p,d,k,c)=\frac{1-\left(p^{c-d-1}\sum_{i=-c+d+1}^{-c+k+1}p_ip^i+p^{k-d}\left(1-\sum_{i=-c+d+1}^{-c+k+1}p_i\right)\right)}{1-p}$$

and the average output length is

$$L_{out}(p,d,k,c)=L_{in}(p,d,k,c)+d+p_c+p^{k-d}\left(1-\sum_{i=-c+d+1}^{-c+k+1}p_i\right).$$

Therefore, the average information rate is

$$I(p,d,k,c)=\left(\frac{L_{in}(p,d,k,c)}{L_{out}(p,d,k,c)}\right)\left((p)\log_2\left(\frac{1}{p}\right)+(1-p)\log_2\left(\frac{1}{1-p}\right)\right).$$

We maximize the average information rate with respect to p in order to find the maximum average information rate of the bit stuff code for a particular (d,k,c) constraint.

Optimal Special Cases

For the (d,∞,∞) constraint, maximizing the equation for the average information rate and setting $p_{opt}=\lambda$, we have the parametric equations for the capacity of a (d,∞,∞) code.

For the $(d,d+1,\infty)$ constraint, maximizing the equation for the average information rate and setting $p_{opt}=\lambda^{d+2}$, we have the parametric equations for the capacity of a $(d,d+1,\infty)$ code.

For the $(2c-2,\infty,c)$ constraint, maximizing the equation for the average information rate and setting $p_{opt}=\lambda^{2c}$, we have the parametric equations for the capacity of a $(2c-2,\infty,c)$ code.

Sub-Optimality of Remaining Cases

Now we argue that the remaining cases are sub-optimal. All bit stuff (d,k,c) constrained sequences that we have not shown to be optimal are in the category $0<d+1<k+1<2c-1\le\infty$. We consider the two equally probable data sequences $[0]^{k-d}[0]^{k-d}111$ and $[0]^{k-d-1}1[0]^{k-d-1}1010$. If the bit stuff encoder starts in state $-c+1$, then it will end in state $-c+1$ and output the sequences $1[0]^k1[0]^k1[0]^d1[0]^d1[0]^d1[0]^d$ and $1[0]^{k-1}1[0]^{k-1}1[0]^{d+1}1[0]^{d+1}$. Since both sequences have the same starting and ending state, both sequences have an identical set of predecessors and successors. However, the output sequence for the first data sequence is longer than the output sequence for the second data sequence. Therefore, in order to maximize the average information rate, the probability of the first data sequence must be less than the probability of the second data sequence, which contradicts the fact that the data sequences are equally probable.

Although these remaining cases are sub-optimal, numerical maximization of the average information rate shows that they are nearly optimal.

Conclusions

We have presented a simple and efficient algorithm for (d,k,c) constraining data. We have shown that the algorithm is optimal for (d,∞,∞), $(d,d+1,\infty)$, and $(2c-2,\infty,c)$ constraints and nearly optimal for all other (d,k,c) constraints.

References

[1] P. Lee, Combined error-correcting/modulation recording codes, Ph.D. Dissertation, University of California at San Diego, 1988.

[2] K. Norris and D. S. Bloomberg, "Channel capacity of charge-constrained run-length-limited codes," IEEE Transactions on Magnetics, vol. MAG-17, pp. 3452-3455, November 1981.

[3] K. J. Kerpez, A. Gallopoulos, and C. Heegard, "Maximum entropy charge-constrained run-length codes," IEEE Journal on Selected Areas in Communications, vol. 10, pp. 242-253, January 1992.

Design of some new Balanced Codes.

L. Tallini
Department of
Computer Science,
Oregon State University,
Corvallis, OR 97331,
tallini@cs.orst.edu

R. M. Capocelli
Dipartimento di
Informatica,
University of Rome
"La Sapienza",
Via Salaria 113,
00100 Roma, ITALY

B. Bose
Department of
Computer Science,
Oregon State University,
Corvallis, OR 97331,
bose@cs.orst.edu

A binary word of length $n \in \mathbb{N}$ is called balanced when it has $\lceil \frac{n}{2} \rceil$ ($\lfloor \frac{n}{2} \rfloor$) 1's and $\lfloor \frac{n}{2} \rfloor$ ($\lceil \frac{n}{2} \rceil$) 0's. A code $\mathcal{C}$ is a balanced code with r check bits and k information bits iff:

1. $\mathcal{C}$ has fixed length $n = k + r$,
2. each word $X \in \mathcal{C}$ is balanced,
3. $|\mathcal{C}| = 2^k$.

In [4], Knuth showed that if a balanced code with r check bits and k information bits exists, then $r > \frac{1}{2}\log_2 k + 0.326$; he has designed serial encoding and both serial and parallel decoding schemes with $k = 2^r$ and $k = 2^r - r - 1$ respectively. In both methods, for each given information word, some appropriate number of bits, starting from the first bit, are complemented; then a check is assigned to this modified information word to make the entire word balanced. In the sequential decoding the check represents the weight of the original information word whereas in the parallel decoding the check directly indicates the number of information bits complemented.

In [1], [2] and [3] improved design methods are given.

In this paper, we divide the set of information words into two subsets: 1)the subset of words that are close to balanced and 2)the subset of words that are not close to balanced; then we encode words in each subset with different methods. More precisely, given $t \in \mathbb{N}$, let ($w(X)$ is the weight of X):

$$\mathcal{U}_t \stackrel{\text{def}}{=} \{X \in \mathbb{Z}_2^k : 0 \le w(X) \le t \text{ or } k - t \le k - w(X) \le k\}$$

and:

$$\mathcal{B}_t \stackrel{\text{def}}{=} \mathbb{Z}_2^k - \mathcal{U}_t$$

be the subsets of information words close to balanced and not close to balanced respectively. The words are made balanced by encoding $\mathcal{U}_t$ using tail-maps and encoding $\mathcal{B}_t$ using single maps defined by Knuth's complementation method.

Three different tail-maps are given and here one of these maps is briefly described.

Tail-map construction 1: Here, the word is divided into $\lceil \frac{k}{2} \rceil$ two bits and each part is encoded into unary with the function $u : \mathbb{Z}_2 \cup \mathbb{Z}_2^2 \longrightarrow \mathbb{Z}_2^*$. More formally, given $k \in \mathbb{N}$, let:

$$X \stackrel{\text{def}}{=} x_1 x_2 x_3 x_4 \dots x_{k-2} x_{k-1} x_k = b_1^X b_2^X \dots b_{\lceil \frac{k}{2} \rceil - 1}^X b_{\lceil \frac{k}{2} \rceil}^X,$$

where:

$$b_i^X \stackrel{\text{def}}{=} \begin{cases} x_{2i-1}x_{2i} & \text{if } i \in [1, \lfloor \frac{k}{2} \rfloor], \\ x_k & \text{if } i = \lceil \frac{k}{2} \rceil \neq \lfloor \frac{k}{2} \rfloor. \end{cases} \quad (1)$$

Each b_i^X can be considered as the binary encoding of an integer number between 0 and 3. In this way X can be identified by a sequence of $\lceil \frac{k}{2} \rceil$ integer numbers between 0 and 3 and so we can encode X using the unary representation of such sequence. Let:

$$U(X) \stackrel{\text{def}}{=} u(b_1^X)u(b_2^X)\dots u(b_{\lceil \frac{k}{2} \rceil}^X), \quad (2)$$

and $\left(\bigcup_{i=0}^{t} \mathcal{S}_i^k\right) \cup \left(\bigcup_{i=k-t}^{k} \mathcal{S}_i^k\right) \stackrel{<Y>}{\longrightarrow} \mathcal{S}_{\frac{k}{2}}^{\kappa}$ the tail-map defined as follows ($\overline{X}$ is the complement of X):

$$< Y > (X) \stackrel{\text{def}}{=} \begin{cases} U(X)0^{(k-1)-l(U(X))}0 & \text{if } X \in \bigcup_{i=0}^{t} \mathcal{S}_i^k, \\ \overline{U(\overline{X})}1^{(k-1)-l(U(X))}1 & \text{if } X \in \bigcup_{i=k-t}^{k} \mathcal{S}_i^k \end{cases}$$

Balanced code construction 1: Given $k, r \in \mathbb{N}$ $k \ge 6$, $k \le 2^{r+1} - 2$ and $t = t(k) \stackrel{\text{def}}{=} \lfloor \frac{k}{4} \rfloor$, the encoding scheme is:

1. Encode the information words $X \in \left(\bigcup_{i=0}^{t} \mathcal{S}_i^k\right) \cup \left(\bigcup_{i=k-t}^{k} \mathcal{S}_i^k\right)$ using the tail-map defined above
2. Encode the other information words using single maps defined by the Knuth's complementation method.

Two other improved tail-maps and balanced codes based on these maps are also designed in the paper. In particular Balanced codes with r check bits, $k \le 3 \cdot 2^r - 8$ and $k \le 5 \cdot 2^r - 10r + Constant$ ($Constant \in \{-15, -10, -5, 0, +5\}$) information bits are designed. In the first two cases the Tail-maps can be computed with a parallel scheme.

[0] This work is supported by the grant from National Science Foundation MIP-9016143. The first author's work is supported by the Italian National Research Council (CNR 203.01.59).

References

[1] S. Al-Bassam and B. Bose, *Design of Efficient Balanced Codes*, in Proc. IEEE 19th Int. Symp. Fault Tolerant Computing, June 1989.

[2] S. Al-Bassam and B. Bose, *On Balanced Codes*, IEEE Trans. Inform. Theory, vol. 36, pp. 406-408, March 1990.

[3] B. Bose, *On Unordered Codes*, IEEE Trans. on Computers, vol. 40, pp. 125-131, Feb. 1991.

[4] D. E. Knuth, *Efficient Balanced Codes*, IEEE Trans. Inform. Theory, vol. IT-32, pp. 51-53, Jan. 1986.

[5] G. Longo, *Teoria dell'Informazione*, Serie di Informatica, Boringhieri, 1980.

[6] E. Sperner, *Ein Satz über Untermenge einer endlichen Menge*, Math. Zhais., vol. 27, pp. 544-548, 1928.

IRREDUCIBLE COMPONENTS OF CANONICAL DIAGRAMS FOR SPECTRAL NULLS

Hiroshi Kamabe
Department of Electrical and Electronic Engineering, Mie University
1515 Kamihama-cho, Tsu-shi 514 Japan.
e-mail: kamabe@elecom.mie-u.ac.jp

Abstract

Irreducible components of canonical diagrams for spectral null constraints at $f = f_S k/n$ are studied, where k and n are relatively prime integers with $0 \leq k < n$ and f_S is the symbol frequency.

To identify systematically all irreducible components of the canonical diagrams for first-order spectral nulls at f, we give a set of channel symbol sequences specifying all of them. If n is a prime number, then each sequence in the set corresponds to exactly one irreducible component up to label-preserving graph isomorphism. We also give a set of channel symbol sequences specifying all irreducible components of canonical diagrams for second-order spectral nulls at dc (i.e.,$f = 0$).

Introduction

A spectral null constraint requires that channel symbol sequences should have no frequency content at a specified frequency f. Codes for the constraint were characterized in terms of finite directed graphs with labeled edges. Marcus and Siegel [1], however, have defined canonical diagrams for first-order spectral null constraints at f, which are countable-state directed graphs with labeled edges, and they also have shown that every finite subdiagram of a canonical diagram has a first-order spectral null at f. Recently, order-K spectral null constraints have been introduced as extensions of first-order spectral null constraints and canonical diagrams for them also have been given. We can construct a spectral null code by choosing an irreducible finite subdiagram from the canonical diagram and by applying the code construction schemes given in [4], [2], [3].

When we design a code in such a way, a choice of an irreducible finite subdiagram from the canonical diagram has an effect on the code we get. Hence, we must be able to identify all irreducible finite subdiagrams of the canonical diagram in order to obtain an optimal code in some sense. However, canonical diagrams have infinitely many states and configurations of them are not so simple that we can understand them intuitively. Moreover, a canonical diagram consists of disjoint irreducible countable-state subdiagrams. Therefore, in this paper we investigate irreducible components of a canonical diagram and give a systematic way to identify all of them.

We assume that the channel symbol alphabet is $\{1, -1\}$.

Irreducible Components for First-Order Spectral Null at f

Let p be an integer with $0 \leq p \leq n-1$. Let G_p^f be a countable-state-transition-diagram(CSTD) which is period-p canonical for a spectral null at f [1]. We assume that $f > 0$ (i.e., $n \geq 2$) because it is trivial that G_0^0 is irreducible. In the case where n is prime, we have identified all irreducible components of G_p^f.

Let $E = \{1, -1, 1\cdot -1, 1\cdot -1\cdot -1, \cdots, 1\cdot \underbrace{-1\cdot -1\cdots -1}_{n-2 \text{ times}}\}$.

Let $\tilde{G}_p^f$ be an irreducible component of G_p^f which contains the state 0. For $a \in E$, let $I_p^f(a)$ be the irreducible component of G_p^f such that a is generated by a cycle in $I_p^f(a)$. For two CSTD's I and I', if there is a label-preserving graph isomorphism of I to I' then we write $I \cong I'$.

Theorem 1 G_p^f is the union of $I_p^f(a)$, $a \in E$ and $\tilde{G}_p^f$.

Theorem 2 Assume that n is prime. Then

- for every pair of sequences $a, b \in E$ with $a \neq b$ we have $I_p^f(a) \ncong I_p^f(b)$;
- for every irreducible component I of G_p^f, we have $I \cong \tilde{G}_p^f$ or $I = I_p^f(a)$ for some $a \in E$.

For every $a = a_0 \cdots a_{L-1} \in E$, $I_p^f(a)$ contains the state $-\sum_{i=0}^{L-1} \exp(-2\pi i\sqrt{-1}k/n)a_i$. Therefore we can generate all irreducible components of G_p^f by applying the state transition rule for G_p^f recursively.

Irreducible Components for Second-Order Spectral Null at DC

We have identified all irreducible components of canonical diagrams for second-order spectral nulls at dc.

Let p be a positive integer. Let $G_p^{(2)}$ be a CSTD which is period-p canonical for a second-order spectral null at dc [5]. The set of states in $G_p^{(2)}$ is $\mathrm{Z} \times \mathrm{Z}$, where Z is the set of integers. Let σ be a state in $G_p^{(2)}$. Define a diagram $L_p(\sigma)$ to be a subdiagram of $G_p^{(2)}$ which consists of all states τ (and all edges connected to those states) such that there are paths from τ to σ or paths from σ to τ. Then

Proposition 1 For every state σ in $G_p^{(2)}$ $L_p(\sigma)$ is irreducible.

Theorem 3 For every irreducible componet I of $G_p^{(2)}$, we hvave $I \cong L_p((i,0))$ for some i with $0 \leq i \leq p-1$.

Thus we can generate all irreducible components of $G_p^{(2)}$ by applying the state transition rule of $G_p^{(2)}$ recursively to states $(0,0),(1,0),\ldots,(p-1,0)$.

References

[1] B. Marcus and P. Siegel, "On codes with spectral nulls at rational submultiples of the symbol frequency," *IEEE Trans. Inform. Theory.*, vol. IT-33, pp. 557–568, 1987.

[2] B. Marcus, "Sofic systems and encoding data," *IEEE Trans. Inform. Theory.*, vol. IT-31, pp. 366–377, 1985.

[3] R. Karabed and B. Marcus, "Sliding-block coding for input-restricted channels," *IEEE Trans. Inform. Theory.*, vol. IT-34, pp. 2–26, 1988.

[4] R. Adler, D. Coppersmith and M. Hassner, "Algorithms for sliding block codes," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 5–22, 1983.

[5] R. Karabed and P. Siegel, "Matched spectral null codes for partial response channels," *IEEE Trans. IT.*, vol. IT-37, pp. 818–855, 1991.

Conservative Codes.

S. Al-Bassam
Computer Science Departement,
King Fahd University of
Petroleum and Minerals,
Dhahran, Saudi Arabia 31261.

B. Bose
Department of Computer
Science,
Oregon State University,
Corvallis, OR 97331,
bose@cs.orst.edu

In a conservative code of dimension n, every word has $\lfloor n/2 \rfloor$ transitions. A transition occurs when two adjacent bits are the complements of each other. A $(1 \rightarrow 0)$ and $(0 \rightarrow 1)$ transitions are treated indistinguishably.

Conservative codes can be used for bit synchronization in a high-speed communication channels [1]. Non-systematic conservative codes were introduced and analyzed in [1]. This paper gives efficient constructions of conservative codes which have the same efficiency as the balanced codes reported in [3].

Notation:

$W(\mathbf{x})$: the Hamming weight of the binary vector $\mathbf{x}$; $W(\mathbf{x}) = W(x_1 \cdots x_n) = \sum_{i=1}^{n} x_i$.

$NT(\mathbf{x})$: the number of transitions in $\mathbf{x}$, $NT(\mathbf{x}) = \sum_{i=1}^{n-1} x_i \oplus x_{i+1}$.

$N(n,t)$: the number of n-bit vectors with t transitions, i.e. $N(n,t) = |\{\mathbf{x} \in \{0,1\}^n \mid NT(\mathbf{x}) = t\}|$.

For instance, if $\mathbf{x} = 1101$, then $W(\mathbf{x}) = 3$ and $NT(\mathbf{x}) = 2$. To find $N(4,2)$ we see that only the words 0010, 0100, 0110, 1001, 1011, and 1101 have 2 transitions; therefore, $N(4,2) = 6$.

Definition 1: Let $\mathbf{x} = x_1 \cdots x_n$ be a binary vector, then let $\mathbf{x}^{[j]}$ denote $\mathbf{x}$ where the first j even bits are complemented; i.e. $\mathbf{x}^{[j]} = x_1 \bar{x}_2 x_3 \bar{x}_4 \cdots x_{2j-1} \bar{x}_{2j} x_{2j+1} \cdots x_n$. For instance, if $\mathbf{x} = 1000\,1101$ then $\mathbf{x}^{[3]} = 1101\,1001$. □

Properties of $\mathbf{x}^{[j]}$:

1. $NT(\mathbf{x}^{[j]}) = NT(\mathbf{x}^{[j-1]}) + i$ for $1 \leq j < \lceil n/2 \rceil$, where $i = 0, 2$, or -2 and $NT(\mathbf{x}^{[\lfloor n/2 \rfloor]}) = n - 1 - NT(\mathbf{x})$
2. For any integer t, of the same parity as $NT(\mathbf{x})$ (i.e. $t \equiv NT(\mathbf{x}) \bmod 2$), where $\min(NT(\mathbf{x}), n-1-NT(\mathbf{x})) \leq t \leq \max(NT(\mathbf{x}), n-1-NT(\mathbf{x}))$, there exist a j such that $NT(\mathbf{x}^{[j]}) = t$ (where $0 \leq j < \lceil n/2 \rceil$).

Given that $NT(\mathbf{x}^{[\lfloor n/2 \rfloor]}) = n-1-NT(\mathbf{x})$ and that $NT(\mathbf{x}^{[j]}) = NT(\mathbf{x}^{[j-1]}) + i$ where $i = 0, 2$, or -2, we see that every integer of the same parity as $NT(\mathbf{x})$ in the range $\min(NT(\mathbf{x}), n-1-NT(\mathbf{x})) \leq t \leq \max(NT(\mathbf{x}), n-1-NT(\mathbf{x}))$ is obtained after the complementation of some even bits. □

Code Design: Before starting the construction, one more definition is required.

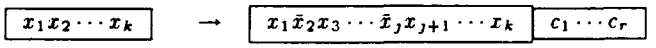
Definition 2: A complement check (c-check), C consists of two r-bit vectors and positive or negative sign as follows: $C \triangleq \{\mathbf{c}_0, \mathbf{c}_1\}^{+/-}$ where $\mathbf{c}_0 = \bar{\mathbf{c}}_1$ and the left most bit of $\mathbf{c}_0$ is 0 (and the left most bit of $\mathbf{c}_1$ is 1). Furthermore, let

$$NT(C) \triangleq \begin{cases} NT(\mathbf{c}_0) & \text{if C has a negative sign} \\ NT(\mathbf{c}_0) + 1 & \text{if C has a positive sign} \end{cases}$$ □

The set $\{0,1\}^r$ of check symbols are grouped into 2^r c-checks. For example, the c-checks obtained from $\{0,1\}^2$ are $\{00,11\}^-$, $\{00,11\}^+$, $\{01,10\}^-$, $\{01,10\}^+$.

[0]This work is supported by the grant from KFUPM and National Science Foundation MIP-9016143.

Figure 1: Conservative code structure of Construction I.



These c-checks will be used to construct a conservative code. When a c-check is appended to an information vector, it is possible to generate (or not to generate) a transition between the information vector and the c-check depending on whether $\mathbf{c}_0$ or $\mathbf{c}_1$ was appended. A c-check carries a positive (or negative) sign depending on whether it is always used to generate (or not to generate) a transition at the time of encoding.

Let R_t denote the set of information vectors with t transitions; i.e. $R_t = \{\mathbf{x} \in \{0,1\}^k \mid NT(\mathbf{x}) = t\}$.

The encoding and decoding process is captured in the following notation:

$$C :: R_{t_1} \cup R_{t_2} \cup \cdots \cup R_{t_q} \rightarrow R_t.$$

This notation means that the information words with $t_1, t_2, \cdots$, and t_q transitions are transformed to some k-bit vector with t transitions by complementing some even bits. One of the two check symbols in C will then be appended to the right-end of the transformed word to obtain the final code word, as depicted in fig. 1. The vector $\mathbf{c}_0$ is appended if $x_k = 0$ and C has a negative sign or $x_k = 1$ and C has a positive sign; otherwise, $\mathbf{c}_1$ is appended. This allows the creation of one (or zero) transition when C has a positive (or negative) sign. To obtain a conservative code, the final code word must have $\lfloor (k+r)/2 \rfloor$ transitions; therefore, all maps must satisfy

$$NT(C) + t = \lfloor (k+r)/2 \rfloor. \quad (1)$$

In decoding, the check symbol and the presence (or absence) of a transition between the information and the check are used to obtain the original information vector. If $\mathbf{yc}$ is the received code word, then $\mathbf{c}$ and the presence/absence of a transition between $\mathbf{y}$ and $\mathbf{c}$ will uniquely determine the c-check (C) and hence its associated map, say $C :: R_{t_1} \cup \cdots \cup R_{t_q} \rightarrow R_t$. The even bits of $\mathbf{y}$ will then be complemented until $NT(\mathbf{y}^{[j]})$ is equal to $t_1, t_2, \cdots$, or t_q.

Based on these concepts, codes with k up to $2^{r+1} - r - 1$, using r check bits are designed.

References

1. Y. Ofek, "The Conservative Code for Bit Synchronization," IEEE Trans. on Communications, vol. 38, July 1990.
2. D. Morris, *Pulse Code Formats for Fiber Optical Data Communication*. NY: Marcel Decker, 1983.
3. D. E. Knuth, "Efficient Balanced Codes," IEEE Trans. on Information Theory, vol. 32, Jan. 1986.
4. S. Al-Bassam and B. Bose, "Design of Efficient Balanced Codes," IEEE Trans. on Computers (to appear).
5. S. Al-Bassam and B. Bose, "On Balanced Codes," IEEE Trans. on Information Theory, vol. IT-36, March 1990.

ADAPTIVE OS LOCAL DETECTION FOR DATA FUSION

M. Longo, M. Lops
Dipartimento di Ingegneria Elettronica, Università di Napoli, Via Claudio 21, 80125 Napoli, Italia

Abstract

In a decentralized detection scheme, we consider adaptive Order Statistic (OS) thresholding for local decisions: each local detection threshold is a linear combination of ranked samples from the reference window centered on the spatial cell being probed. Letting $\underline{c}_m$ the vector of coefficients of the linear combination at the m-th node, $m = 1, \dots M$, how should the $\underline{c}_m$'s be chosen in order to maximize the overall power of a given fusion rule for a fixed overall type-I error probability? Assuming exponential observations, we solve this problem for AND and OR fusion rules, and we compare the respective performance.

Problem formulation

Decentralized detection is based on the concept of data fusion: local decisions taken at spatially separated remote sensors are transmitted to a central processor to make a binary decision about the state of the sensed environment. In practical situations, the environment is sequentially scanned in time-space on a cell-by-cell basis, and a decision is made for each cell. Adaptive processing is required to track various changes of the disturbance. Examples are presented in [1,2]: a common model of disturbance is assumed at each sensor, and adaptation is accomplished by estimating a distributional parameter based on a reference set of cells surrounding the cell under test.

Estimation procedures based on Order Statistics (OS) are preferrable in regard to robustness against possible non-homogeneities in the reference set. We assume that the local estimates of the disturbance activity are achieved as linear combinations of OS's, and that each local decision is based on the logical variable

$$h_m = u\left(z_m - \gamma_m \underline{c}_m^T \underline{y}_m\right), \qquad m = 1, \dots, M,$$

where $u(\cdot)$ is the unit-step function, z_m is the observation from the cell being tested available at the $m-$th sensor, $\underline{y}_m$ is a ranked version of the N_m samples from the reference set of the $m-$th sensor, γ_m is a threshold coefficient determining the local type-I error probability, $\underline{c}_m$ is a set of coefficients of size N_m (if censoring is adopted, the last r_m entries of $\underline{c}_m$ are set to zero). The final decision is taken according to a preassigned fusion rule of the h_m's, e.g. AND, OR, which determines the global performance.

Optimisation of the local disturbance estimators is of crucial importance if the inherent detection loss due to the adaptive processing is to be kept at a minimum. This requires a statistical model for the observables.

We assume that the sample from the cell being tested is exponential with parameter σ_{m_1} (alternative hypothesis) or σ_{m_0} (null hypothesis), while the samples in the reference sets are independent exponential variates with parameter σ_{m_0}. The local error probabilities are

$$\alpha_m = \prod_{h=1}^{N_m - r_m} \frac{1}{1+\gamma_m \zeta_{mh}}; \qquad \beta_m = 1 - \prod_{h=1}^{N_m - r_m} \frac{1}{1+\frac{\gamma_m \zeta_{mh}}{1+S_m}}$$

where $\zeta_{mh} = \sum_{l=r_m}^{N_m - h} a_{m N_m - l}/(N_m + 1 - h)$ and where $S_m = \sigma_{m_1}^2/\sigma_{m_0}^2$. The goal is to maximize, with respect to $\underline{c}_m$, the overall power $1-\beta$ for constrained α.

AND fusion rule

The global performance reduces to mere products of local α_m's and β_m's. Lagrangian maximization then yields the optimum $\underline{c}_m$ in terms of products $\gamma_m \zeta_{mh}$, that are solutions to the system of equations

$$\frac{1+\gamma_1\zeta_{11}}{1+S_1+\gamma_1\zeta_{11}} = \frac{1+\gamma_m\zeta_{mh}}{1+S_m+\gamma_m\zeta_{mh}}$$

For constant S_m, say $S_m = S$, the solution is $\gamma_m \zeta_{mh} = A$, say, and the optimum performance is

$$\alpha = (1+A)^{-N_{eq}}; \qquad 1-\beta = \left(1+\frac{A}{1+S}\right)^{-N_{eq}},$$

with $N_{eq} = \sum N_m - r_m$. That is, the same performance of a single detector whose threshold is adapted through a minimum variance estimate based on a reference set of N_{eq} i.i.d. exponential variates.

OR fusion rule

The global performance can still be expressed in simple terms of the α_m's and the β_m's, but the equations for the optimum coefficients can only be solved through numerical techniques. A suboptimal approach leading to analytical solution is to minimize the individual β_m with respect to the ζ_{mh}'s (whence $\zeta_{mh} = 1/(N_m - r_m)$), and subsequently to minimize β with respect to the γ_m's. In particular, assuming $S_m = S$ and $N_m - r_m = L$ leads to

$$\gamma_m = L[(1-(1-\alpha)^{1/M})^{-1/L} - 1] \equiv \gamma,$$

say, whence,

$$\beta = [1-(1+S)^L/(1+S+\gamma/L)^L]^M.$$

References

[1] M. Barkat, P. K. Varshney, "Decentralized CFAR Signal Detection", *IEEE Trans.*, Vol. AES-25, March 1989, pp. 141-149.

[2] A. R. Elias-Fusté, A. Broquetas-Ibars, J. P. Antequera, J. C. Marin Yuste, "CFAR Data Fusion Center with Inhomogeneous Receivers", *IEEE Trans.*, Vol. AES-28, January 1992, pp. 276-284.

ASYMPTOTIC REFINEMENTS IN BAYESIAN DISTRIBUTED DETECTION[1]

Adrian Papamarcou and Po-Ning Chen

Electrical Engineering Department and Institute for Systems Research
University of Maryland, College Park MD 20742

Abstract

The performance of a parallel distributed detection system is investigated as the number of sensors tends to infinity. It is assumed that the i.i.d. sensor data are quantized locally into m-ary messages and transmitted to the fusion center for Bayesian binary hypothesis testing. Large deviations techniques are employed to show that the equivalence of absolutely optimal and best identical-quantizer systems is not limited to error exponents, but extends to the actual Bayes error probabilities up to a multiplicative constant. This is true as long as the two hypotheses are mutually absolutely continuous; no further assumptions, such as boundedness of second moments of the post-quantization log-likelihood ratio, are needed.

Summary

Consider a parallel distributed system consisting of n geographically dispersed sensors, noiseless one-way communication links, and a fusion center. Each sensor makes an observation (denoted by Y_i) of a random source, quantizes Y_i into an m-ary message $U_i = g_i(Y_i)$, and then transmits U_i to the fusion center. Upon receipt of $(U_1, \ldots, U_n)$, the fusion center performs a binary hypothesis test (H_0 against H_1) about the nature of the random source. A Bayesian setup is assumed throughout, and the Bayes error probability is denoted by $\gamma_n(\pi)$, where π is the prior probability of H_0.

It was shown by Tsitsiklis [1] that even when the observations are i.i.d., the optimal m-ary quantizers g_i need not be identical. Thus the absolutely optimal system (*) does not, in general, coincide with the best identical-quantizer system ($^\circ$). Since the latter is much easier to design than the former, it is natural to seek an estimate of the performance loss resulting from using identical quantizers.

Tsitsiklis supplied a result of this type in the i.i.d. case by showing, under a fairly general assumption, that the two systems are asymptotically exponentially equivalent. More precisely, if P and Q are mutually absolutely continuous distributions of the i.i.d. observations under H_0 and H_1 respectively, then

$$\lim_{n\to\infty} -\frac{1}{n}\log\gamma_n^*(\pi) \;=\; \lim_{n\to\infty} -\frac{1}{n}\log\gamma_n^\circ(\pi)$$

(i.e., the two error exponents coincide) provided that the second moments—under P and Q—of the post-quantization log-likelihood ratio $\log[P_g(U)/Q_g(U)]$ are bounded as the quantizer mapping g varies. The optimal error exponent is the supremum (over g) of the Chernoff exponent associated with the m-ary post-quantization distributions P_g and Q_g. It has also been shown that this supremum is achieved by a g^* taken from the class of deterministic likelihood-ratio quantizers; and that such quantizers are optimal in the nonasymptotic (fixed n) setting.

Two questions arose from Tsitsiklis' work:

1. Is the aforementioned boundedness assumption really necessary?
2. Does the nonnegative quantity $\log[\gamma_n^\circ(\pi)/\gamma_n^*(\pi)]$ admit an upper bound tighter than $O(n)$ (implied by the equality of error exponents)?

Tsitsiklis [2] conjectured that the answer to the first question is negative, and in this paper we give proof to his conjecture. As to the second question, we show that the upper bound $O(n)$ on $\log[\gamma_n^\circ(\pi)/\gamma_n^*(\pi)]$ can be tightened to $O(1)$, hence the ratio $\gamma_n^\circ(\pi)/\gamma_n^*(\pi)$ is bounded from above (trivially, it is also lower-bounded by unity). We therefore have:

Theorem *If P and Q are mutually absolutely continuous, then for all $\pi \in (0,1)$,*

$$\limsup_{n\to\infty} \frac{\gamma_n^\circ(\pi)}{\gamma_n^*(\pi)} \;<\; \infty\,. \qquad \square$$

We employ large deviations techniques for proving this theorem. Using a refinement (due to Esseen) of the central limit theorem for independent but not identically distributed summands, we show the following: if all quantizers in the optimal system are "regular," in that they yield—at their output—log-likelihood ratios that satisfy certain uniform boundedness constraints, then the Bayes error probability $\gamma_n^*(\pi)$ can be lower-bounded by $cn^{-1/2}\exp\{-\rho_m n\}$, where ρ_m is the optimal error exponent. The same expression—only with a larger value of c—is also an upper bound on $\gamma_n^\circ(\pi)$, so the conclusion that $\gamma_n^\circ(\pi)/\gamma_n^*(\pi)$ is bounded from above is close at hand. It remains to show that the number of "irregular" quantizers in the optimal system is bounded. As it turns out, these quantizers yield an error exponent smaller than ρ_m, and thus heuristically, they can only exist in small numbers. We give rigorous proof to this fact using a technical argument based on conditioning.

Our simulations of Bayesian distributed detection have shown that the ratio $\gamma_n^\circ(\pi)/\gamma_n^*(\pi)$ is in many instances close to unity. It is quite possible that under conditions as yet unknown to us, the ratio $\gamma_n^\circ(\pi)/\gamma_n^*(\pi)$ tends to unity as n approaches infinity. This, however, is not true in general, and we give a counterexample in which the ratio $\gamma_n^\circ(\pi)/\gamma_n^*(\pi)$ is greater than $r > 1$ infinitely often in n.

References

[1] J. N. Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals and Systems*, 1(2): 167–182, 1988.

[2] J. N. Tsitsiklis. On threshold rules in decentralized detection. In *Proc. 25th Conference on Decision and Control*, pages 232–236, Athens, Greece, 1986.

[1] Research supported by the Institute for Systems Research (a National Science Foundation Engineering Research Center) at the University of Maryland, College Park.

DISTRIBUTED CELL-AVERAGING CFAR DETECTION OF DEPENDENT SIGNAL RETURNS *

Rick S. Blum
Electrical Engineering and Computer Science Department
Lehigh University, Bethlehem, PA 18015

Saleem A. Kassam
Department of Electrical Engineering
University of Pennsylvania, Philadelphia, PA 19104

Abstract

Constant false alarm rate (CFAR) detection techniques have been of much interest in radar and sonar applications. In this paper we consider CFAR detection in a decentralized, multisensor context and show some interesting characteristics of cell averaging (CA) in this setting. We investigate cases with observations which are dependent from sensor to sensor, for which results have been lacking. The in-phase and quadrature components of the received narrowband observation at each sensor consist of a weak random signal in additive clutter and noise. Each sensor transmits a single binary decision to a fusion center which uses either an AND or an OR fusion rule to develop a final decision. The forms of the best sensor detector rules are found and a set of necessary conditions are given for their thresholds. Solutions are obtained for some representative cases and their detection probability performance is studied. Their ability to maintain constant false alarm probability in the face of clutter edges is also studied. We show that solutions which use different fusion rules may excel for each of these different criteria.

I. Introduction

Previous studies of cell-averaging constant false alarm rate (CA-CFAR) distributed detection techniques have focused on cases with independent observations [1, 2]. Here we investigate some cases with dependent observations from sensor to sensor. For our model of the dependency between the narrowband returns received at the remotely located sensors, we initially assume the in-phase and quadrature signal components of the returns have a jointly Gaussian probability density function (pdf). This appears to be a reasonable extension to the common radar return model of Swerling type I target fluctuations often used for the single sensor case. The narrowband signal components are observed in the presence of additive noise and clutter with the combined noise and clutter observations initially assumed to be Gaussian distributed and independent from sensor to sensor. The power of the combined noise and clutter observations at each sensor is unknown and so a CA-CFAR scheme is employed at each sensor.

Due to the typical structure of radar and sonar receivers, our decisions will be based on processing the envelope of the observed returns. Our signal detection schemes will be optimized for the case of weak signals so we use locally optimum detection techniques [3]. While we provide some results for cases with N sensors, due to the complexity of the problem we focus on the two-sensor case with sensor decisions based on a single observation, augmented with N_j reference samples taken at each sensor $j = 1, 2$. The two sensors are each constrained to transmit only a single binary decision to a fusion center and we consider only non-randomized fusion rules; specifically, the AND and the OR fusion rules.

*This material is based upon work supported by the National Science Foundation under Grant No. MIP-9211298 and by the Air Force Office of Scientific Research under Grant 90-0050

II. Summary of Results

The necessary conditions for the locally optimum (LO) sensor detector thresholds for either an AND or an OR fusion rule have been obtained for N-sensor cases. Example solutions have been found for some representative two-sensor cases. The LO AND rule solution was found to always use equal thresholds at each sensor detector. The LO OR rule solution used equal thresholds if the number of reference cells at each of the sensors was the same, but different thresholds if the number of reference cells at the two sensors was different. We found that the AND rule solution was better than the OR rule solution for a range of small false alarm probabilities and that this range tended to shrink as the number of reference samples used at each sensor was increased.

We have also provided some analysis and results on the ability of our two-sensor distributed CFAR detection systems to maintain constant false alarm probability in the presence of clutter edges across their groups of reference samples. We considered the LO schemes using the AND and OR fusion rules for a particular false alarm probability. We found that the OR rule scheme was superior to the AND rule scheme if both sensors had the same clutter edge applied to them. The detection probability of one of our two-sensor LO distributed CA-CFAR detection schemes was shown to be larger than that for a single sensor CA-CFAR scheme for cases with a wide range of signal-to-noise ratios and dependent observations from sensor to sensor.

Cases with non-Gaussian signals, noise, and clutter were also considered for the particular case where the pdf of the combined noise and clutter contains an unknown scale parameter. We suggest schemes which use sensor tests which are LO [3] among all tests satisfying a specific constraint which insures a CFAR test. This technique generates tests which can be considered a generalization of the CA-CFAR scheme for non-Gaussian combined noise and clutter pdfs since the CA-CFAR scheme is generated by assuming a Gaussian pdf for the combined noise and clutter observations. The form of the CFAR sensor test statistics were found for a specific example with a Cauchy noise pdf. The weak-signal performance of our scheme was shown to be better than that of a CA-CFAR scheme for a single sensor case with Cauchy noise.

References

[1] M. Barkat and P. K. Varshney, "Decentralized CFAR Signal Detection," *IEEE Transactions on Aerospace and Electronic Systems*, AES-2, vol. 25, pp. 141-148, March 1989.

[2] M. Barkat and P. K. Varshney, "Adaptive Cell-Averaging CFAR Detection in Distributed Sensor Networks," *IEEE Transactions on Aerospace and Electronic Systems*, AES-27, No. 3, pp. 424-429, May 1991.

[3] R. S. Blum and S. A. Kassam, "Locally optimum distributed detection with dependent sensors," *IEEE Transactions on Information Theory*, IT-38, pp. 1066-1079, May 1992.

Integration of Complementary Detection-Localization Systems

T. T. Kadota
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, New Jersey 07974

Consider two surveilance systems: one using the visible-band satellite (called the primary) and the other using the infrared (called the secondary). The primary system, because of its shorter wavelength, gives more acurate and detailed results than the secondary. Yet, it is more susceptible to atmospheric interference (e.g., clouds). This complementary aspect suggests integration of the two systems. Performance improvement by the integration depends on how they are combined and, even if the improvement is substantial, cost of the integration may not justify it. Thus, it is desirable to have some rational methods of integration and to evaluate the merit of such integration (i.e., performance improvement vs. cost). By using a simple example, we illustrate how this might be done.

The primary system consists of two-dimensional sampled data, say, an $n \times n$ data array, and a signal processing algorithm to detect and localize a target. The data contain a target signal, if present, the usual white background noise and localized random disturbances representing cloud coverage. In the absence of clouds, the optimum processor is the matched filter and the detection and localization is done by maximizing the matched filter output with respect to the possible target location and thresholding it. When random clouds appear, they overshadow the target and severly reduce the effective S/N. With the use of a Poisson-distributed cloud model, we derive the log-likelihood ratio which consists of the matched filter and a "cloud remover". The latter involves a linear combination of exponentials in the data and requires far more computation than the former. However, it significantly reduces the effect of the cloud. The secondary system consists of a similar two-dimensional data covering the same physical area but with coarser sampling, say, an $m \times m$ data array with $m < n$. Unlike the primary data, the random disturbances are assumed negligible. Hence, the optimum processor comprises of a matched filter only.

The simplest integration of the two systems is the "decision-decision" integration where the decisions of the indivisual systems are combined to make a joint decision. If both systems agree on the target presence and the location, this integration is satisfactory and the cost is minimal. If they disagree, some hierachical order must be established to arrive at a joint decision. In the event of a weak target, if both decide that there is no target inspite of its presence, the joint decision will be "no target" and the target will be missed. The most thorough integration is the "data-data" integration where the secondary data are combined with the primary and they are jointly processed to produce a single decision. With the optimum processing this integration gives the highest performance improvement, though the cost of integration is also the highest since the entire data, instead of the decisions only, are combined and processed. Between these two, there is an "data-decision" integration where the decision of the secondary system is combined with the data of the primary and they are jointly processed to produce the final decision. This represents a compromise between the performance improvement and the additional cost. Such an integration is advantagenous when the secondary system has high S/N. Then the detection is virtually done by the secondary and its localization result specifies a small region in the primary data where the target is likely to be found. On the other hand, if the secondary S/N is low, the decision by the secondary may mislead the primary. Thus we need a criterion for deciding when and how to integrate the two systems.

For measuring the performance improvement by integration, we introduce "db-gains" in detection and localization (or resolution), which is the gain in the equivalent S/N measured in db. The equivalent S/N for detection is defined as the S/N required by the matched filter to achieve the same detection probability (P_d) in the absence of clouds given the same false-alarm probability (P_f). The equivalent S/N for resolution is defined as follows:Ideally, P_d at the target location should be 1.0 and it should drop to the preassigned P_f value everywhere else. Calculate the average slope of this peak ($P_d = 1.0$) with respect to everywhere else ($P_d = P_f$) and use this as a reference. Define the global resolution as a ratio of the average slope calculated from the actual values of P_d at all locations to this ideal average slope. There are two types of cost incurred by integration: the one which increases with the size of data and the other which is constant, such as fixed overhead cost. We divide the first into the cost of additional storage space and the cost of additional computation load. As unit-free relative figures, we choose the ratio of the cost with the integration to the cost without the integration on all three elements: storage space, computation load and overhead. Then we average these three ratios with appropriate weighting. For example, in the decision-decision integration the additional space is for the information on the location of a target obtained by the secondary system and the additioanl computation load is for combining this information with a similar one obtained by the primary. This may be equivalent to the space and time needed for handling one data point. In the data-data integration, however, the additional space is for the entire secondary data which are $m \times m = m^2$ data points. Thus, the space ratio is $(n^2 + m^2 + n^2)/(n^2 + m^2)$. On the other hand, the computation load is not increased since the matched filtering on the secondary data is done anyway without the integration. In the case of the data-decision integration, the space ratio is about the same as in the case of the decision-decision integration, but the time ratio actually decreases by the factor $[(n/m)^2 + m^2]/(n^2 + m^2)$ since only the section of the primary data corresponding to the declared target location in the secondary data is processed.

As the figure of merit which evaluates the relative merit of integration, we propose the ratio of the db-gain to the average cost for both detection and resolution. In the paper, this ratio is numerically evaluated (via Monte Carlo simulation) for six detection-localization algorithms utilizing the data-data and the data-decision integrations for various values of parameters (e.g., occurrence probability of cloud, attenuation constant of cloud and reflected-light intensity from cloud).

On the Relationship Between Suboptimal Detectors and Measures of Discrimination

Geoffrey C. Orsak[1] *and Bernd-Peter Paris*[2]
Department of Electrical and Computer Engineering
Center of Excellence in C3I
George Mason University
Fairfax, VA 22030-4444

Summary. There are many instances in detector design where one can not implement the minimum probability of error detector for deciding between measures P_0 and P_1. As such, there has been increasing interest in the development of design techniques for determining "good" suboptimal detection strategies.

One effective approach to designing detection strategies has been the optimization of a statistical distance measure between competing hypotheses. The advantage of this approach over minimizing the total probability of error is that these distance measures can be computed while for most problems, the minimum error rate is analytically intractable. Unfortunately, system parameters derived in this manner are not necessarily optimal in the minimum probability of error sense. To address this issue, we develop sufficient conditions under which solutions obtained by optimizing arbitrary distance measures results in the minimum probability of error detector over a chosen class of detectors.

We restrict our attention to the general class of measures of discrimination between probability measures P_0 and P_1 known as f-divergences [3] or Ali-Silvey distances [1]. Mathematically, these distance measures are given by

$$d(P_0, P_1) = h\left(E_0\left[C\left(\frac{dP_1}{dP_0}\right)\right]\right)$$

where E_0 indicates that the expectation is taken with respect to P_0 and where $C(\cdot)$ is a convex real function and $h(\cdot)$ is an increasing real function of a real variable. It is well established that many well known measures of discrimination including the J-divergence, the Battacharyya distance, the Kullback-Leibler distance, and Kolmogorov's measure of variational distance are elements of this class.

Relationships between the *minimum probability of error* for deciding between P_0 and P_1 and various f-divergences have been studied at great length [1, 2, 4]. Most of this work has focused on developing bounds for the minimum probability of error in terms of several f-divergences. Other work has centered on utilizing these relationships to optimize communication systems with respect to specific distance measures rather than the less analytically tractable probability of error [5, 6]. However, there has been no work to our knowledge on studying the relationship between the probability of error for suboptimal detection strategies and f-divergences.

While it is well known that the minimum probability of error between P_0 and P_1 is a f-divergence between these measures, we show that the probability of error derived from *any* decision strategy not equivalent to the likelihood ratio test is not equivalent to a f-divergence (i.e., not a measure of discrimination between P_0 and P_1). This implies that designing suboptimum detection strategies by maximizing the distance between the input statistics may not only be inappropriate but may also lead to inconsistent solutions.

This result seems to suggest that f-divergences have limited applicability to the design of suboptimal detectors. However, we demonstrate a direct linkage in the performance of suboptimal detectors and f-divergences through a form of the data processing theorem. Specifically, we show that the loss in performance of any suboptimal detector over the minimum probability of error detector is bounded below by the loss of "information" across the detector as determined by a specific f-divergence. As such, when this lower bound holds with equality, minimizing the loss of information with respect to this specific f-divergence results in the minimum probability of error solution.

Unfortunately, the f-divergence in the lower bound is just as difficult to compute as the probability of error. Thus, we develop sufficient conditions under which all f-divergences have a common extremum over a class of probability measures. The significance of this result to the problem of detector design is that one may maximize the most analytically tractable f-divergence between the statistics of the detector to minimize the lower bound on the performance loss of the detector. Most importantly, when this lower bound holds with equality, the resulting solution minimizes the probability of error over the class of allowable detectors. To demonstrate the applicability of this theory, we determine the optimal linear detector in the presence of a specific non-Gaussian noise and show in fact that the matched filter is not always the "best" linear receiver. In a second example, we apply this theory to the problem of signal design and determine the optimal signal set in the presence of a specific non-Gaussian noise.

[1]Supported in part by the National Science Foundation under Grant NCR-9109858 and by Rome Laboratories under contract F30602-92-C-0053.

[2]Supported in part by Rome Laboratories under contract F30602-92-C-0053.

References

[1] S. M. Ali and D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another," *J. Royal Stat Soc.*, vol. 28, pp. 131–142, 1966.

[2] M. Basseville, "Distance Measures for Signal Processing and Pattern Recognition," *Signal Processing*, vol. 18, pp. 349–369, 1989.

[3] I. Csiszar, "Information-Type Measures of Difference of Probability Distributions and Indirect Observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299–318, 1967.

[4] M. E. Hellman and J. Raviv, "Probability of Error, Equivocation, and the Chernoff Bound," *IEEE Trans. Inform. Theory*, vol. IT-16, no. 4, pp. 368–372, July 1970.

[5] D. H. Johnson and G. C. Orsak, "Relation of Signal Set to the Performance of Optimal Non-Gaussian Detectors," *to appear IEEE Trans. Commun.*, 1992.

[6] H. V. Poor, "Robust quantization of ϵ-contaminated data," *IEEE Trans. Commun.*, vol. COM-33, pp. 218–222, 1985.

ASYMPTOTIC EXPANSIONS FOR SAMPLE SIZE IN SIGNAL DETECTION

Marat V. Burnashev
Institute for Problems of Information Transmission
Ermolovoy str. 19
Moscow 101447 RUSSIA

H. Vincent Poor
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544 USA

ABSTRACT

The number of samples required for signal detection is considered as a function of the error probabilities. This problem is treated in the context of detecting a constant signal in additive, independent and identically distributed noise. Detectors that base their decisions on the comparison with a threshold of accumulated, nonlinearly transformed observations are treated. Asymptotic expressions are derived for the relationship between sample size and error probabilities for this model in two situations: that in which the nonlinearity has a partially absolutely continuous output distribution; and that in which it has a lattice output distribution. Traditional analyses of such problems have involved only the lowest-order terms of such relationships (i.e., central limit results), leading to performance indices such as the Pitman asymptotic relative efficiency (ARE). Such indices are known to be of limited accuracy in predicting performance for more moderate sample sizes. Here, the behavior of sample size as a function of error probabilities is considered in more detail, leading to more accurate indices of relative efficiencies for such detection problems. Several specific examples are examined in detail, and numerical results are included to illustrate the significantly improved performance estimation afforded thereby for even small sample sizes.

Introduction and Overview

In this paper, we consider the following pair of statistical hypotheses concerning a set $x_1, \ldots, x_n$ of random observations:

$$H_0 : x_k = \xi_k, \quad k = 1, \ldots, n$$

versus

$$H_1 : x_k = \theta + \xi_k, \quad k = 1, \ldots, n$$

where $\{\xi_k\}$ is a sequence of independent and identically distributed (i.i.d.) random variables (r.v.'s) with marginal probability density function f. In order to test between these hypotheses, we consider threshold tests based on statistics of the form

$$T^{(n)}(g, x_1, \ldots, x_n) = \sum_{k=1}^{n} g(x_k)$$

where g is a measurable real-valued function.

A traditional way of comparing two detectors that operate in this way is to consider the relative sample sizes they require to achieve the same performance in terms of the false-alarm and miss probabilities. These required sample sizes are usually estimated through the use of the central limit theorem (CLT) in describing the behavior of the test statistics $T^{(n)}$. Such comparisons are conventionally made in terms of the asymptotic value of the ratio between required sample sizes, in the limits as θ approaches zero at an appropriate rate (see, e.g., Poor [1]). With fixed error probabilities this limit forces the sample size to infinity, and the corresponding limiting ratio is the (Pitman) asymptotic relative efficiency (ARE).

There are several practically interesting noise densities f (Gaussian, Laplacian, sech) and detection functions g (linear, signum, deadzone) for which it is possible to calculate exact values of necessary sample sizes. Studies [2-6] of such cases have shown that the ARE gives a quite good approximation (within let us say 10%) to the actual relative efficiency (RE) of two detectors, when the sample size is rather large ($n \geq 10^3$). For moderate sample sizes (say, $n = 20 - 100$) the ARE is much less accurate (within 50%), but it is still reasonable.

In order to better approximate RE for moderate sample sizes it is quite natural to consider refinements of the CLT using asymptotic expansions. The applicability of such asymptotic expansions in related problems was shown in pioneering work of Cramer (see, e.g., [7]). However, this approach has not been developed in the context of signal detection, although alternative intermediate estimates for RE (without asymptotic expansions in the CLT) have been considered in [4, 8].

The contribution of this paper is to develop, for a given error probabilities α and β, detector function g and signal strength θ, asymptotic expansions for a necessary sample size $n(\alpha, \beta, g, \theta)$, as $\theta \to 0^+$, through the use of asymptotic expansions in the CLT, and to explore the accuracy of approximations based on these expansions. The presentation of these developments is organized as follows. First, we provide a brief review of relevant results on the asymptotic expansion of distributions of sums of i.i.d. random variables. Then, we develop the desired expansions for sample size for two basic cases: that in which the distribution of $g(x_k)$ is of "density" type (meaning that its characteristic function converges to a value less than unity with increasing argument), and that in which the distribution of $g(x_k)$ is of lattice type. Finally, we consider several specific examples and illustrate the accuracy of the developed expansions numerically.

*This research was supported by the U. S. National Science Foundation under Grant NCR-90-02767.

References

[1] H.V. Poor, *An Introduction to Signal Detection and Estimation*, New York: Springer-Verlag, (1988).

[2] J.H. Miller and J.B. Thomas, "Numerical results on the convergence of relative efficiencies", *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 11, pp. 204-209, (March 1975).

[3] R.J. Marks, G.L. Wise, D.G. Haldeman, and J.L. Whited, "Detection in Laplace noise," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 14, pp. 866-872, (November 1978).

[4] D.L. Michalsky, G.L. Wise and H.V. Poor (1982), "A relative efficiency study of some popular detectors," *J. Franklin Inst.* , Vol. 313, pp. 135-148, (1982).

[5] M.I. Dadi and R.J. Marks, "Detector relative efficiencies in the presence of Laplace noise," *IEEE Trans. Aerosp. Electron. Syst.*, Vol. 23, pp. 568-582, (July 1987).

[6] C.W. Helstrom, "Detectability of signals in Laplace noise," *IEEE Trans. Aerosp. Electron. Syst.* , Vol. 25, pp. 190-196, (March 1989).

[7] V.V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, (1975).

[8] R.S. Blum and S.A. Kassam, "Approximate analysis of the convergence of relative efficiency to ARE for known signal detection," *IEEE Trans. Inform. Theory* , Vol. IT-37, pp. 199-206, (January 1991).

Robust Detection of Weak, Known Signals Using Higher-Order Moments

Kevin R. Kolodziejski
Dept. of Electrical and
Computer Engineering
Northeastern University
Boston, MA 02115

John W. Betz
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730

John G. Proakis
Dept. of Electrical and
Computer Engineering
Northeastern University
Boston, MA 02115

Locally optimal detection of a weak, known signal in independent, identically distributed (i.i.d.) noise with known probability density function involves a nonlinear correlator — a memoryless nonlinearity that depends on the noise density function followed by a correlator [1]. In practice, the noise distribution may not be known precisely; but, several moments of the noise may be known. We determine a robust detector using a limited number of moments that describe the ϵ-contaminated noise.

A unique robust maximin efficacy detector has been found for a nonlinear correlator and an ϵ-contaminated noise model when ϵ and the nominal noise density are completely specified. The least-favorable density for a strongly unimodal nominal density [2] and for a nonstrongly unimodal nominal density [3] were obtained from the efficacy saddle-point property. For both the strongly and nonstrongly unimodal nominal density noise models, the robust nonlinearity depends on the least-favorable noise density in the same way that the locally optimal nonlinearity depends on the assumed known noise density. In contrast, we assume a parametric form for the nominal density of a mixture model and find the parameters that yield maximin efficacy, given a limited number of moments of the noise.

The problem is modeled as deciding between the null hypothesis $\mathbf{X} = \mathbf{N}$ and the alternative hypothesis $\mathbf{X} = \theta\mathbf{s} + \mathbf{N}$ where $\mathbf{X}$ is a n-element random observation vector, $\mathbf{N}$ is a vector of i.i.d. noise random variables with univariate density f, $\mathbf{s}$ is a vector of a known signal with finite, nonzero power, $\theta = \frac{K}{\sqrt{n}}$, for some unknown $K > 0$, and n is the number of samples. Let the set of admissible noise densities be the absolutely continuous, ϵ-contaminated densities $\mathcal{F} = \{f | f(x) = (1-\epsilon)g_\alpha(x) + \epsilon h(x)\}$ where g_α is an even symmetric, nonstrongly unimodal density with unknown parameter vector α, h is any even symmetric density from the convex class defined in [3], and ϵ is the unknown contamination parameter. Let Ψ be the set of nonlinearities with derivatives almost everywhere with respect to Lebesgue measure. The robust maximin nonlinear correlator, in terms of efficacy $\eta(\psi, f)$, is at the saddle-point (ψ_o, f_o) that satisfies

$$\max_{\psi \in \Psi} \eta(\psi, f_o) = \min_{f \in \mathcal{F}} \eta(\psi_o, f) \tag{1}$$

where ψ_o is the robust nonlinearity and f_o is the least favorable density in terms of efficacy. From the saddle-point property in equation (1) and the Cauchy-Schwartz inequality, the robust nonlinearity is the locally optimal nonlinearity $\psi_o = -f_o'/f_o$, and the least favorable density minimizes the Fisher information and is given by [3]

$$f_o(x) = \begin{cases} (1-\epsilon)g_\alpha(a)\exp\left[-k(|x|-a)\right], & a < |x| < b \\ (1-\epsilon)g_\alpha(x), & \text{otherwise} \end{cases} \tag{2}$$

where $k = -g_\alpha'(a)/g_\alpha(a)$, a can be uniquely determined from ϵ, and b is chosen so that f_o is absolutely continuous.

Rather than assuming that the nominal noise density and ϵ are known, we select the parametric form of the nominal density and numerically determine ϵ and α by finding a minimum of the Fisher information of the least favorable density while satisfying the moments of the noise. Our results use the variance and kurtosis of the noise and a normalized, truncated Cauchy nominal density with $\alpha = [\alpha_1, \alpha_2]$, where α_1 is the scale parameter and α_2 the truncation point. Detection performance is examined using Monte Carlo simulations with noise distributions that differ from the Cauchy contaminated noise model. The noise realizations are from normalized densities truncated at $\pm\alpha_2$ with either a Gaussian-Gaussian mixture density, a Johnson-S_u density, or the least favorable density with parameters chosen to satisfy the given variance and kurtosis. The number of samples n comprising the test statistic at the output of the nonlinear correlator is equal to 1000. The input signal-to-noise ratio (SNR) is -25 dB. The output SNR is the simulation performance measure since it is approximately proportional to efficacy for a weak signal [1].

Performance of the robust detector compares favorably with the linear detector, sign detector, and the locally optimal detector of the simulation noise. For very heavy-tailed noise, the robust detector performs as well as the sign detector, and significantly better than the linear detector. For moderately heavy-tailed noise, the robust detector performs better than the sign detector and, in some cases, better than the linear detector. In many cases, the robust detector's performance approaches that of the locally optimal detector derived with knowledge of the complete noise statistics.

References

1. Kassam, S. A., 1988, *Signal Detection in Non-Gaussian Noise*, New York: Springer-Verlag.
2. Huber P. J., March 1964, "Robust Estimation of a Location Parameter," *Ann. Math. Statist.*, Vol. 35, pp. 73–101.
3. Warren D. J. and J. B. Thomas, May 1991, "Asymptotically Robust Detection and Estimation for Very Heavy-Tailed Noise," *IEEE Trans. Inform. Theory*, Vol. IT-37, no. 3, pp. 475–481.

Acknowledgement

This work was supported in part by the National Science Foundation under Grant MIP 9115526 and in part by a MITRE grant to Northeastern University's Research Center for Communications and Digital Signal Processing. John Betz's work was supported by the MITRE Sponsored Research Program.

ROBUST CONTINUOUS-TIME DETECTION OF LINEAR PROCESSES*

P. Srinivasa Rao *Don H. Johnson*

Electrical & Computer Engineering Department
Computer & Information Technology Institute
Rice University
Houston, Texas 77251-1892

ABSTRACT

Linear processes are suitable for modeling random received waveforms in a scattering medium, which represents radar, sonar and multipath communication channels. We address a continuous-time detection problem where both the noise (hypothesis H_0) and signal-plus-noise (H_1) waveforms are modeled as linear processes. Uncertainty in the nominal model is considered in the form of classes of probability distributions induced on a function space by the processes under the two hypotheses. By embedding the linear processes in the larger class of infinitely divisible processes and using an integral representation for the latter class, we identify the pair of distributions that are least favorable for the discrimination of linear processes; an optimal detector designed for these distributions is robust for the uncertainty classes considered.

1. PRELIMINARIES

A *linear process* $y(t)$ is defined by

$$y(t) = \int_0^T f(t,s)\,dx(s), \quad t \in I \triangleq [0,T], \qquad (1)$$

$x(s)$ being an independent increment process.

The characteristic function of $y(\underline{t}) = (y(t_1), \ldots, y(t_n))$ is

$$\ln \Phi_{y(\underline{t})}(\underline{u}) = \int_0^T \int_{\mathbb{R}} (e^{iwv} - iwv - 1)\, G(ds \times dv), \quad (2)$$

where $w = \sum_{k=1}^n u_k f(t_k, s)$ and $G(\cdot)$ is a finite measure given by $E\,x^2(t) = \int_0^t \int_{\mathbb{R}} v^2\, G(ds \times dv)$. Defining a measure $\Lambda_\theta(A) = G((s,v) : (vf(t_1,s), \ldots, vf(t_n,s)) \in A)$ on $\mathbb{R}^\theta$, $\theta = \{t_1, \ldots, t_n\}$, (2) can be written as

$$\ln \Phi_{y(\underline{t})}(\underline{u}) = \int_{\mathbb{R}^\theta} (e^{i\langle \underline{u}, v_\theta\rangle} - i\langle \underline{u}, v_\theta\rangle - 1)\, \Lambda_\theta(dv_\theta), \quad (3)$$

which is a canonical form of an infinitely divisible random vector's characteristic function. Hence, $y(t)$ is an *infinitely divisible process* [1].

A *projective limit measure* $\Lambda(\cdot)$ on $\mathbb{R}^I$ can be constructed from the family $\{\Lambda_\theta(\cdot)\}$ such that $\Lambda(g_{\theta I}^{-1} A) = \Lambda_\theta(A)$, where $g_{\theta I}(\cdot)$ is the projection mapping from $\mathbb{R}^I$ to $\mathbb{R}^\theta$. We assume that $\Lambda(\cdot)$ is restricted to $\mathcal{X} \triangleq L_2(I)$. The projective limit measure of $y(t)$ is $\Lambda(B) = G((s,v) : vf(\cdot,s) \in B)$, $\forall B \in \mathcal{B}(\mathcal{X})$, the Borel sets of $\mathcal{X}$ [1].

Maruyama (cf. [2]) obtains an integral representation for infinitely divisible processes by considering a *Poisson random measure* $\Pi(\cdot)$ on $\mathcal{B}(\mathcal{X})$ with intensity $\Lambda(\cdot)$; For disjoint $B_j \in \mathcal{B}(\mathcal{X})$, $\Pi(B_j)$ are independent Poisson random variables with $E\,\Pi(B_j) = \Lambda(B_j)$ and $\Pi(\cdot, \omega)$ is a measure on $\mathcal{B}(\mathcal{X})$ a.s. We then have

$$y(\cdot) = \int_{\mathcal{X}} z\,[\Pi(dz) - \Lambda(dz)]. \qquad (4)$$

Equation (4) is symbolic and denotes equality in distribution of $\langle y, \psi\rangle$ and $\int_{\mathcal{X}} \langle z, \psi\rangle\,[\Pi(dz) - \Lambda(dz)]$, $\forall \psi \in \mathcal{X}$.

Denote by P_0 and P_1 the probability measures induced by $y(t)$ on $\mathcal{X}$ under H_0 and H_1. Using (4), we find that $P_1 \ll P_0$ and that under H_0,

$$\frac{dP_1}{dP_0}(y) = \exp\left\{-K + \int_{\mathcal{X}} \ln \frac{d\Lambda_1}{d\Lambda_0}(z)\Pi_0(dz)\right\}, \qquad (5)$$

where $K = \Lambda_1(\mathcal{X}) - \Lambda_0(\mathcal{X})$. Note that (5) is only a *representation* formula and is not always computable in terms of $y(t)$. The mapping of $\Pi_0(\cdot)$ into $\mathcal{X}$ defined by the integral in (4) is not one-to-one and hence may not be invertible.

2. ROBUST DETECTION OF LINEAR PROCESSES

We address the minimax robust problem

$$\min_{\phi} \sup_{P_1 \in \mathcal{P}_1} R(P_1, \phi) \text{ subject to } \sup_{P_0 \in \mathcal{P}_0} R(P_0, \phi) \le \alpha, \quad (6)$$

where $R(P_j, \phi), j = 0, 1$ are the expected risks. The classes $\mathcal{P}_j$ correspond to the ϵ-contamination or total variation neighborhoods ($\mathcal{L}_j^\epsilon$ or $\mathcal{L}_j^{tv}$) of nominal measures $\bar{\Lambda}_j$. If a *least favorable* pair of distributions P_j' satisfying $R(P_j', \phi') \ge R(P_j, \phi')\ \forall P_j \in \mathcal{P}_j$, $\forall$ likelihood ratio tests ϕ' between P_0' and P_1' exists, (6) is solved by the likelihood ratio test between P_0' and P_1' with a statistic of the form (5).

Consider the robust discrimination of Poisson random measures on $\mathcal{X}$ with classes of distributions $\mathcal{P}_{\Lambda_j}$ generated by intensity measures in $\mathcal{L}_j$. Suppose that (Λ_0', Λ_1') is the least favorable pair identified by Huber's theory after normalizing $\mathcal{L}_j$ to classes of probability distributions. It follows that the Poisson distributions $P_{\Lambda_0'}$ and $P_{\Lambda_1'}$ corresponding to Λ_0' and Λ_1' are least favorable [3].

The robust detector for linear processes now follows. Using representation (4) of linear processes in terms of Poisson random measures, and the fact the likelihood ratio (5) is identical to that between Poisson random measures, we can show that the probability measures P_0' and P_1' corresponding to Λ_0' and Λ_1' are least favorable [2].

*Supported by ONR Grant N00014-89-J-3152.

REFERENCES

[1] P. L. Brockett. The likelihood ratio detector for non-Gaussian infinitely divisible, and linear stochastic processes. *Ann. Stat.*, 12(2):737-744, 1984.

[2] P. Srinivasa Rao. *Robust Continuous-Time Detection in Linear Process Noise.* PhD thesis, Rice University, Houston, TX, April 1992.

[3] J. S. Sadowsky. On the robust discrimination of Poisson random measures. *IEEE Trans. Info. Theory*, IT-33(3):415-419, 1987.

On the Detection of Gaussian Cyclostationary Random Processes

John W. Betz
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730

Abstract

This paper extends previous work on the likelihood detection of cyclostationary processes in stationary Gaussian noise. In contrast to previous developments, we use a Gaussian cyclostationary signal assumption rather than a weak signal assumption. Under the assumption of completely known statistics for signal and noise, the likelihood ratio detector is derived for two related cases: signal detection and detection of cyclostationarity. The difference between these two cases involves different models for the stationary statistics under the two hypotheses.

Summary

We formulate the detection problem based on the complex n_s-element observation vector $\boldsymbol{W}$, whose elements are samples of a bandlimited process $w(t)$ with the first sample at index $t = 0$, and subsequent samples spaced by d. The noise, $\boldsymbol{Z}$, is a zero-mean complex, stationary Gaussian random vector with Toeplitz, Hermitian, autocovariance matrix F. The signal, $\boldsymbol{X}$, is a zero-mean complex, cyclostationary Gaussian random vector with a finite set of cycle frequencies that are harmonically related with fundamental frequency α. (The extension to polycyclostationarity, with incommensurate fundamental frequencies, is straightforward.) The signal autocovariance matrix is $\mathrm{C} \triangleq \mathrm{C}_0 + \mathrm{C}_*$, where C_0 is the Toeplitz, Hermitian component of C that corresponds to a stationarized version of $\boldsymbol{X}$, and $\mathrm{C}_* \triangleq \sum_{m,m\neq 0} \mathrm{C}_m$, where C_m corresponds to the mth cycle frequency. Each matrix C_m is the Hadamard product of a Toeplitz Hermitian matrix and a Hankel matrix that is periodic on its diagonals. The (ρ,χ) element of C_m is $(\mathrm{C}_m)_{(\rho,\chi)} = c_m((\rho-\chi)d)\exp\{im(\psi+2\pi\alpha(\rho-1)d)\}$, where $c_m(\tau)$ is the mth Fourier series coefficient of the autocovariance function, at lag τ, of the continuous-time signal from which $\boldsymbol{X}$ is derived, and ψ is the cyclic phase: the phase offset of the fundamental cycle frequency relative to the sampling instant for the first element in $\boldsymbol{X}$.

The signal detection problem involves detecting a cyclostationary signal added to stationary noise [1]. The observation vector under the null hypothesis is $\boldsymbol{W} = \boldsymbol{Z}$, while the observation vector under the alternative hypothesis is $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{Z}$. When the signal and noise statistics are completely known, the likelihood ratio test yields a sufficient statistic that is the quadratic form $\Lambda(\boldsymbol{W}) = \boldsymbol{W}^\dagger \mathrm{L} \boldsymbol{W} = \mathrm{trace}\{\mathrm{L}\boldsymbol{W}\boldsymbol{W}^\dagger\}$, where † represents the adjoint, and $\mathrm{L} = \mathrm{F}^{-1}\mathrm{C}(\mathrm{F}+\mathrm{C})^{-1}$. The test involves knowledge of the signal power and the noise power, as well as the cyclic phase of the signal.

When the signal is weak, the sufficient statistic reduces to

$$\Lambda(\boldsymbol{W}) \approx \Lambda_w(\boldsymbol{W}_w) \triangleq \boldsymbol{W}_w^\dagger \boldsymbol{W}_w + \sum_{m,m\neq 0} \mathrm{trace}\{\mathrm{C}_0^{-\frac{1}{2}} \mathrm{C}_m \mathrm{C}_0^{-\frac{1}{2}} \boldsymbol{W}_w \boldsymbol{W}_w^\dagger\}$$

where $\boldsymbol{W}_w$ is the linear time-invariant filtering of the observation vector: $\boldsymbol{W}_w \triangleq \mathrm{C}_0^{\frac{1}{2}} \mathrm{F}^{-1} \boldsymbol{W}$, and $\mathrm{C}_0^{\frac{1}{2}}$ is the positive definite square root of C_0. The test is the sum of an energy detector and the coherent sum of detectors for each cycle frequency [2].

When the signal is strong, the sufficient statistic reduces to $\Lambda(\boldsymbol{W}) \approx \boldsymbol{W}^\dagger \mathrm{F} \boldsymbol{W}$. The linear operator depends only on the statistics of the noise, and does not involve cyclostationary statistics of the signal.

One deviation from the assumption of known signal statistics is when the cyclic phase is unknown but constant over the observation. The likelihood ratio test statistic $\Lambda(\boldsymbol{W}) = \int \Lambda(\boldsymbol{W}|\psi) f_\Psi(\psi) d\psi$ where $\Lambda(\boldsymbol{W}|\psi)$ is conditioned on a fixed cyclic phase, and $f_\Psi(\psi)$ is the probability density function of the cyclic phase. When the cyclic phase is uniformly distributed over $(0, 2\pi)$, the likelihood ratio test does not depend on the C_m, but only on C_0.

The cyclostationary detection problem involves determining whether an observation has periodic statistics when the stationary statistics are the same under the two hypotheses. The observation vector under the null hypothesis is $\boldsymbol{W} = \boldsymbol{Y} + \boldsymbol{Z}$, where $\boldsymbol{Y}$ is zero-mean, Gaussian, and stationary with autocovariance matrix C_0 while the observation vector under the alternative hypothesis is $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{Z}$ where the statistics of $\boldsymbol{X}$ are the same as for the signal detection problem. When the signal and noise statistics are completely known, the likelihood ratio test statistic is a quadratic form with operator $\mathrm{L} = (\mathrm{F}+\mathrm{C}_0)^{-1}\mathrm{C}_*(\mathrm{F}+\mathrm{C}_0+\mathrm{C}_*)^{-1}$. When the noise is much stronger than the signal, the linear operator becomes $\mathrm{L} \approx \mathrm{F}^{-1}\mathrm{C}_*\mathrm{F}^{-1}$, which does not depend on the stationary statistics of the signal.

References

1. Chen, C. K., "Spectral Correlation Characterization of Modulated Signals With Application to Signal Detection and Source Location," Ph.D. Dissertation, University of California, Davis, (1989).
2. Gardner, W. A., and Spooner, C. M., "Signal Interception: Performance Advantages of Cyclic-Feature Detectors," Vol. 40, No. 1, *IEEE Trans. Comm.* (1992).

Acknowledgment

This work was supported by the MITRE Sponsored Research Program.

REDUCED-STATE SEQUENCE DETECTORS ARE NOT SIMPLER THAN THE VITERBI ALGORITHM WITH GOOD CONVOLUTIONAL CODES

J.B. Anderson* and E. Offer**

ECSE Department
Rensselaer Polytechnic Institute
Troy, New York 12180-3590 USA

Inst. Communications Eng.
German Aerospace Research
D-8031 Oberpfaffenhofen, Germany

Reduced-stated sequence detection is a method of reducing the state trellis of channel code to a smaller structure. We show that it does not reduce the complexity of BSC decoding for good convolutional codes.

The reduced-state sequence detector, or RSSD, has been introduced in works of Eyuboglu, Qureshi, Chevillat, Elephtheriou, Aulin and Larsson. Because there exists some confusion over precisely how an RSSD works, we will briefly review the procedure here. First, we define some ground rules for the encoder design. We consider only decoders that never backtrack, as does the stack algorithm for instance; furthermore, they retain a fixed number of survivors at each trellis level, as do the Viterbi and M-algorithms; finally, the decoding is bounded-distance decoding, which means that channel error correction is guaranteed so long as the error sequence is of some size d/2 or less. With these assumptions, the RSSD idea works as follows. Code trellis states at level n are grouped into classes, defined by the condition that no code words leading into states in a class are closer than d to any other code words leading into other states in the class. When the decoder search moves on to the next trellis level, only one survivor leading into each class is kept. This contrasts with the usual state trellis decoding, in which one survivor is kept into each state. RSSD works because if the noisy received word satisfies the bounded-distance condition, the transmitted path has to be among the survivors. The full bounded-distance potential of the code may be obtained if the d parameter is set equal to the code free distance.

RSSD's should not be confused with reduced-search decoders, which search only a small part of the original, large trellis; in an RSSD, the trellis is reduced a priori and all of it is searched.

We give a new algorithm that forms the optimally reduced trellis for a convolutional code and a given d, and we show what happens when the algorithm is applied to good codes. The class-forming algorithm depends on the linear-code symmetries of convolutional codes and falls into three parts. The first two parts act to form the class that contains trellis state 0. In part I, trellis state i is tested to see whether it may be classified with state 0. The procedure reduces to a dynamic program (i.e., the Viterbi algorithm), run on the code trellis until the distance into each state in the trellis reaches a steady state. If the least-weight path into state i in this steady-state condition is heavier than d, then state i may be added to the class containing state 0; otherwise, it may not. Part II tests whether a state that may be classed with state 0 may be combined with other states j, k, ... that already have been classed with state 0. Part III forms the other classes, based on the class that contains state 0.

If the algorithm just described cannot find any state that may be classed with state 0, then by code linearity, no states in the code trellis may be classed with any other states, and the RSSD idea fails to produce a smaller trellis structure than the original state trellis. We have applied the algorithm to a large number of good codes of rates 1/3, 1/2, 2/3 and 3/4, and have found no code that admits of any RSSD trellis reduction. Only when the class-finding algorithm is applied to codes with much poorer free distance, such as QLI codes and feed-forward systematic codes, do non-trivial classes get formed. Then an interesting phenomenon occurs: The number of classes so formed is invariably almost the same as the number of states in the best-free-distance code with free distance equal the parameter in the class formation. In this way, the RSSD idea seems to convert bad codes into good ones, so far as the trellis size needed to attain a d is concerned.

In conclusion, RSSD seems to point to some interesting structural properties of codes, but it does not create a simpler decoder for codes that are already good. It is also of interest to compare the RSSD to the M-algorithm, which obeys the same ground rules. It is easy to see that the M-algorithm is the optimal non-backtracking decoder that keeps a fixed number of survivors. A more subtle proof shows that RSSD cannot keep fewer survivors while attaining the same bounded-distance d parameter; we show this in [1]. Examples can be given that show that the M-algorithm actually retains many fewer survivors for the same d. For rate 1/2 coding, RSSD must retain about $4^{d/2}$ survivors, while the M-algorithm needs only $2.414^{d/2}$. The difference is much more extreme in intersymbol interference problems. These facts make sense when one considers that RSSD makes its trellis reduction a priori, while the M-algorithm and other reduced-search decoders make their reductions after viewing the received channel sequence.

[1] J.B. Anderson and E. Offer, "Reduced-state sequence detection with convolutional codes," in submission, IEEE Trans. Information Theory, August 1992.

This work supported by the Humboldt Foundation, Bonn, Germany.

Mapping the Boundaries Established by State Diagram Connectivity

Oliver Collins
Johns Hopkins University

This talk will analyze how the intrasystem communication between the different parts of a partitioned Viterbi decoder is affected by removing the restriction that states be permanently assigned to particular modules. It is the strains on intrasystem communication that limit the coding gains achievable with VLSI technology. Some of the techniques outlined will also be useful for improving concatenated coding systems using non-partitioned decoders. The talk will require that the fundamental, atomic element of a decoder, i.e., the add compare select unit, remain unchanged but will impose no restriction on which state a given processor handles at any time. The lower bounds on intrasystem communication will result from following the flow of imaginary tokens which move along the same paths as the accumulated metrics. The tokens will flow through an encoder state diagram which has been unraveled in time. Each column from left to right represents the next decoded bit time and the nodes within a column are the different possible states at that time.

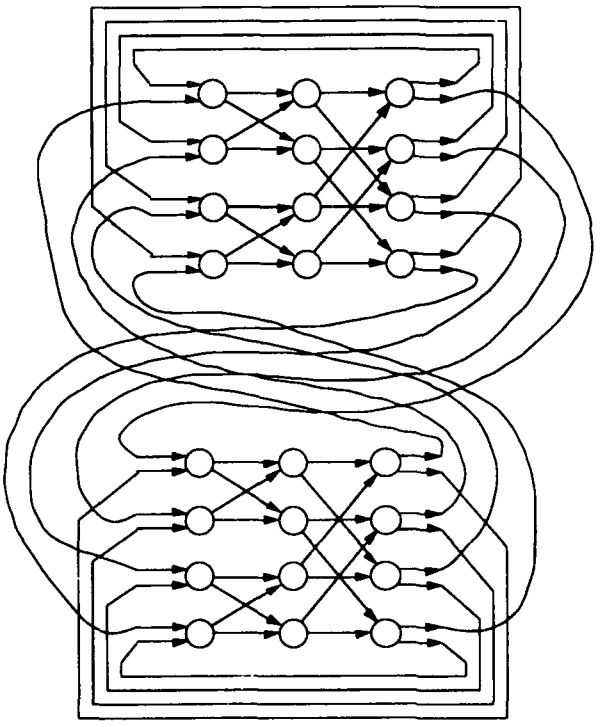
The easiest way to discover what this new time-flow graph looks like is to examine a logically equivalent but structurally different encoder, i.e., given the same sequence of input bits this new encoder always produces the same sequence of output bits that a conventional convolutional encoder would; however, the internal mechanics are very different. The state of this encoder will, of course, depend on the most recent K bits of the data stream, but instead of shifting all of the bits to the right to make room for a new incoming bit, the bits which make up the state are replaced sequentially, and no shifting takes place. The new encoder is time varying; it contains a pointer which starts out at the first memory cell and then moves on to the second, third and so on. When it reaches the last cell it cycles back to the first. The pointer indicates which bit of the state will be replaced by the new information bit. A suitably rotating set of generator polynomials will clearly allow this new encoder to reproduce the same sequence of symbols as the original. Although, technically, time is now a part of the encoder state, the receiver does not have to estimate this variable.

The time unraveled state diagram of this encoder and, of course, the conventional shift register encoder follows directly from its picture and looks like an endless succession of FFT's placed one after the other in a line with the right hand circles of one overlapping the left hand circles of the next. If a suitable decoding procedure is used it is easy to show by using this time unraveled state diagram that the memory of an outer code interleaver can be reduced without increasing the probability of there being an error somewhere in the interleaving block.

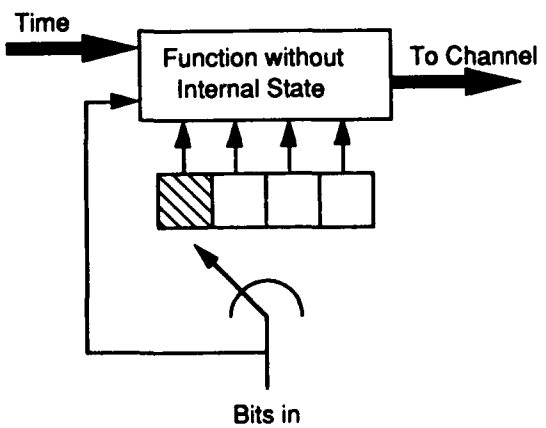
Only one special property of the time unraveled state diagram is necessary in order to analyze the information flow in the decoder, *viz.* the two paths which leave a node remain completely separate for K-2 time steps. This observation is sufficient to show the following upper bound on the average residence time of a token in a module where X is the total number of state pairs which the module can hold:

$$L_{\max avg} \le \frac{1+\log(x)+\sum_{n=\log(x)+1}^{K-2} 2^{(\log(x)-i)}}{1-2^{(\log(x)-K+2)}}$$

Surprisingly, in certain special cases this very coarse bound can be achieved with equality as the following time unraveled state diagram illustrating the split of a *K=5 decoder* into two equal parts shows:



K=5 Equivalent Encoder



ON THE MAXIMUM DIFFERENCE BETWEEN PATH METRICS IN A VITERBI DECODER

Andries P. Hekstra*

Abstract

The number of bits to be used in the path metric calculus of Viterbi decoders depends logarithmically on the maximum possible difference between any two path metrics. Here, a recent upper bound of Alston and Chau is generalised. In addition, we obtain an easy-to-compute exact expression for the maximum path metric difference.

Introduction

Correct operation of the well-known Viterbi algorithm depends only on differences of path metrics. As shown in [1], two's complement arithmetic can faithfully represent these metric differences, which leads to an efficient implementation. The number of bits to be used in this calculus depends on the maximum possible difference $\max\{\Delta pm\}$ of any two path metrics at an arbitrary depth L in the trellis. Let s_{max} denote the maximum symbol metric. Assuming nonnegative symbol metrics, an elementary upper bound on the maximum path metric difference $\max\{\Delta pm\}$ is [1, 2, 3]

$$\max\{\Delta pm\} \leq s_{max} nm \tag{1}$$

where m denotes the memory order of the rate $R(= k/n)$ convolutional code, i.e. the maximum shift register length [3].

Recently, Alston and Chau obtained a new upper bound for decoders of binary $R = 1/n$ codes, under certain assumptions on the metric function [2],

$$\max\{\Delta pm\} \leq s_{max} \lfloor n(m+1) - \frac{d_{free}}{2} \rfloor. \tag{2}$$

The assumptions of Alston and Chau can be simplified to the following.

I Symbol metrics are nonnegative integers.

II A branch metric is the sum of n symbol metrics.

III If transmission is over a noiseless channel and the transmitted bit differs from the hypothesis bit the maximum symbol metric s_{max} is assigned, otherwise the symbol metric equals zero.

If negative branch metrics can occur, the addition of a constant to all branch metrics will not affect the operation of the Viterbi decoder. For binary codes, it can be shown that Assumption III does not entail a loss of generality either, again because path selection is determined by the difference of symbol metrics. The difference of a symbol metric given that a zero was sent and the symbol metric given that a one was sent can always be negative, irrespective of Assumption III.

We show that reception of the all zeroes sequence constitutes the worst case for $\max\{\Delta pm\}$, for any depth $L \geq m$. As a result, we obtain an easy-to-compute, exact expression for $\max\{\Delta pm\}$ (Theorem I) and a generalisation of the Alston and Chau bound (Theorem II). By the *hard decision path metric*, we mean the path metric if the (soft decision) Viterbi decoder were of the hard decision type and the same signal were received. Since these results were obtained obtained, we have learned that an exact expression for maximum path metric differences can also be found in a NASA internal report.

Theorem I

Under the Assumptions I-III, for any (n, k, m, M, d_{free}) binary convolutional code and for an arbitrary received sequence, the maximum path metric difference $\max\{\Delta pm\}$ equals s_{max} times the maximum hard decision path metric of a final state in a trellis L branches deep, starting from the all-zero state (for arbitrary $L \geq m$).

Let M denote the total number of shift register cells in the encoder. Throughout the paper, a realization of the encoder as a parallel combination of k shift registers $i = 1, 2, \ldots, k$ for which the i-th shift register is m_i cells long [3] is assumed. The memory order m is defined as the maximum of the m_i, $M = \sum_{i=1}^{k} m_i$. Then, ΔM is defined as

$$\Delta M = km - M. \tag{3}$$

Of course, for direct evaluation of Theorem I the case $L = m$ is the easiest to evaluate. However, values $L > m$ can be used to obtain an analytical upper bound to $\max\{\Delta pm\}$.

Theorem II (Generalized Alston and Chau bound)

Under the Assumptions I-II, for any (n, k, m, M, d_{free}) binary convolutional code, the maximum path metric difference is upper bounded by

$$\max\{\Delta pm\} \leq s_{max} \min\{ \lfloor n(m+\delta) - d_{free}(1 - 2^{-(\Delta M + k\delta)}) \rfloor \mid \delta = 0, 1, \ldots \}. \tag{4}$$

References

[1] A.P. Hekstra, "An alternative to metric rescaling in Viterbi decoders," IEEE Trans. Commun., vol. COM-37, pp. 1220-1222, Nov. 1989.

[2] M.D. Alston, P.M. Chau, "An improved analytical bound on the maximum difference between path metrics in Viterbi decoders of binary tree convolutional codes", submitted to IEEE Trans. on Commun, vol. COM-40, 1992.

[3] S. Lin, D.J. Costello, Error control coding fundamentals and applications, New Jersey: Prentice Hall, 1983.

[4] P.H. Siegel, C.B. Shung, T.D. Howell, H.K. Thaper, "Exact bounds for Viterbi detector path metric differences", Proc. of IEEE Int. Conf. Acoust., Speech and Signal Proc., pp. 1093-1096. May 13-16, 1991.

*PTT Research, P.O. Box 421, 2260 AK Leidschendam; email: A.P.Hekstra@research.ptt.nl

LIST OUTPUT AND SOFT SYMBOL OUTPUT VITERBI ALGORITHMS: EXTENSIONS AND CONNECTIONS

Christiane Nill
European Space Agency
ESRIN
Frascati, Italy

Carl-Erik W. Sundberg
Signal Processing Research Department
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, New Jersey 07974

ABSTRACT

The Viterbi Algorithm (VA) is the optimum decoding algorithm for a convolutional code. Improvements in the performance of a concatenated coding system that uses VA decoding (inner decoder) can be obtained when, in addition to the stanard output, an indicator of the reliability of the VA decision is delivered to the outer stage of processing. Two different approaches of extending the VA are considered. In the first approach, the VA is extended with a Soft Output unit (SOVA) that calculates, based on the difference between the cumulative metrics of the two paths merging at each time instant and state, reliability values for each of the decoded information *symbols*. In the second approach, coding gains are obtained by delivering, in addition to the best path, the next $L-1$ best estimates of the transmitted data *sequence*. Here, the output format is a list of size L. This is a List VA (LVA). In this work, we evaluate LVA and SOVA in comparison to each other and attain extended versions of LVA and SOVA with low complexity that implement the other algorithm. We construct and evaluate a List-SOVA using the reliability information of the SOVA to generate a list of size L and that also has a lower complexity than the LVA for a long list size. Further, we introduce a low complexity algorithm that accepts the list output of the LVA and calculates for each of the decoded information bits a reliability value. The complexity and the performance of this Soft-LVA is a function of the list size L. The performances of Soft-LVA and SOVA are compared in concatenated coding systems.

SUMMARY

The Soft-Output Viterbi Algorithm (SOVA) [1] and the Generalized Viterbi Algorithm (GVA) or List output VA (LVA) [2] are further extended and compared in this paper. The LVA produces a list of the L best estimates of the transmitted *data sequence*. The SOVA, however, generates sequentially and continuously soft output *symbols*, where the amplitude of each symbol contains the reliability information for that specific symbol. Two different units are developed, both using the reliability information to produce a list of size L. Assuming an ideal outer code, the performances of these two List-SOVAs (SOVA and List Generating Units) are compared to the performance of the LVA for the Gaussian and independent Rayleigh fading channels. A Soft Symbol Output Algorithm is defined, using the differences in the accumulated metrics of the best path and the L best paths ($1 < l \leq L$) of the LVA as a measurement for the reliability of each decoded bit. The serial LVA (SLVA) [2] generates this list iteratively. A new software implementation of the SLVA is presented. The new Soft-LVAs and the SOVA are tested in a concatenated coding system, where a convolutional code is chosen as the outer code. The two algorithms of interest — LVA and SOVA — evaluate the reliability of the VA decisions and deliver the attained information to an outer stage of processing. In case of erroneous decisions of the VA, we observe in the VA outputs correlated information bit errors and the SOVA outputs correlated symbol reliability values.

Motivated by the idea of finding algorithms with lower complexity, which produce list output (long list size L) or soft output, we extended the two algorithms by additional units to produce output according to the other algorithm. We defined a low complexity list generating unit that accepts the deinterleaved SOVA output of length N and generates, by using the reliability values as "differences" for a 1-state SLVA (new implementation) decoding process, a list of length L. LVA and List-SOVA achieve equal coding gains (for the Gaussian channel) of about 1.0 dB for a list size $L = 2$ and 1.5 dB for a list size $L = 3$ compared to the VA performance. Due to the higher slope of the error probability of the List-SOVA, the inferior performance for low SNRs changes into a superior performance for high SNRs when compared to the LVA. We explain these results due to the use of interleaver for the List-SOVA. We attain with the List-SOVA an alternative List Output Viterbi Algorithm that achieves due to costs of higher decoding delays (interleaving) than the LVA a superior performance for high SNRs. For a short list size the complexity of the List-SOVA is higher than the complexity of the LVA. We propose in future work to study the List-SOVA performance as a function of the update length δ_{up} to obtain possibly even for a short δ_{up} an acceptable performance.

We introduce a low complexity Soft Symbol Output Viterbi Algorithm, based on the LVA (Soft-LVA) that uses the knowledge of the positions where the L best path differ and the cumulative metrics at state s_N of the L best paths, to produce reliability information for each of the decoded information bits. Due to the fact that the L best path sequences only differ at a limited number of information bits we discovered that with "soft" initialization of the reliability values coding gains can be achieved versus a scheme with fixed initialization values where smaller coding gain could be obtained. For the preliminary "soft" initialization method that is based on a SOVA update (obtained from the SLVA) with update length $\delta_{up} = \nu + 1$, we achieve, e.g for a list size of $L = 2$ (low complexity) in comparison to hard output decoding a coding gain of 0.5 dB at 10^{-3} for codes with memory 3 for the inner and outer code and code rates 1/2 (inner) and 2/3 (outer). With list size $L = 8$, 1 dB coding gain is achieved. We assume that in the first 8 estimates of the sequence the significant error events in the VA output are considered for block lengths $N = 32, 64, 128, 512$. Compared to the SOVA the proposed Soft-LVA has 0.2 dB lower coding gain at a bit error probability of 10^{-3}, $N = 64$, $L = 8$. Concerning the complexity of the algorithms, the Soft-LVA, especially when the SLVA is used, has a lower complexity than the SOVA, [3].

References

[1] J. Hagenauer, P. Hoeher and J. Huber, "Soft-Output Viterbi and Symbol-by-Symbol MAP Decoding: Algorithms and Applications." In submission to *IEEE Transactions on Com.*

[2] N. Seshadri and C-E. W. Sundberg, "List Viterbi Algorithms and their Applications to Speech and Data Transmission." In submission to *IEEE Transactions on Com.*

[3] C. Nill and C-E. W. Sundberg, "List and Soft Symbol Output Viterbi Algorithms: Extensions and Comparisons." In submission to *IEEE Transactions on Com.*

NEW LOW COMPLEXITY SOFT MAXIMUM LIKELIHOOD DECODING OF PARTIAL UNIT MEMORY CODES.

V.V.Zyablov*, B.Honary, G.Markarian****

*) Institute for Problems of Information Transmission, Russian Academy of Sciences, 19 Ermolova Str., 101447, Moscow, C.I.S.

**)Hull-Lancaster Communication Research Group, Lancaster University, Lancaster, LA1 4YR, UK.

Recent publication by Forney [1] has increased an interest paid to trellis decoding of block codes and different combined coding and modulation techniques. Partial unit memory (PUM) codes introduced in [2] have advantages of both block and convolutional codes. Soft maximum likelihood decoding (SMLD) based on trellis structure can be obtained for these codes, however such a decoder will have huge number of states and branches. The problem of reducing complexity of such decoders has been investigated in [3], where trellis decoding derived was based on technique described by J.Wolf [4]. In [5] a new simple algorithm for trellis design, using generator matrix of array codes was proposed. This algorithm allows to derive the trellis diagram for any array code with reduced number of states and branches.

In this paper a new low complexity SMLD for PUM codes, based on [5] is introduced. It is shown that, the new technique will provide the lowest implementation complexity together with better distance properties in comparison with conventional techniques. Let $\{X\}=\{(x_1,x_2,...x_k)\}$ be the sequence of input information symbols of length k, and the sequence of output codewords of length n (n>k) $\{Y\}=\{(y_1,y_2,...y_n)\}$ is defined as:

$$Y_i = X_i G_0 + X_{i-1} G_1 \qquad (1)$$

where, G_0 and G_1 are kn matrices. If $\text{rank}(G_0)=k$ and $\text{rank}(G_1)=k_1<k$, such code is known as PUM code [2]. Matrices G_0 and G_1 are defined as follows:

$$G_0 = ||G_{00}G_{01}||^T \qquad G_1 = ||0G_{11}|| \qquad (2)$$

whrer, G_{00} is a generator matrix of (n,k_0) block code ($k_0=k-k_1$); G_{01} and G_{11} are (kn) matrices of rank k_1; $||0||$ is $k_0 n$ all zero matrix and $||.||^T$ is the transpose of matrix $||.||$. In order to design the PUM code with $d_{free}=d_{min}$ of block code, we choose the G_{00} as a generator matrix of array code, but matrices G_{01} and G_{11} must satisfy to following conditions:

(i) the distances of block codes, generated by matrices G_0 and G_1 must be no less than $d_{min}/2$ of array code;

(ii) matrix

$$G_n = ||G_{00}G_{01}G_{11}|| \qquad (3)$$

must be non-singular.

Example. We describe an array code (9,4,4) with generator matrix:

$$G = \begin{matrix}101000101\\011000011\\000101101\\000011011\end{matrix} \qquad G_{00} = \begin{matrix}101000000\\011000000\end{matrix} \qquad G_{11} = \begin{matrix}000100001\\001000100\end{matrix}$$

Using the trellis diagram of array code [5], the trelis structure of PUM code can be derived easily. This code has the following parameters: n=9, k=6 and $d_{free}=4$. The designed trellis diagram allows to implement the SMLD of PUM code with much lower complexity comparing with conventional techniques. Table 1 compares the complexity of trellis diagrams and number of operations for conventional Viterbi decoder (VD), SMLD using J.Wolf's trellis structure of block codes [3] and for PUM code, proposed in the above Example. As it follows from Table 1, the proposed algorithm allows to achieve the lowest complexity.

Table 1.

Decoding Algorithm	No of Additions	No of Comparis.	No of Branches	No of Nodes
VD	2048	240	256	8
SMLD[3]	296	148	296	152
Example	172	75	100	28

In addition, the technique described above allows to increase the distance properties of PUM.

REFERENCES

1. Forney G. Coset codes-Part 1: Introduction and geometrical classification.-"IEEE Transaction on Information Theory", vol.34,1988,p.p.1123-1151.
2. Lee L.N. Short unit-memory byte oriented convolutional codes, having maximal free distance.-"IEEE Transactions on Information Theory", vol.22, No 3, 1976, p.p.349-355.
3. Zyablov V.V., Sidorenko V.R. Soft maximum likelihood decoding for PUM codes.-"Problems of Information Transmission", No 1, 1992.
4. Wolf J.K. Efficient maximum likelihood decoding of linear block codes, using a trellis.-"IEEE Transactions on Information Theory", vol.24, No 1, 1978, p.p.76-84.
5. Honary B., Markarian G., Darnell M. Trellis decoding technique for array codes.- "Proceedings of Eurocode'92", Udine, Italy.

On the Evaluation of the Error Performance of Trellis Codes

Christian Schlegel
Digital Communications Group
University of South Australia
The Levels, SA 5095, Australia

Summary

The evaluation of the first event error probability P_e of trellis codes is a difficult and complex problem. The best known approach is the truncated union bound P_{ub} on P_e, but even the evaluation of P_{ub} is rather complex for most codes, due to the nonlinear or nonregular structure of typical trellis codes, which requires a double summation over all sequences, i.e.,

$$P_{ub} = \sum_{\mathbf{x},\mathbf{x}'} Q\left(\sqrt{\frac{d^2(\mathbf{x},\mathbf{x}')}{2N_0}}\right),$$

where $\mathbf{x}, \mathbf{x}'$ are the sequences of the code. The complexity of the search algorithm is thus proportional to N^2, where N is the number of states in the code. The above equation is often written in terms of the distance spectrum as

$$P_{ub} = \sum_{d \geq d_{free}} A_d Q\left(\sqrt{\frac{d^2}{2N_0}}\right),$$

where the infinite set of pairs $\{d^2, A_d\}$ is the distance spectrum and d_{free} is the minimum squared Euclidean distance of the code. We thus concentrate on evaluating the distance spectra of these codes.

A lot of effort has gone into designing regular trellis codes and the condition for regularity (or geometric uniformity) are well understood now [1]. The reason why regular codes are so popular is that their error performance can be evaluated by regarding a single, arbitrary correct sequence, i.e., the double summation above is reduced to a single summation. The complexity gain thus achieved is significant and searching a trellis with N states, where N is the number of code states is considered acceptable. In this sense, calculating the error performance of regular codes is equivalent in complexity to calculating the error performance of linear codes.

Various researchers [2, 3, 4, 5] have successfully looked at reducing the complexity of calculating the union bound of trellis codes, using the linear structure of the underlying generating trellis. All these methods are essentially equivalent [6], in the sense that the linearity of the code generating the trellis is extended to the average symbol sequence.

In this paper we look at this problem from the perspective of graph theory and finite-state machines. The original problem of calculating the union bound can be formulated as finding all possible paths through a graph with N^2 states, where N is the number of code states. We show that the number of vertices of this graph can be reduced in size in many cases. We show that the conditions for quasi-regularity in [3] (or row and column uniformity in [4]) lead to sets of equivalent states, reducing the size of the graph to N vertices. This approach does not explicitly use the linearity of the generating code generating trellis and is thus applicable to nonlinear trellis codes such as some of the rotationally invariant codes.

The distance spectrum for some of these trellis codes can actually be calculated using a graph with fewer than N states, as in the case of the 4-state 8-PSK Ungerboeck code, whose associated Euclidean distance graph has only 3 states. We expound on this and look at how the distance graph can be used to obtain bounds for other distance measures which can be computed with small complexity.

Ways of loosening the bound which leads to complexity reductions and the tightness of the bounds will also be addressed.

References

[1] G.D. Forney, "Geometrically Uniform Codes," *Trans. on Inform. Theory*, vol. 18, pp. 1241-1260, September 1991.

[2] E. Zehavi and J.K. Wolf, "On the performance evaluation of trellis codes," *Trans. on Inform. Theory*, vol. 33 no. 2, pp. 196–201, March 1987.

[3] M. Rouanne and D.J. Costello, Jr., "An Algorithm for Computing the Distance Spectrum of Trellis Codes," *IEEE Journal on Selected Areas in Communications*, vol. 7, No. 6, pp. 929–940, August 1989.

[4] E. Biglieri and P.J. McLane, "Uniform distance and error probability properties of TCM schemes," *ICC'89*, Boston, Mass., June 11–14, 1989.

[5] E. Biglieri et. al. *Introduction to Trellis-Coded Modulation with Applications*, Macmillan, New York, 1991.

[6] C. Schlegel, "Evaluating Distance Spectra and Performance Bounds of Trellis Codes on Channels with Intersymbol Interference", *Trans. on Inform. Theory*, May, 1991.

SOFT SYNDROME DECODING OF TRELLIS CODED MODULATION CODES

Meir Ariel and Jakov Snyders

Department of Electrical Engineering - Systems,
Tel-Aviv University, Ramat-Aviv 69978, Israel.

Abstract

A recursive algorithm is presented for accomplishing maximum likelihood soft syndrome decoding of trellis coded modulation codes. It consists of signal-by-signal hard decoding, followed by a search for the most likely error pattern. An error trellis, alternatively a decoding table, is devised for describing the decoding procedure. Methods for degenerating the error trellis enable identification of the surviving error path by a relatively small number of real additions. In comparison with the Viterbi algorithm, the syndrome decoder achieves substantial reduction in the computational complexity, especially for moderately noisy channels.

Summary

Trellis coded modulation (TCM) codes are usually generated by employing an $k/(k+1)$ rate binary convolutional encoder. The $k+1$ coded bits select one of 2^{k+1} subsets of a redundant 2^{k+m+1}-ary signal set, while the m uncoded bits determine which of the 2^m signals of this subset is to be transmitted. The signals are transmitted through an additive white Gaussian noise channel, hence maximum likelihood (ML) decoding is equivalent to minimum squared Euclidean distance decoding.

The ML decoding of TCM codes is customarily accomplished in two steps: a) within each subset of signals, the nearest neighbour to the received signal is determined by a procedure called *subset decoding*, then b) the Viterbi algorithm is applied for finding the signal path through the codeword trellis with the minimum squared Euclidean distance from the sequence of received signals [2]. We replace step b) by an efficient ML syndrome decoding algorithm, suited to deal with nonbinary modulation signals.

Let H be an infinite dimensional parity check matrix of an $(k+1,k)$ binary convolutional code C employed for a TCM scheme. Given some received sequence of channel output signals, $\mathbf{r}=(r_1,r_2,r_3,\cdots)$, signal-by-signal *hard decision* yields $\mathbf{v}=(v_1,v_2,v_3,\cdots)$. By *hard decision* we mean finding the closest code signal to the received signal in terms of squared Euclidean distance. The subset decoding is also accomplished as part of the hard decision procedure. We then expand each signal in $\mathbf{v}$ into its $k+1$ coded bits representation (i.e., discarde the uncoded bits). Subsequently, we compute the corresponding sequence of syndrome bits $\mathbf{z}=(z_1,z_2,z_3,\cdots)^t$, defined by $\mathbf{z}=H\mathbf{v}^t$ (where the superscript t indicates transposition). A measure of confidence, named *confidence value*, is assigned to each of the 2^{k+1} signals,

$$c_i^j \ ; j=0,1,2,\cdots,2^{k+1}-1 \ ,$$

belonging to the reduced signal set (i.e., the signals determined by the subset decoding). The confidence value, denoted μ_i^j, is defined by

$$\mu_i^j = d^2(c_i^j, r_i) - d^2(v_i, r_i),$$

where $d^2(\cdot,\cdot)$ stands for squared Euclidean distance. The columns of H are partitioned into sets of $k+1$ consecutive columns each, called *bands*, then μ_i^j is also regarded as the confidence value of the jth *subband* (i.e., subset) of the ith band of H. The *weight* of a collection of subbands belonging to different (not necessarily consecutive) bands, is defined to be the sum of the confidence values of the elements of the collection.

The decoding procedure starts with the computation of $\mathbf{z}$. If $\mathbf{z}=0$ then $\mathbf{v}$ is the most likely sequence of coded bits. Otherwise, ML decoding is achieved by finding the *least weighing error collection of subbands that sum up to* $\mathbf{z}$ and then complementing the bits of $\mathbf{v}$ at the positions corresponding to the columns included in this collection. An error trellis is devised for compactly describing all the possible error collections. The trellis is degenerated according to the composition of the syndrome sequence.

The following Table exhibits a comparison between computational complexities of the Viterbi algorithm and the proposed syndrome decoding algorithm, applied to a simple four state trellis code for 8-PSK modulation. The two algorithms will decode to the same code sequence and thus give identical performance. However, the Viterbi decoder's computational complexity is independent of the channel signal to noise ratio (SNR) while the syndrome decoder's average computational complexity decreases as the SNR increases (in similarity with the case of syndrome decoding of binary convolutional codes [3]). The complexity is measured in terms of the total numbers of real additions and comparisons, required for decoding 300-bit truncated sequences. The average complexities were obtained by simulations. The worst case complexities are fixed and independent of the SNR.

SNR [dB]	Viterbi algorithm	Syndrome worst case	Syndrome average
3	1800	1500	670
6	1800	1500	320
8	1800	1500	124

This research was supported in part by the Basic Research Foundation administered by the Israel Academy of Sciences and Humanities.

References

[1] G.D. Forney, Jr., "Convolutional codes II. Maximum likelihood decoding," *Inform. Contr.*, vol. 25, *pp.* 222-266, July 1974.

[2] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, *pp.* 56-67, Jan. 1982.

[3] M. Ariel, and J. Snyders, "Soft syndrome decoding of binary convolutional codes," submitted for publication.

[4] J.P.M. Schalkwijk, A.J. Vinck, and K.A. Post, "Syndrome decoding of binary rate k/n convolutional codes," *IEEE Trans. Inform. Theory*, vol. IT-24, *pp.* 553-562, Sept. 1978.

[5] J. Snyders, and Y. Be'ery, "Maximum likelihood soft decoding of binary block codes and decoders for the Golay codes," *IEEE Trans. Inform. Theory*, vol. IT-35, *pp.* 963-975 Sept. 1989.

[6] H. Miyakawa and T. Kaneko, "Decoding algorithms of error-correcting codes by use of analog weights," *Electronics and Communications in Japan*, vol. 58-A, pp. 18-27, Jan. 1975.

[7] J. Snyders, "Reduced lists of error patterns for maximum likelihood soft decoding," *IEEE Trans. Inform Theory*, vol. IT-37, pp. 1194-1200, July 1991.

[8] A.J. Viterbi "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun. Technol.*, vol. COM-19, *pp.* 751-772, Oct. 1971.

ERROR PROBABILITY ANALYSIS IN REDUCED STATE TCM

Carlos Valdez † Hiroyuki Fujiwara † Ikuo Oka ‡

†Univ. of Electro-Communications
Chofugaoka 1-5-1, Chofu-shi,
Tokyo 182, JAPAN

‡Osaka City University
Sumiyoshi-ku, Sugimoto 3-3-138,
Osaka 558, JAPAN

In Full State (FS) Viterbi decoding, $N = N_\epsilon \times N_{ch}$ is the number of states of the FS trellis and of a *super-encoder* which includes the convolutional encoder and the channel number of states, given by N_ϵ and N_{ch} respectively. In Reduced State (RS) schemes [1], reduction of the channel states to N'_{ch} ($1 \le N'_{ch} \le N_{ch}$), yields the size $|\mathcal{F}_s|$ of the FS super-encoder mapper also diminished, with a N'-state RS super-encoder with mapper's size $|\mathcal{F}'_s|$. Thus, per each state (or symbol)in the RS trellis we have $p = N_{ch}/N'_{ch} = |\mathcal{F}_s|/|\mathcal{F}'_s|$ unknown states (or symbols), where the symbols are estimated by the Viterbi decoder and then fed back to equalize the unknown ones in the next decoding step. If among the p potential survivors, an incorrect sequence is selected, the *error propagation* (EP) effect occurs. The feedback mechanism prevents the attainment of error probability bounds from the RS state transition diagram. To overcome this difficulty, the N' reduced states can be combined with the p unknowns, to obtain the N-state FS trellis ($N'p = N$), with labels including symbol estimations. Since in RS decoding at least two error events are involved, we consider the occurrence of multiple error events in the FS trellis resultant from splitting of the RS states, where all the states $s_0, s_1, \ldots, s_{p-1}$ are seen the same by the decoder. Its upper bound is obtained as

$$P_E < \sum_{K=1}^{\infty} \sum_{L=1}^{\infty} \sum_{S_{K,L}} \sum_{S'_{K,L}} P_c[S_{K,L}] \sum_{i=0}^{p-1} P_K[s_i] \sum_{j=1}^{p-1} P[S_{K,L} \to S'_{K,L}/s_i \to s_j] \tag{1}$$

where the correct sequence is denoted by $S_{K,L}$ and its occurrence probability as $P_c[S_{K,L}]$. K stands for error event order and L represents its length. $P_K[s_i]$ is the K-th error event starting state s_i probability, and $P[S_{K,L} \to S'_{K,L}/s_i \to s_j]$ is the pairwise error probability for the incorrect path $S'_{K,L}$ between s_i and s_j. From the first error event definition, the term ($K = 1$) in Eq. (1) carries no EP effect. By assuming that the same starts from the transmitted state s_0, $P_1[s_0] = 1$ and $P_1[s_i] = 0$ ($i = 1, 2, \ldots, p-1$). For $K \ge 2$, the error events may start from any s_i but from s_0, since at that time at least one error event have already happened (i.e. $P_K[s_0] = 0$ for $K \ge 2$). These terms will contain the EP effect. Moreover, all error events will end at any s_i ($i = 1, 2, \ldots, p-1$), where s_0 is excepted since at the time of RS decision, the transmitted and decoded states are not fully coincident.

Now, we calculate Eq. (1) by two methods. The first considers the *error weight matrix* transfer function [2]. We define two transfer function matrices : $\mathbf{T}_a(z)$ which represents the transfer function of the first error event starting from s_0 and ending at s_j ($j = 1, 2, \ldots, p-1$), and $\mathbf{T}_b(z)$ that corresponds to the error events starting from s_i and ending at s_j ($i, j = 1, 2, \ldots, p-1$). z is a parameter resulting from the Chernoff bound. An element (i, j) of the matrices is an error event probability upper bound associated to the paths between s_i and s_j. Then, we define two other transfer function matrices $\mathbf{T}'_a(z)$ and $\mathbf{T}'_b(z)$ with elements (i, j) containing the pairwise error (rather than the error event) probability upper bound. The total transfer function matrix becomes then, $\mathbf{T}(z) = \mathbf{T}_a(z) + \mathbf{T}'_a(z)[\mathbf{I} - \mathbf{T}'_b(z)]^{-1}\mathbf{T}_b(z)$ where all the factor multiplying $\mathbf{T}_b(z)$ gives the starting state probabilities for ($K \ge 2$) and $\mathbf{I}$ is a $N \times N$ identity matrix. To obtain the bit error probability upper bound, we extend $\mathbf{T}_a(z)$ and $\mathbf{T}_b(z)$ to include the number ϵ of incorrect input bits resulting $\mathbf{T}_a(z, I)$ and $\mathbf{T}_b(z, I)$, where I is an indeterminate whose exponent is ϵ. The total extended transfer function results then $\mathbf{T}(z, I) = \mathbf{T}_a(z, I) + \mathbf{T}'_a(z)[\mathbf{I} - \mathbf{T}'_b(z)]^{-1}\mathbf{T}_b(z, I)$ and the upper bound of the bit error probability is obtained as in [2]. In the calculation of $\mathbf{T}(z, I)$, the Chernoff bound of the pairwise error probability is taken, and the elements of all the defined matrices are calculated. We call this the Union Bound (UB).

The second method consists in the estimation by Gaussian Quadrature Rules (GQR), of the CDF associated to the pairwise error probability of Eq. (1). This is based on the calculation of a set of moments of the r.v. related to the branch metrics. Since the path metric is constructed by succesive branch metric additions, the moments of the correspondent r.v. are obtained from succesive binomial expansions in the form of moments convolution. When an error event is defined, such moments are used to estimate the CDF, from where the pairwise error probability is computed. In this case, we use an algorithm and include all the error events by traversing in a general N^2-state trellis diagram based on pairwise states [3]. The bound thus obtained is called Moments Bound (MB).

Although the pairwise error probability of Eq. (1), based on the path metric condition, holds equally for trellises with or without parallel transitions, in the first case, an additional condition is satisfied since a decoded branch symbol is, among the members of the parallel transition, the nearest in Euclidean distance to the received signal. By union bounding all the sequences arising from parallel transitions concatenation, an upper bound is obtained. An element taken from it will have a branch metric that is calculated from a truncated Gaussian noise pdf, with limits depending on how the decision space is divided by the parallel transition symbols. The tighter bound like this obtained is called Elementary Bound (EB).

The results shown below correspond to a 16-QAM TCM scheme, with $N_\epsilon = 4$, $N_{ch} = 8$, $N'_{ch} = 1$. The tight upper bound (TUB) [2] shown can be applied only for complete Gaussian noise pdf or UB.

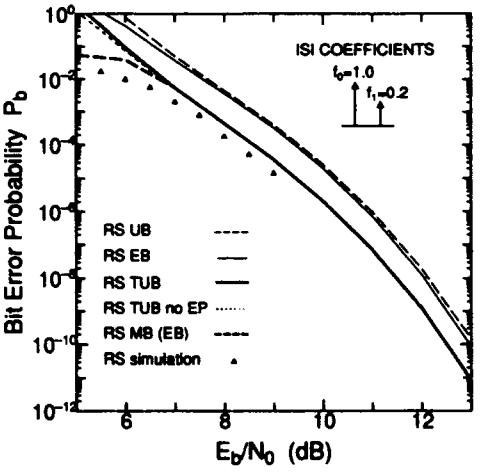


Figure 1: RS 16-QAM TCM performance bounds

References

[1] P.R.Chevillat, E.Eleftheriou, *Decoding of Trellis-Encoded Signals in the Presence of Intersymbol Interference*, IEEE Trans. Commun. Vol.COM-37, No.7, pp.669-676, July 1989.

[2] Y.J.Liu, I.Oka, E.Biglieri, *Error Probability for Digital Transmission Over Nonlinear Channels with Application to TCM*, IEEE Trans. on Inform. Theory, Vol.IT-36, No.5, pp.1101-1110, September 1990.

[3] C.Valdez, H.Fujiwara, I.Oka, H.Yamamoto, *Error Probability Analysis in Reduced State Viterbi Decoding*, IEICE Technical Report, Vol.92, No.193, pp.55-60, August 1992.

This work was partially supported by granted aid of Fujitsu Laboratories, Ltd.

EFFICIENT MAXIMUM-LIKELIHOOD SOFT-DECISION DECODING OF LINEAR BLOCK CODES USING ALGORITHM A^*[1]

Yunghsiang S. Han, Carlos R. P. Hartmann, and Chih-Chieh Chen[2]

Abstract

In this paper we present a novel maximum-likelihood soft-decision decoding algorithm for linear block codes. The approach used here is to convert the decoding problem into a search problem through a graph which is a trellis for an equivalent code of the transmitted code. Algorithm A^* is employed to search through this graph. This search is guided by an evaluation function f defined to take advantage of the information provided by the received vector and the inherent properties of the transmitted code. This function f is used to drastically reduce the search space and to make the decoding efforts of this decoding algorithm adaptable to the noise level. Simulation results for the (104, 52) binary extended quadratic residue code and the (128,64) binary extended BCH code are given.

Summary

The use of block codes is a well-known error-control technique for reliable transmission of digital information over noisy communication channels. Linear block codes with good coding gains have been known for many years; however, these block codes have not been used in practice for lack of an efficient soft-decision decoding algorithm.

In this paper we present a novel maximum-likelihood soft-decision decoding algorithm for linear block codes. This algorithm uses Algorithm A^* [3], which is a generalization of Dijkstra's algorithm [2] to search through the trellis for a code equivalent to the transmitted code. This search is guided by an evaluation function f defined for every node m in the trellis. $f(m) = g(m) + h(m)$, where $g(m)$ is an estimate of the cost of the minimum cost path from the all-zero node at depth 0 to node m, and where $h(m)$ is an estimate of the cost of the minimum cost path from node m to the all-zero node at depth n. The function f is defined to take advantage of the information provided by the received vector and the inherent properties of the transmitted code. The use of this priority-first search strategy for decoding drastically reduces the search space and results in an efficient optimal soft-decision decoding algorithm for linear block codes. Furthermore, in contrast with Wolf's algorithm [4], the decoding efforts of our decoding algorithm are adaptable to the noise level.

The proposed algorithm is applicable to any linear block code and does not require the availability of a hard-decision decoder. Furthermore, any stopping criterion ensuring that a solution has been found can be easily incorporated into this algorithm.

Simulation results for the (104,52) binary extended quadratic residue code and the (128,64) binary extended BCH code when these codes are transmitted over the AWGN channel are given in tables 1 and 2, respectively. These results were obtained by simulating 35,000 samples for each SNR.

Table 1: Simulation for the (104,52) code

γ_b	5 dB		6 dB		7 dB		8 dB	
	max	ave	max	ave	max	ave	max	ave
$N(r)$	142123	19	2918	1	221	1	0	0
$C(r)$	32823	5	519	2	35	2	1	1
$M(r)$	13122	4	1912	1	155	1	0	0

Table 2: Simulation for the (128,64) code

γ_b	5 dB		6 dB		7 dB		8 dB	
	max	ave	max	ave	max	ave	max	ave
$N(r)$	216052	42	13603	2	1143	1	0	0
$C(r)$	38219	8	1817	2	91	2	1	1
$M(r)$	16626	7	856	1	965	1	0	0

where

$N(r)$ = the number of nodes visited during the decoding of r;

$C(r)$ = number of codewords constructed in order to decide on the closest codeword to r;

$M(r)$ = maximum number of nodes stored during the decoding of r;

max = maximum value among 35,000 samples;

ave = average value among 35,000 samples;

$\gamma_b = E_b/N_0$.

Simulation results for the above linear block codes attest to the fact that this decoding technique drastically reduced the search space, especially for the majority of practical communication systems where the probability of error is less than 10^{-3} (γ_b greater than 6.8 dB) [1]. For example, the results of Table 2 at 6 dB indicates that for the 35,000 samples tried, this decoding algorithm is approximately 15 orders of magnitude more efficient, in time and space, than Wolf's. Thus, this decoding procedure has not only resulted in an efficient soft-decision decoding algorithm for hitherto intractable linear block codes, but an algorithm which is in fact optimal as well.

References

[1] G. C. Clark, Jr., and J. B. Cain, *Error-Correction Coding for Digital Communications*. New York, NY: Plenum Press, 1981.

[2] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambridge, MA: The MIT Press, 1991.

[3] N. J. Nilsson, *Principle of Artificial Intelligence*. Palo Alto, CA: Tioga Publishing Co., 1980.

[4] J. K. Wolf, "Efficient Maximum Likelihood Decoding of Linear Block Codes Using a Trellis," *IEEE Trans. on Information Theory*, pp. 76–80, January 1978.

[1]This work was partially supported by the National Science Foundation under Grant No. NCR-9205422, and used the computational facilities of the Northeast Parallel Architectures Center (NPAC) at Syracuse University.

[2]Y. S. Han and C. R. P. Hartmann are with the School of Computer and Information Science at Syracuse University, Syracuse, NY 13244-4100. C.-C. Chen was with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218. He is now with the Department of Computer Science, University of California, Los Angeles, CA 90024. Mr. Chen participated in this research during the summer of 1991 while he was working under the Syracuse Center of Computational Science Research Experience for Undergraduate Program grant from the National Science Foundation (NSF-REU award No. CDA-9100633).

ON THE PERFORMACE EVALUATION OF APPROXIMATE APP DECODING

S. A. RAGHAVAN

COMSTREAM CORPORATION
10180 BARNES CANYON ROAD
SAN DIEGO, CA 92121

Abstract

The performance evaluation of a priori probability (APP) decodable codes is considered. In particular, we are interested in a decoding method, which uses an approximation [2] to the weight functions involved in the decoding process. The channel is assumed to be additive white Gaussian, and the modulation format is assumed to be binary phase shift keying (BPSK). Two cases are considered: in the first case, analog (unquantized) demodulator output samples are used in soft decision approximate APP (AAPP) decoding. In the second case, we assume that the demodulator output samples are quantized using an analog-to-digital (A/D) converter, and these quantized samples are utilized in the AAPP decoding process. In both the cases, expressions for the probability of first decoding error are derived using characteristic functions. We compute the probability of first decoding error numerically for both block and convolutional APP decodable codes using the analytical expressions derived, and these results show good agreement with simulation results. Finally some interesting aspects of large block length threshold decodable codes are discussed [5].

Summary

APP decoding was introduced by Massey [1]. It provides a soft decision decoding algorithm that minimizes symbol error rate for threshold decodable codes [1, 3]. In APP soft decision decoding, a set of orthogonal parities are computed from the hard decisions, and each of these parities are assigned "weights" based upon the channel reliability information. In exact APP decoding, these weights are complex non-linear functions of the channel reliabilities of the components of the parity equations. Hence in reality, exact APP decoding is rarely implemented.

In [2], an approximation to the weight function which depends only on the least reliable component of the parity check equation was suggested, and shown *via simulations* to perform extremely well (degradation relative to exact APP decoding is less than a tenth of a dB) even at low Eb/No for a large set of threshold decodable codes. This approximation is widely used in practical realizations of soft decision threshold decoders [4].

In our work, we derive *analytical expressions* for the probability of first decoding error for AAPP soft decision decoding. In our analysis, we assume that already decoded symbols and their reliabilities are *not* fed back in the decoding of subsequent symbols. We also assume for the purpose of analysis, a "Type-II" decoder [1, 3] where any orthogonal parity

$$B = r_1 + r_2 + \ldots\ldots\ldots + r_n \qquad (1)$$

is an estimate of the orthogonal symbol C. In (1), r_i represent the hard decisions on the demodulated matched filter samples y_i, and the symbol $\oplus$ represents modulo-2 addition. In AAPP decoding, the weight of the parity B, denoted w(B), is given by

$$w(B) = \min \{|y_1|, |y_2|, \ldots\ldots, |y_n|\}. \qquad (2)$$

We proceed to evaluate the conditional characteristic function of the random variable w(B)(2B-1), conditioned upon the orthogonal symbol C. We then make use of the fact that given the orthogonal symbol, the parity check equations are independent, and derive conditional characteristic function of the decision variable in Type-II AAPP decoding. An efficient technique to numerically evaluate the probability of first decoding error from the characteristic function of the decision variable is also described.

We evaluate the probability of first decoding error for threshold decodable block codes using the analysis described above. We compare the numerical results to simulation results, and show that they are in good agreement. It is shown that when the demodulator output samples are quantized to 8-levels and AAPP decoding is used, the degradation relative to unquantized case is about 0.25 dB only when the length of the threshold decodable code is small. When the length of the threshold decodable code is large, the degradation is close to 1.0 dB. A new decoding algorithm [5] for the quantized sample case is proposed, which eliminates the above mentioned drawback for threshold decodable block codes with large block length.

Discussions with my collegue Y. Hebron are greatly appreciated during the course of this work.

References

[1] J. L. Massey, *Threshold Decoding*, MIT Press, 1963.

[2] H. Tanaka et al., "A novel approach to soft decision decoding of threshold decodable codes," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 244-246, March 1980.

[3] S. Lin and D. J. Costello, *Error Control Coding*, Prentice Hall, 1983.

[4] P. Lavoie et al., "New architechtures for fast soft-decision threshold decoders," *IEEE Transactions on Communications*, vol. COM-39, pp. 200-207, February 1991.

[5] S. A. Raghavan et al., "On the application of approximate APP decoding to digital video transmission," Accepted for publication, *IEEE Journal on Selected Areas in Communications*, Special issue on HDTV and digital video communications, December 1992.

Maximum–Likelihood Soft Decision Decoding of BCH codes

Alexander Vardy
IBM Research Division, Almaden Research Center
650 Harry Road, San Jose, CA 95120

Yair Be'ery
Tel-Aviv University, Department of Electrical Engineering
Ramat-Aviv 69978, Tel-Aviv, Israel

Abstract. The problem of efficient maximum-likelihood soft decision decoding of binary BCH codes is considered. It is known that those primitive BCH codes whose designed distance is one less than a power of two, contain subcodes of high dimension which consist of a direct sum of several identical codes. We show that the same kind of direct-sum structure exists in *all* the primitive BCH codes, as well as in the BCH codes of composite block length. We also introduce a related structure termed the "concurring-sum", and then establish its existence in the primitive binary BCH codes. Both structures are employed to upper bound the number of states in the proper minimal trellis of BCH codes, and develop efficient algorithms for maximum-likelihood soft decision decoding of these codes.

In [2] Forney has shown that the binary Reed-Muller codes contain direct-sum subcodes of high dimension. It is well known that certain BCH codes, namely the primitive binary BCH codes with designed distance one less than a power of two, are supercodes of punctured Reed-Muller codes. Hence these BCH codes evidently share the direct-sum structure of the RM codes. This fact was used by Kasami et al. [3] to construct efficient trellis diagrams for the (64,24,16) and (64,45,8) extended BCH codes, and also several double-error correcting BCH codes. The following question, hence, arises: do other BCH codes also contain direct-sum subcodes of high dimension? We settle this question affirmatively for *all* the primitive BCH codes, and also for the BCH codes of composite block length. The direct-sum structure is in a sense a counterpart of the concept of "zero-concurring" codewords of [1, 4], obtained by substituting a code for each codeword. We also study a different structure, where we allow the constituent codes to overlap over a fixed set of coordinates. This *concurring-sum* structure is the corresponding counterpart of the "concurring" codewords of [1]. We show the existence of concurring-sum structures in all the primitive binary BCH codes. Both the direct- and the concurring-sum structures make it possible to set nontrivial upper bounds on the number of states in the minimal proper trellis of BCH codes, and provide a clue for efficient soft-decision decoding.

Let C be a binary BCH code of length n and dimension k, let α be a primitive n^{th} root of unity, and let $\mathcal{I}$ be a subset of $\{0, 1, \ldots n-1\}$. Denote by $C[\mathcal{I}]$ the subcode of C which consists of all those codewords that are nonzero only on the positions contained in $\mathcal{I}$. Let $C(\mathcal{I})$ be the code obtained from $C[\mathcal{I}]$ by puncturing out all the positions not in $\mathcal{I}$.

Proposition 1. *Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be subsets of the set $\{0, 1, \ldots n-1\}$, such that for some $a \in \{0, 1, \ldots n-1\}$ we have $\{\alpha^i : i \in \mathcal{I}_2\} = \{\alpha^a \cdot \alpha^i : i \in \mathcal{I}_1\}$. Then $C(\mathcal{I}_1) = C(\mathcal{I}_2)$.*

Now assume that the block length of C is composite, say $n = n_1 \cdot n_2$, and let Z be the set of zero frequencies of C. Define $S = \{s \equiv z \pmod{n_1} : z \in Z\}$.

Proposition 2. *Let $\mathcal{I}_1 = \{0, n_2, 2n_2, \ldots (n_1-1)n_2\}$. Then the code $C(\mathcal{I}_1)$ is a BCH code of length n_1 and dimension $k_1 = n_1 - |S|$. The zeros of $C(\mathcal{I}_1)$ lie at $\{\beta^s : s \in S\}$, where $\beta = \alpha^{n_2}$ is a primitive n_1^{th} root of unity.*

In order to obtain direct-sum subcodes of high dimension in BCH codes of composite block length, it would now suffice to partition the set $\{0, 1, \ldots n-1\}$ into n_2 disjoint subsets satisfying the condition of Proposition 1 with respect to the set $\mathcal{I}_1$ defined in Proposition 2. Note that the sets Z and S are unions of cyclotomic cosets modulo n and n_1, respectively. Thus the definition of S in conjunction with Proposition 2 induces "coset aliasing" between the cyclotomic cosets modulo n and modulo n_1. In particular, certain high frequencies of C alias as low frequencies in $C(\mathcal{I}_1)$. This is intuitively plausible since $\mathcal{I}_1$ is just the "time-domain sampling" of C.

In the sequel we consider the primitive BCH codes. Henceforth let C denote an extended primitive narrow-sense BCH code of length $n + 1 = 2^m$.

Proposition 3. *Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be subsets of the set $\{0, 1, \ldots n-1, \infty\}$, such that for some $a \in \{0, 1, \ldots n-1\}$ we have $\{\alpha^i : i \in \mathcal{I}_2\} = \{\alpha^a + \alpha^i : i \in \mathcal{I}_1\}$. Then $C(\mathcal{I}_1) = C(\mathcal{I}_2)$.*

Proposition 3 may be thought of as the "addition counterpart" of Proposition 1. Thus we can exhibit the existence of direct-sum subcodes in the extended primitive BCH codes by partitioning the set $\{\alpha^0, \alpha^1, \ldots \alpha^{n-1}, \alpha^\infty\}$ into disjoint subsets satisfying the condition of Proposition 3 with respect to some given subset. Yet this set is just the field $GF(2^m)$. Thus it would suffice to regard $GF(2^m)$ as a vector space, and partition it into a subspace and its cosets. Notably, Proposition 3 may be also employed for the derivation of the concurring-sum structure in the primitive binary BCH codes. For more details on this see [5].

We now indicate how the direct-sum and the concurring-sum structures may be employed for efficient maximum-likelihood soft decision decoding. Let s be the logarithm of the maximum number of states in the minimal proper trellis of a linear code C. This parameter governs the complexity of maximum-likelihood decoding of C using the trellis diagrams of [2]. It follows from the trellis construction of Wolf [6], that $s \leq min\{K, N-K\}$, where N and K are the block length and the dimension of C. We employ the direct-sum and the concurring-sum structures of C to substantially improve upon this upper bound. Assume that C contains a subcode which is a direct-sum of h identical codes, each of dimension k. Then by arranging the coordinates of C in alignment with its direct-sum structure, it follows that $s \leq K-(h-1)k$. Substituting the parameters of the direct-sum structures, that we were able to find using the techniques described herein, into this expression yields upper bounds on s which are often tighter than the bound of Wolf. Arranging the coordinates of C in alignment with its concurring-sum structure also yields low values of s in all the primitive binary BCH codes. Some of the bounds on s, resulting from the direct- and concurring-sum structures, are listed in the table below. The table also lists the complexity of decoding the primitive binary BCH codes using the proposed techniques, as compared to the complexity of the the conventional decoders (Viterbi decoding based on the trellis of Wolf [6] for high-rate codes, and Fast Hadamard Transform decoding [1] for low-rate codes). These figures are given in terms of the number of real operations per bit of information. The computational gain obtained reaches several orders of magnitude in many cases. For instance for the (64,30,14) extended BCH code the proposed techniques are about 1,000 times more efficient.

Code	Wolf bound	DS and CS structures	Lower bound	Conventional decoding	Proposed techniques
BCH[8,4]	4	3	3	16	6
BCH[16,11]	5	4	4	66	26
BCH[16,7]	7	6	5	128	42
BCH[16,5]	5	4	4	32	13
BCH[32,26]	6	5	5	160	66
BCH[32,21]	11	10	9	3413	1094
BCH[32,16]	16	9	9	20480	251
BCH[32,11]	11	10	9	2048	398
BCH[32,6]	6	5	5	64	27
BCH[64,57]	7	6	6	$3.48 \cdot 10^2$	$1.32 \cdot 10^2$
BCH[64,51]	13	12	11	$1.91 \cdot 10^4$	$6.76 \cdot 10^3$
BCH[64,45]	19	14	11	$9.67 \cdot 10^5$	$2.19 \cdot 10^4$
BCH[64,39]	25	20	11	$4.04 \cdot 10^7$	$8.81 \cdot 10^5$
BCH[64,36]	28	19	15	$2.16 \cdot 10^8$	$3.77 \cdot 10^5$
BCH[64,30]	30	21	16	$6.08 \cdot 10^8$	$6.06 \cdot 10^5$
BCH[64,24]	24	16	14	$1.68 \cdot 10^7$	$1.96 \cdot 10^4$
BCH[64,18]	18	17	16	$2.62 \cdot 10^5$	$3.37 \cdot 10^4$
BCH[64,16]	16	15	14	$6.55 \cdot 10^4$	$1.16 \cdot 10^4$
BCH[64,10]	10	9	9	$1.02 \cdot 10^3$	$4.61 \cdot 10^2$
BCH[64,7]	7	6	6	$1.28 \cdot 10^2$	$5.49 \cdot 10^1$

References

[1] Y. Be'ery and J. Snyders, "Optimal soft decision block decoders based on Fast Hadamard Transform," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 355–364, 1986.

[2] G.D. Forney, Jr., "Coset Codes II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1152–1187, 1988.

[3] T. Kasami, T. Takata, T. Fujiwara, and S. Lin, "Trellis diagram construction for some BCH codes," *IEEE Int. Symp. Inform. Theory and Appl.*, Hawaii, 1990.

[4] A. Vardy and Y. Be'ery, "On the problem of finding sero-concurring codewords," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 180–187, 1991.

[5] A. Vardy and Y. Be'ery, "Maximum-likelihood soft decision decoding of BCH codes," *IEEE Trans. Inform. Theory*, submitted for publication.

[6] J.K. Wolf, "Efficient maximum-likelihood decoding of linear block codes," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 76–80, 1978.

Fast Generalized-Minimum-Distance Decoding

Ulrich K. Sorger
Inst. für Netzwerk- und Signaltheorie, Technical University of Darmstadt
Merckstraße 25, 6100 Darmstadt, Germany; email: uli@nesi.e-technik.th-darmstadt.de

Abstract

We propose a Reed-Solomon code decoding algorithm based on Newton's interpolation to speed up Generalized-Minimum-Distance (GMD) decoding. This algorithm uses a modified Berlekamp-Massey algorithm to perform all necessary GMD decoding steps in only one run. The solutions generated by a Berlekamp-Massey algorithm if i least reliable symbols are erased are used to generate the solutions for 2 erasures less. By then using a time domain decoder the overall asymptotic GMD decoding complexity becomes $\mathcal{O}(dn)$ with n the length and d the distance of the code. It can be shown that this GMD decoding complexity is asymptotically minimal.

Summary

Up to now the coding and decoding of Reed-Solomon codes is based on the Fourier transform. The approach proposed here uses Newton's interpolation. To use interpolation for *coding* was already proposed by Mandelbaum [4] back in 1979. Newton's interpolation has the advantage that if one wants to add a new interpolation value then only one additional coefficient has to be calculated. We use this for GMD *de*coding.

We assume that the distance of the Reed-Solomon code is odd (d=$2t$+1) and w.l.o.g. that the Reed-Solomon codewords over GF(q) are defined by the evaluation of polynomials of degree less or equal $n-1-2t$ with $n \leq q$. (I.e. the generator polynomial has all zeros at the highest locations.)

Let $\Psi_{2i}(x) = \prod_{j=n-2(t-i)}^{n-1}(x - z_j)$ be the erasure locator polynomial erasing the least reliable $2(t-i)$ locations. We get the key equation on erasing these locations [2]:

$$\Psi_{2i}(x)\Lambda_i(x)E(x) = (x^n - 1)\Omega_i(x) \qquad (1)$$

with $E(x)$ the transform of the error vector, $\Lambda_i(x)$ the error locator polynomial and $\Omega_i(x)$ the error value polynomial. Let $E(x)$ be given in Newton coefficients:

$$\begin{aligned} E(x) &= \mathcal{E}_0 + \sum_{i=1}^{n-1} \mathcal{E}_i \prod_{j=0}^{i-1}(x - z_j) \\ &= \bar{S}(x) + S(x) \prod_{j=0}^{n-2t-1}(x - z_j). \end{aligned}$$

I.e. $S(x) = \mathcal{S}_0 + \sum_{i=1}^{2t-1} \mathcal{S}_i \prod_{j=n-2t}^{n-1-2t+i}(x - z_i)$ and $\mathcal{S}_i = \mathcal{E}_{i+n-2t}$. Note that for a received word $R(x) = E(x) + C(x)$ and that then $S(x)$ is a (known) Newton syndrome. With

$$K_i(x) = \frac{\bar{S}(x)\Lambda_i(x)}{\prod_{j=0}^{n-2t-1}(x - z_j)} \qquad (2)$$

and as $\Psi_0(x) = (x^n - 1)/\prod_{j=0}^{n-2t-1}(x - z_j)$ the subproblems of GMD decoding become: If $2(t-i)$ least reliable locations are erased find the polynomial $\Lambda_i(x)$ of smallest degree solving

$$S(x)\Lambda_i(x) = \frac{\Psi_0(x)}{\Psi_{2i}(x)}\Omega_i(x) - K_i(x). \qquad (3)$$

with $\deg(\Lambda_i(x)) :> \deg(K_i(x))$ (see (2)). This solution is then necessarily unique up to a constant factor.

If we have solved (3) we wish to solve the next problem (i+1$\rightarrow i$) using the old solution.

However, we do not only need the minimal solution of (3) but also another second solution.

We can show the following:

Let $\Lambda_i^-(x)$, $deg(\Lambda_i^-(x))$=l be the minimal polynomial solving (3) and $\Lambda_i^+(x)$ of degree $2i-l+1$ be another solution of (3) that is not divided by $\Lambda_i^-(x)$. Then the nonzero polynomials

$$\Lambda_{i+1}^-(x) = \frac{\Psi_{2i}(x)}{\Psi_{2i+2}(x)B(x)}\Lambda_i^-(x) \qquad (4)$$

$$\Lambda_{i+1}^+(x) = A(x)\Lambda_i^-(x) - B(x)\Lambda_i^+(x) \qquad (5)$$

with

$$B(x) = \gcd\left[\Psi_{2i}(x)/\Psi_{2i+2}(x), \Omega_i^-(x)\right] \qquad (6)$$

and $A(x)$ the minimal solution of

$$A(x)\Omega_i^-(x) - B(x)\Omega_i^+(x) = \frac{\Psi_{2i}(x)}{\Psi_{2i+2}(x)}\Omega_{i+1}^+(x) \qquad (7)$$

solve (3) with $i \leftarrow i+1$. One of them is minimal with degree k. The other has degree $2(i+1)-k+1$ and it is not divided by the minimal solution.

This proves that there exists an algorithm that solves the GMD decoding problem in only one run. By transferring this algorithm into the time domain as in [2] the overall complexity of the algorithm becomes $\mathcal{O}(nt)$.

It is easy to see that this asymptotic complexity is minimal: Even if there existed an algorithm that generates the GMD list without any operation, only the operation to search for the nearest codeword would already take $\mathcal{O}(nt)$ operations. Thus is cannot be better.

References

[1] Elwyn R. Berlekamp: *Algebraic Coding Theory*, London, McGraw-Hill, 1986, pp. 176-199

[2] Richard E. Blahut: *Theory and Practice of Error Control Codes*, Addison-Wesley Publishing Company, 1984, pp. 207-245

[3] G. David Forney: *Concatenated Codes*, M.I.T. Press, Cambridge, Mass.

[4] D. M. Mandelbaum: "Construction of Error Correcting Codes by Interpolation", *IEEE Trans. Inform. Theory*, vol. IT-25, No.1, Jan. 1979

SUBOPTIMAL SOFT DECISION DECODING OF LINEAR CODES

Ilya I. Dumer
Institute for Problems of Information Transmission, Moscow
& Manchester University, El. Eng. Lab., Manchester, M13 9PL, UK

Abstract
Suboptimal decoding of linear codes in "symmetric" memoryless channels is considered. For the q-ary codes of length $n \to \infty$ and code rate R the number of decoding operations is upper bounded by the value $q^{n(c+o(1))}$, where $o(1) \to 0$ and $c = \min(R(1-R), (1-R)/2)$. The decoding error probability ϵ is upper bounded by the double error probability ϵ_e of maximum likelihood (ML) decoding, while $\epsilon \sim \epsilon_e$, when $n \to \infty$. For the channels with discrete (quantified) output the better estimate $c = R(1-R)/(1+R)$ is obtained.

Suboptimal decoding
Wolf's trellis algorithm [1] provides ML-decoding for all linear codes in memoryless channels with decoding complexity $q^{n(c+o(1))}$, where $n \to \infty$, $o(1) \to 0$ and the exponent $c = \min(R, 1-R)$. Below we consider the decoding of linear codes in "symmetric" memoryless channels with similar correcting capacity and smaller complexity. These channels include as examples the discrete symmetric memoryless channels [2], AWGN-channel or the memoryless channel with 2-dimensional wight Gaussian noise and q-PSK modulation. We consider also the complete minimum distance (MD) decoding algorithms and construct the corresponding suboptimal coverings with polynomial complexity.

Consider a channel with a discrete set X of Q inputs and an arbitrary output set Y ($|Y| \leq \infty$). Let $P_{Y|X}(y|x)$ define the probability measure for each $x \in X$. For any finite output set $Y_\alpha \subset Y$, $|Y_\alpha| = J_\alpha$, consider $(Q \times J_\alpha)$-matrix $P_\alpha = P(y|x), x \in X, y \in Y_\alpha$, using inputs as rows and outputs as columns. Following [2], the channel with an arbitrary output set Y is defined to be symmetric if Y can be partitioned into disjoint finite subsets Y_α, $Y = \cup_\alpha Y_\alpha$, in such a way that in any matrix P_α each row is a permutation of each other row and each column (if more than one) is a permutation of each other column.

For any output $y \in Y$ order all Q elements $x \in X$ into (any) set $X_y = \{x(1), x(2), \ldots, x(Q)\}$, where $P(y|x(i)) \geq P(y|x(i+1))$ for all $i = 1, \ldots, Q-1$. Let $N(x)$ denote the number of the vector $x \in X$ in the ordered set X_y. Let any subset A of $S = |A|$ inputs be used with equal probability $1/S$. For any received output $y \in Y$ let $D(y)$ be the ML-decoding solution: $D(y) = x' \in A : N(x') < N(x), \forall x \in A, x \neq x'$ For the given subset A of M inputs define the decoding algorithm D_M by the following rule:

$$D_M(y) = \begin{cases} x' & \text{if } N(x') \leq M \\ \emptyset & \text{otherwise.} \end{cases}$$

Let $\epsilon(M)$ be the error probability of D_M. Obviously, $\epsilon(Q) = \epsilon_e$ is the probability of ML-decoding. The following theorem generalizes lemma 1 [3] and gives an upper estimates on $\epsilon(M)$ for $M \geq N, N = \lceil Q/S \rceil$.

Theorem 1 *The error probability of D_M-decoding in any symmetric channel can be upper estimated for any $M = iN, i = 2, \ldots, S-1$ as $\epsilon(M) \leq \epsilon_e + \epsilon_e/(i-1)$.*

The following theorem generalizes algorithms [3], [4] for the suboptimal soft decision decoding in symmetric memoryless channels.

Theorem 2 *Virtually all q-ary linear codes of length $n \to \infty$ and code rate $R, 0 < R < 1$, can be decoded in memoryless symmetric channel with error probability, which is equivalent to the error probability of ML-decoding, and complexity $\kappa = q^{n(c+o(1))}$, where $o(1) \to 0$ and $c = \min(R(1-R), (1-R)/2)$.*

Moreover, the complexity exponents $c = R(1-R)$ and $c = (1-R)/2$ are valid for all linear cyclic codes and for all linear codes respectively. Suboptimal decoding can also decrease the complexity for the short code lengths. For example, the binary (24,12) Golay code can be decoded using the most reliable 64 trellis nodes on two information sets of the first 12 positions and the last 12 positions. Note, that the complete trellis diagram includes 4096 nodes.

Theorem 3 *Virtually all q-ary linear codes of length $n \to \infty$ and code rate $R, 0 < R < 1$, can be decoded in memoryless discrete symmetric channel with error probability, which is equivalent to the error probability of ML-decoding, and complexity $\kappa = q^{n(c+o(1))}$, where $o(1) \to 0$ and $c = R(1-R)/(1+R)$.*

Suboptimal coverings
The known information set decoding algorithms are found on suboptimal coverings in Hamming metric. These coverings can be constructed by random search [5] and provide asymptotically $\epsilon \sim \epsilon_e$, when $n \to \infty$. We construct suboptimal coverings with polynomial complexity, providing therefore the complete minimum distance decoding with the error probability ϵ_e and the same complexity exponent.

Let $S(n,t)$ be the set of vectors of Hamming weight t in F_2^n. A subset of $S(n,t)$ is called a covering $T(n,t,l)$, $t > l$, if any vector in $S(n,l)$ is covered by some vector(s) from $T(n,t,l)$. We call the covering $T(n,t,l)$ suboptimal if it has the lowest exponential order: $\log_2 |T(n,t,l)| \sim \log_2(\binom{n}{l}/\binom{t}{l})$, when $n \to \infty$. If the covering vector in $T(n,t,l)$ can be constructed in polynomial time $c(n)$ for any vector in $S(n,l)$, we call $T(n,t,l)$ a polynomial covering of complexity $c(n)$.

Theorem 4 *Suboptimal covering $T(n,t,l)$ of complexity $O(n \log_2 n)$ can be constructed, when $n \to \infty, t = \alpha n, l = \beta n, 0 < \beta < \alpha < 1$.*

Conclusion
The minimum distance decoding algorithms of the papers [3], [4] are generalized for the suboptimal decoding in an arbitrary memoryless channel, providing the new estimates of the soft decision decoding complexity.

References

[1] J.K. Wolf, "Efficient maximum likelihood decoding of linear codes using a trellis", *IEEE Trans. Inform. Theory*, vol. IT-24, pp.76-80, 1978.

[2] R.G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[3] G.S. Evseev, "On the complexity of decoding linear codes", *Problemy Peredachi Informatsii*, vol. 19, no. 1, pp.3-8, 1983.

[4] I.I. Dumer, "Two algorithms for linear codes decoding', *Problemy Peredachi Informatsii*, vol. 25, no. 1, pp.24-32, 1989.

[5] P.Erdos and J.Spencer.*Probabilistic methods in combinatorics*. Akademiai Kiado,Budapest, 1974.

An Efficient Soft Decision Decoding Algorithm For Array Codes

Xiao-Hong Peng & P.G. Farrell
Department of Electrical Engineering
University of Manchester
Manchester, UK.

ABSTRACT

A soft decision decoding algorithm for array codes, using less computation and with better performance than a previous algorithm, is introduced. The new algorithm uses received symbol hard decision and confidence values to optimise which sub-codes are selected for full soft-minimum-distance decoding, and corrects more error patterns than the previous algorithm.

SUMMARY

A number of soft decision algorithms exist which aim to reduce the number of computations required to perform Maximum Likelihood soft decision decoding (MLSDD) with as little loss in performance as possible. They are designed, according to different purposes and applications, to minimise either the symbol error rate or the codeword error rate.

A soft decision decoding algorithm for array codes is introduced in this paper. By taking advantage of the array code's characteristics, this algorithm is designed to optimise decoding of the array's sub-codes, with overall performance which is close to MLSDD under certain conditions. As an improved scheme from the original proposals [1][2], it decodes only selected sub-codes instead of all the sub-codes of an array code, and also has the ability the old algorithms lack to correct full hard errors.

Concept Of New Algorithm:

It is possible to reduce the number of soft decision computations required due to the fact that the distribution of the number of errors within one codeword is dynamic, and the number of errors is very small compared with the full code array size. There always exist some sub-codes of the code with few errors or error free. We can begin by applying full soft decision decoding on the sub-code with the fewest errors. After being decoded successful, all symbols of this sub-code are considered to have the highest confidence, and are used to decode other sub-codes. The procedure continues until all symbols have been involved rather than all sub-codes have been decoded. For the error free sub-codes, of course, soft computation is not needed.

Decoding Algorithm:

The new efficient soft decision decoding algorithm is stated as below (for simple two-coordinate parity check array codes):

1.Compute the row sub-code and column sub-code confidence sums.

2.Compute the syndrome for each sub-code using the hard decision algorithm. Sub-codes with syndrome '0' and syndrome '1' are labeled 'matched' and unmatched', respectively.

3.Rule out sub-codes with full confidence sums and the 'matched' sign, i.e., satisfying, $\left\{ \begin{array}{l} s_i = H_i b^T = 0 \\ \sum_{i=0}^{n-1} \left| \log\left[\frac{p(r/1)}{p(r/0)} \right] \right| \end{array} \right. \rightarrow \max$

These sub-codes do not need soft decision decoding, and their symbols are taken as correct.

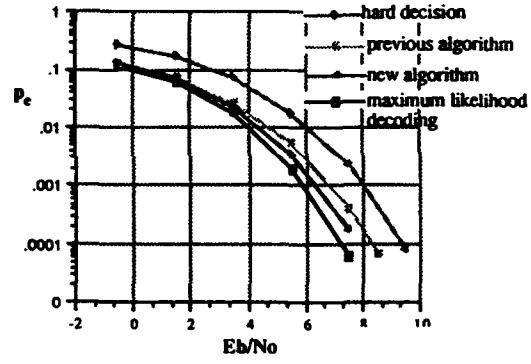
4.Decode the sub-codes, using full minimum soft distance, i.e., in the following order: high confidence +'matched' ---> high confidence +'unmatched' ---> low confidence +'matched' ---> low confidence +'unmatched'. (if both factors are same choose the one with the larger number of rows/columns)

5.After decoding each sub-code, re-calculate confidence sums of related sub-codes and change syndrome signs if necessary.

6.Repeat 3, 4, and 5 until all symbols in the array have been involved in the decoding procedure.

Performance Simulations:

The simulations assume a binary input, 8-ary output AWGN channel. The results obtained for the (16, 9) array code are shown in the following figure.



Due to the ability of being able to correct full hard errors and its other advantages, the new algorithm has a lower error rate with respect to E_b/N_o than that of the old algorithm.

With regard to decoding complexity, the new algorithm reduces the number of soft decision computations required over the old algorithm to a fraction of 0.5 or less. Alternatively, a more powerful code can be used with no increase in complexity. The new algorithm can be developed for use with more complex array codes and other types of error-control code.

REFERENCES:

[1] P.G. Farrell and S.J. Hopkins, "Decoding algorithm for a class of burst-error correcting array codes," ISIT, Les Arcs, France, 1982.

[2] J.S. Daniel, "Synthesis and decoding of array error control codes," Ph.D. Thesis, University of Manchester, 1985.

A NEW EFFICIENT ERROR-ERASURE LOCATION SCHEME IN GMD DECODING

Ralf Kötter
Dept. of Electrical Engineering, Linköping University
S-581 83 Linköping, SWEDEN

Introduction

We consider Generalized Minimum Distance (GMD) decoding, proposed by Forney in [1]. Let a received vector r and an ordering of the positions in r according to some reliability information be given. One of the key problems in GMD decoding is to decode a collection of vectors r_i obtained from r by erasing more and more positions. We solve the problem of finding error and erasure positions in different r_i efficiently by using relations between the decoding branches correcting different numbers of erasures.

Error-Erasure Location

Given a linear code C and a received vector r we want to correct t errors and ρ erasures. Formally erasure positions are set to zero. An error-erasure location for C can be done with a so called error locating pair of vector spaces (U, V) ([3],p.347). The dimension k_U of U is greater than $t + \rho$, the minimum distance d_V of V is greater than t and

$$U * V^{\perp} \subseteq C^{\perp}, \text{ with}$$

$$U * V^{\perp} = \{(u_1 v_1, u_2 v_2, \ldots, u_n v_n) : u \in U, v \in V^{\perp}\}.$$

Let G_U be a generator matrix for U, H_V a parity check matrix for V and let $diag(r)$ be the diagonal matrix containing r in its diagonal. The first step in finding an error-erasure locating vector is to find the space of solutions σ to the key equation

$$\Gamma\sigma^T = 0, \quad \sigma \in \mathbb{F}_q^{k_U} \quad \text{where } \Gamma = H_V diag(r) G_U^T.$$

Denote this space by Σ. Using Σ we find a subspace of U spanned by vectors $\sigma G_U, \sigma \in \Sigma$. We denote this subspace by W. Finally we restrict W to the space of vectors which are zero in erasure positions thus yielding the space of vectors which locate errors and erasures with zeros.

Error-Erasure Location in GMD Decoding

In the case of GMD decoding we have a collection $\{U_i, V_i\}_{i=1}^{l}$ of error erasure locating pairs correcting t_i errors and ρ_i erasures. We assume $U_i \subset U_{i+1}$ and $V_i \subset V_{i+1}$. We then can write $G_{U_{i+1}}$ containing G_{U_i} in the first k_{U_i} rows. In the same way H_{V_i} is contained in the first $n - k_{V_i}$ rows of $H_{V_{i-1}}$. The corresponding set of key equations is

$$\Gamma_i\sigma^T = 0, \quad \sigma \in \mathbb{F}_q^{k_{U_i}} \text{ where } \Gamma_i = H_{V_i} diag(r) G_{U_i}^T.$$

Each key equation has a space of solutions denoted by Σ_i. In a first step we obtain the spaces Σ_i and the corresponding spaces

$$W_i = \{u : u = \sigma G_{U_i}, \sigma \in \Sigma_i\}$$

described by generator matrices G_{W_i}. In a second step we find the subspaces of W_i which are zero in the desired erasure positions.

We define a matrix $\tilde{\Gamma}$ by

$$\tilde{\Gamma} = H_{V_1} diag(r) G_{U_l}^T$$

and denote the submatrix of $\tilde{\Gamma}$ consisting of the elements in the first a rows and first b columns as $\tilde{\Gamma}^{(a,b)}$. The following relation holds:

$$\tilde{\Gamma}^{(n-k_{V_i}, k_{U_i})} = \Gamma_i.$$

The efficiency of the proposed procedure resides from the fact that we can find the solution spaces of all key equations by dealing with only one matrix $\tilde{\Gamma}$. We apply to $\tilde{\Gamma}$ a slightly modified version of the fundamental iterative algorithm (FIA) proposed by Feng and Tzeng in [2] which gives us the spaces Σ_i. The Σ_i satisfy $\Sigma_{i-1} \subseteq \Sigma_i$ and so $W_{i-1} \subseteq W_i$. We find a generator matrix G_{W_i} containing a generator matrix for $G_{W_{i-1}}$ in the leading rows. The second task is now to find the subspaces of W_i that are zero in the desired erasure positions. This is done by applying simple row operations to the matrix G_{W_i}.

We have found an efficient procedure to obtain error-erasure positions in every branch of a GMD decoding scheme. The asymptotic complexity of this procedure is in the general case given by $\mathcal{O}(n^3)$.

References

[1] G.D. Forney Jr., *Generalized Minimum Distance Decoding*, IEEE Trans. on Information Theory, IT-12:125-131, 1966.

[2] G.L. Feng and K.K. Tzeng, *A Generalization of the Berlekamp Massey Algorithm for Multisequence Shift-Register Synthesis with Application to Decoding Cyclic Codes*, IEEE Trans. on Information Theory, 37(5):1274-1287, section III, november 1991.

[3] M.A. Tsfasman and S.G Vlădut, *Algebraic-Geometric Codes*, Dordrecht/Boston/London, Kluwer Academic Publishers, 1991.

THE GENERALIZED SYNDROME POLYNOMIAL AND ITS APPLICATION TO THE EFFICIENT DECODING OF REED-SOLOMON CODES BASED ON GMD CRITERION

Kiyomichi Araki, Masayuki Takada
Department of Electronics Eng., Saitama Univ.
Urawa, Saitama, Japan, 338

Masakatu Morii
Department of Computer Eng., Ehime Univ.
Matsuyama, Ehime, Japan, 790

In this paper, we will provide an efficient algorithm for GMD (Generalized Minimum Distance) decoding in which an algebraic errors-and-erasures decoding procedure, the W-B method, is required to execute only one time, whereas in a conventional GMD decoding at most $\lfloor d/2 \rfloor$ times algebraic decoding must be necessary. (d:minimum distance of code)

I. Generalized syndrome polynomial for errors-and-erasures

We let $R(z)$ a received word polynomial, $C(z)$ a code word polynomial, $E(z)$ an error polynomial, $\varepsilon(z)$ an erasure polynomial and $D(z)=E(z)+\varepsilon(z)$ an errata polynomial, which have the following relation.

$$R(z) = C(z)+E(z)+\varepsilon(z) = C(z)+D(z) \tag{1}$$

Here we will propose a *generalized syndrome polynomial* $S(z)$ [3], as follows;

$$S(z) = \sum_{i=0}^{n-1} R_i \alpha^{ib} \frac{T(z) - T(\alpha^i)}{z - \alpha^i} \tag{2}$$

where $T(z)$ is an *arbitrary* $(d-1)$th degree polynomial and R_i is the i-th coefficient of $R(z)$ and the generator polynomial is given by

$$G(z) = (z-\alpha^b)(z-\alpha^{b+1}) \cdots (z-\alpha^{b+d-1}) . \tag{3}$$

When $T(z)=z^{d-1}$, $S(z)$ becomes a conventional syndrome polynomial.

By (1) and (2) we get the following relation.

$$S(z) = \sum_{i\in\{D\}} D_i \alpha^{ib} \frac{T(z) - T(\alpha^i)}{z - \alpha^i} \tag{4}$$

where $\{D\}$ is a set of indices of errata location. From (4), we have the following key equation for decoding;

$$\sigma(z)S(z)=\eta(z)T(z)+\omega(z) \tag{5}$$

where $$\sigma(z)=\prod_{i\in\{D\}}(z-\alpha^i)=\prod_{i\in\{E\}}(z-\alpha^i)\prod_{i\in\{\varepsilon\}}(z-\alpha^i)=\sigma_E(z)\sigma_\varepsilon(z) \tag{6a}$$

$$\eta(z)=\sum_{i\in\{D\}} D_i\alpha^{ib}\prod_{\substack{j\in\{D\}\\ i\neq j}}(z-\alpha^j),\quad \omega(z)=-\sum_{i\in\{D\}} D_i\alpha^{ib}T(\alpha^i)\prod_{\substack{j\in\{D\}\\ i\neq j}}(z-\alpha^j) \tag{6b}$$

II. Reduced key equation

We will modify the key equation (5). At first, a set of indices $\{T\}$ and a polynomial $\sigma_T(z)$ are defined such that

$$\{T\}=\{i \mid \sigma_\varepsilon(\alpha^i)=0 \text{ and } T(\alpha^i)=0 \text{ , } i\in[0,1,\ldots,n-1]\} \tag{7}$$

$$\sigma_T(z)=\prod_{i\in\{T\}}(z-\alpha^i) \tag{8}$$

$T(z)$ and $\sigma_\varepsilon(z)$ are known polynomials for the receiver, so that $\sigma_T(z)$ is also known before the decoding process. By (5), $\omega(z)$ has also a factor of $\sigma_T(z)$, as for $\sigma(z)$ and $T(z)$, so that we have

$$\tilde{\sigma}(z)S(z)=\eta(z)\tilde{T}(z)+\tilde{\omega}(z) \tag{9}$$

where $\sigma(z)=\tilde{\sigma}(z)\sigma_T(z)$, $T(z)=\tilde{T}(z)\sigma_T(z)$, $\omega(z)=\tilde{\omega}(z)\sigma_T(z)$ (10)

We will call (9) a *reduced key equation*. When the solution of $\tilde{\sigma}(z)$ and $\tilde{\omega}(z)$ are obtained, we can also calculate the errata value as follows;

For $T(\alpha^i)\neq 0$ $$D_i=-[\tilde{\omega}(\alpha^i)/\tilde{\sigma}'(\alpha^i)] \,/\, [T(\alpha^i)\alpha^{ib}] \tag{11}$$

For $T(\alpha^i)=0$ $$E_i=[S(\alpha^i)-\tilde{\omega}'(\alpha^i)/\tilde{\sigma}'(\alpha^i)] \,/\, [T'(\alpha^i)\alpha^{ib}] \tag{12a}$$

$$\varepsilon_i=[S(\alpha^i)-\tilde{\omega}(\alpha^i)/\tilde{\sigma}(\alpha^i)] \,/\, [T'(\alpha^i)\alpha^{ib}] \tag{12b}$$

III. Application of the W-B method

We will apply the W-B method to solve (9). For the index i such that $\tilde{T}(\alpha^i)=0$, we have from (9)

$$\tilde{\sigma}(\alpha^i)S(\alpha^i)=\tilde{\omega}(\alpha^i) \tag{13}$$

The W-B method is a kind of the iterative rational interpolation method in which $\tilde{\omega}(z)$ and $\tilde{\sigma}(z)$ are numerator and denominator polynomial of rational function, respectively, as well as, α^i and $S(\alpha^i)$ corresponds prescribed sampling point and sampled value [2]. In order to stress an iterative meaning, we will write $\tilde{\sigma}^{(k)}(z)$ and $\tilde{\omega}^{(k)}(z)$ by the iterative index k, which means the k-th solutions, i.e. they satisfy

$$\tilde{\sigma}^{(k)}(\alpha^j)S(\alpha^j)=\tilde{\omega}^{(k)}(\alpha^j) \quad (j=i_0,\ldots,i_{k-1}) \tag{14}$$

where $\tilde{T}(\alpha^j)=0$ for $j\in\{i_0,\ldots,i_k,\ldots,i_{m-1}\}$ and m means the number of roots of $\tilde{T}(z)$.

IV. Efficient algorithm for GMD decoding

In the Forney's procedure (GMD decoding) [1], $(d-1)$ most unreliable received symbols, $R_{i_0}, R_{i_1}, \ldots, R_{i_{d-2}}$ are selected with decreasing reliability order, i.e., $\theta_{i_0}\geq\theta_{i_1}\geq\cdots\geq\theta_{i_{d-2}}$. Thus, by using an arbitrariness of $T(z)$, we will choose $T(z)$ such that

$$T(z)=(z-\alpha^{i_0})(z-\alpha^{i_1})\cdots(z-\alpha^{i_{d-2}}) \tag{15}$$

For the $(d-1-k)$ erasures and $\lfloor k/2 \rfloor$ errors-decoding, the erasure locations are $(i_k,\ldots,i_{d-2})$. The errata values can be calculated by (11) or (12).

Note also that $\tilde{\sigma}^{(k)}(z)$ is actually an error location polynomial with $\lfloor k/2 \rfloor$ errors and is not including an errata portion. Following Forney's procedure, we can derive an efficient algorithm for GMD decoding. This procedure is schematically shown in Fig.1.

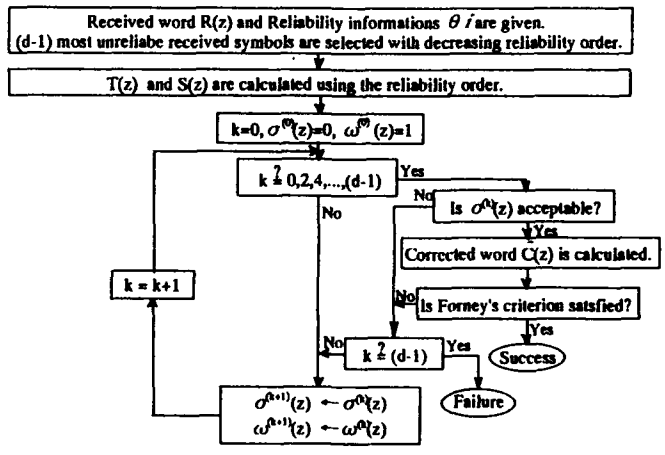


Fig.1 Flow chart of an efficient GMD decoding algorithm

V. Concluding Remarks

In our algorithm, the GMD solution can be found only by one pass using the W-B method.

References

[1] G.D.Forney, Jr., "Generalized minimum distance decoding," IEEE Trans. IT-12, pp.125-131, April 1966

[2] L.R.Welch and E.R.Berlekamp, "Error correction for algebraic block codes," presented at St. Jovites ISIT'82, 1982

[3] K.Araki and I.Fujita, "Generalized syndrome polynomials for decoding Reed-Solomon codes," IEICE Trans. Fundamentals , pp. 1026-1029, August 1992

This work was partly supported by the Telecommunications Advancement Foundation.

A Family of BCH Codes for the Lee Metric

RON M. ROTH* PAUL H. SIEGEL†

Let $C(n,r;p)$ be the (shortened) BCH code of length n over $GF(p)$ with a parity-check matrix $[\alpha_j^i]_{i=0,j=1}^{r-1,n}$, where the α_j's are distinct nonzero elements of the smallest field $GF(p^m)$ of size greater than n. The minimum Lee distance of $C(n,r;p)$ will be denoted by $d_{\mathcal{L}}(n,r;p)$.

Theorem.

$$d_{\mathcal{L}}(n,r;p) \geq \begin{cases} 2r & \text{for } r \leq (p-1)/2 \\ p & \text{for } r \geq (p+1)/2 \end{cases} .$$

Comparing the codes $C(n,r;p)$ with Berlekamp's negacyclic codes, it follows that the theorem yields the same lower bound on the minimum Lee distance as that of extended negacyclic codes; however, given p, r, and redundancy, the maximal length of the codes $C(n,r;p)$ is twice as large as that of their negacyclic counterparts. Furthermore, the decoding algorithm of $C(n,r;p)$ appears to be simpler than Berlekamp's decoding algorithm for the negacyclic case.

For fixed p and r, the codes $C(p^m-1,r;p)$ approach the sphere-packing bound on the minimum Lee distance as m tends to infinity.

When $n \leq p-1$, the codes $C(n,r;p)$ become (generalized) Reed-Solomon codes and the lower bound in the theorem can be improved to

$$d_{\mathcal{L}}(n,r;p) \geq 2r , \qquad (1)$$

which, for $r \geq \frac{6}{7}p$, can further be improved to

$$d_{\mathcal{L}}(n,r;p) \geq \frac{r+1}{2} + \frac{(r+1)^2}{4(p-1-r)} .$$

The codes $C(n,r;p)$ have an efficient decoding procedure, based upon Euclid's algorithm, that corrects all errors up to Lee weight $r-1$ and detect all errors of Lee weight r whenever the $2r$ lower bound applies. An error pattern in the Lee space is viewed as additions of $+1$'s and -1's at (not necessarily distinct) entries of the codeword. The positive (respectively negative) error-locator polynomial $\sigma^+(x)$ ($\sigma^-(x)$) is the product of terms $1-\alpha_j x$ that correspond to all location j, counting multiplicity, in which Lee errors of the $+1$-type (-1-type) have occurred. A key ingredient in the decoding algorithm of $C(n,r;p)$ is computing a polynomial $\phi(x)$ which is congruent modulo x^r to the *error-locator ratio* $\rho(x) = \sigma^+(x)/\sigma^-(x)$. The polynomial $\phi(x)$ is computed using the equality

$$\rho(x)\,S(x) + x\,\rho'(x) = 0 ,$$

taken modulo x^r, where $S(x)$ stands for the syndrome polynomial. The polynomials $\sigma^{\pm}(x)$ are then obtained by applying Euclid's algorithm on $\phi(x)$ and x^r.

One of the applications that motivated this work was analyzing the correction capability of spectral-null codes for partial-response channels [1]. These codes can be modeled as sets $C(n,r)$ of integer vectors $[c_j]_{j=1}^n$ satisfying the equalities $\sum_{j=1}^n j^i c_j = 0$ for $i = 0,1,\ldots,r-1$.

The $2r$ lower bound (1) applies also to $C(n,r)$. In particular, the bound applies to codes with an rth-order spectral null at zero frequency [1]. Furthermore, the decoding algorithm for $C(n,r;p)$ can be adapted to the codes $C(n,r)$ and thus can be used in the scheme suggested in [1] for improving the reliability of information transmission in noisy partial-response channels by matching the spectral nulls of the codes with those of the channel.

*Computer Science Department, Technion — Israel Institute of Technology, Haifa 32000, Israel.

†IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120.

References

[1] R. KARABED, P.H. SIEGEL, *Matched spectral-null codes for partial-response channels, IEEE Trans. Inform. Theory,* IT-37 (1991), 818–855.

On minimum Lee distances of generalized Reed-Muller codes

Tomoharu Shibuya, Hajime Jinushi and Kohichi Sakaniwa

Department of Electrical and Electronic Engineering,
Tokyo Institute of Technology
O-okayama, Meguro-ku, Tokyo, 152 Japan
Tel. +81-3-3726-1111(Ext. 2184), Fax. +81-3-3729-0685, E-mail. sakaniwa@ss.titech.ac.jp

1. Introduction

To meet an increasing demand for high speed communication, high density recording etc., the use of multilevel signaling has been widely considered and developed. In those systems employing multilevel signaling, it is important to develop non-binary error control codes for improving the reliability of the systems. Although binary codes based on the Hamming metric have attracted much attention of most researchers, not a lot has been investigated on non-binary error control codes.

It is well known that the Lee metric is suited to multiple-valued systems such as communication systems employing multi-phase shift keying, etc.[1]. Since it is obvious that the Lee distance is not less than the Hamming distance, in order for a non-binary code to be used as an error control code in a multiple-valued system, it is necessary for a code to have a larger minimum Lee distance than the minimum Hamming distance.

In this paper, we study the minimum Lee distances of generalized Reed-Muller (GRM) codes[2, 3]. (Complete solution is given on the minimum Hamming distance[4].)

A GRM code is defined as follows. Denote by $P_{\nu,m}$ the set of polynomials over $GF(p)$ with m variables $X_1,\ldots,X_m$ and the total degree not greater than ν. Also denote by $\boldsymbol{k}=(k_1,k_2,\ldots,k_m)$ the expression of an integer k in the number system with radix(base) p, *i.e.*,

$$k=\sum_{i=1}^{m}k_ip^{i-1},\quad 0\le k_i<p.$$

In the following, we regard k_i's $(i=1,2,\ldots,m)$ as elements of $GF(p)$.

Definition 1 For $f=f(X_1,\ldots,X_m)\in P_{\nu,m}$, let

$$c_k\triangleq f(\boldsymbol{k})=f(k_1,k_2,\ldots,k_m)\ (0\le k<p^m)$$

and define $\boldsymbol{c}_f$ and $\boldsymbol{c}_f^e$ by

$$\boldsymbol{c}_f=(c_1,c_2,\ldots,c_{p^m-1}),\quad \boldsymbol{c}_f^e=(c_0,c_1,\ldots,c_{p^m-1}).$$

Then the ν-th order generalized Reed-Muller (GRM) code C with code length p^m-1 and the ν-th order extended generalized Reed-Muller (e-GRM) code C^e with code length p^m are defined by[4, 5]

$$C=\{\boldsymbol{c}_f\mid f\in P_{\nu,m}\},\quad C^e=\left\{\boldsymbol{c}_f^e\mid f\in P_{\nu,m}\right\}.$$

2. Main Theorems

Our main results are as follows.

Theorem 1 A lower bound, $\underline{d}_{Lmin}$, of the minimum Lee distance of the ν-th order GRM code with code length p^m-1 is given by

$$\underline{d}_{Lmin}\triangleq d_{LA0}+(p^{m-1}-1)d_{LA1}\qquad(1)$$

where d_{LA0} and d_{LA1} denote the minimum Lee distance of the ν-th order GRM code with code length $p-1$ and the minimum Lee distance of the ν-th order e-GRM code with code length p, respectively. ¶

Theorem 1 implies that a lower bound of the minimum Lee distance of the ν-th order GRM code with code length p^m-1 $(m>1)$ can be obtained only from the minimum Lee distances of the corresponding ν-th order GRM code and e-GRM code with $m=1$.

We also give the true minimum Lee distances for special classes of GRM codes.

Theorem 2 The minimum Lee distance d_{Lmin} of the first order GRM code with code length p^m-1 is given by

$$d_{Lmin}=p^m-1>d_{Hmin}$$

where d_{Hmin} denotes the minimum Hamming distance of the code. ¶

Theorem 3 The minimum Lee distances of the $(p-2)$-th and $(p-1)$-th order GRM codes are given by

$$(p-2)\text{-th order}: d_{Lmin}=2p^{m-1}-1\ (=d_{Hmin})$$
$$(p-1)\text{-th order}: d_{Lmin}=p^{m-1}-1\ (=d_{Hmin})$$

and both equal their minimum Hamming distances. ¶

3. Numerical Examples

Since the true minimum distances of GRM and e-GRM codes with shorter code length can be obtained rather easily by computer search, the expression of lower bound derived in this paper enables us to get a lower bound of the minimum Lee distance of a GRM code having a longer code length.

We show in Tables 1 and 2 the lower bounds of the minimum Lee distances for GRM codes obtained by Theorem 1 together with the minimum Hamming distances for comparison. From Tables 1 and 2, we can see that there are many GRM codes, marked by †, whose minimum Lee distances really exceed the minimum Hamming distances. It is also confirmed that the lower bounds shown in Tables 1 and 2 all agree with the true minimum Lee distances, some of which, marked by *, are obtained by Theorems 2 and 3, and others by finding codewords whose Lee weights are actually equal to the lower bounds. Therefore we may conjecture that $\underline{d}_{Lmin}$(Eq.(1)) gives the true minimum Lee distance of the ν-th order GRM code, while it is a further study to give a rigorous proof.

Table 1: Lower bounds of minimum Lee distances of GRM codes for $m=2$.

order ν	p (code length)	5 (24)	7 (48)	11 (120)	13 (168)	17 (288)	19 (360)
1	$\underline{d}_{Lmin}$	*† 24	*† 48	*† 120	*† 168	*† 288	*† 360
	d_{Hmin}	19	41	109	155	271	342
2	$\underline{d}_{Lmin}$	† 18	† 48	† 120	† 168	† 288	† 360
	d_{Hmin}	14	34	98	142	254	322
3	$\underline{d}_{Lmin}$	*9	† 39	† 120	† 168	† 288	† 360
	d_{Hmin}	9	27	87	129	237	303
4	$\underline{d}_{Lmin}$	*4	† 26	† 120	† 168	† 288	† 360
	d_{Hmin}	4	20	76	116	220	284
5	$\underline{d}_{Lmin}$	—	*13	† 105	† 168	† 288	† 360
	d_{Hmin}	3	13	65	103	203	265
6	$\underline{d}_{Lmin}$	—	*6	† 85	† 151	† 288	† 360
	d_{Hmin}	2	6	54	90	186	246

Table 2: Lower bounds of minimum Lee distances of GRM codes for $m=3$.

order ν	p (code length)	5 (124)	7 (342)	11 (1330)	13 (2196)	17 (4912)	19 (6858)
1	$\underline{d}_{Lmin}$	*†124	*†342	*†1330	*†2196	*†4912	*†6858
	d_{Hmin}	99	293	1209	2027	4623	6497
2	$\underline{d}_{Lmin}$	† 98	† 342	† 1330	† 2196	† 4912	† 6858
	d_{Hmin}	74	244	1088	1858	4334	6136
3	$\underline{d}_{Lmin}$	* 49	† 291	† 1330	† 2196	† 4912	† 6858
	d_{Hmin}	49	195	967	1689	4045	5775
4	$\underline{d}_{Lmin}$	* 24	† 194	† 1330	† 2196	† 4912	† 6858
	d_{Hmin}	24	146	846	1520	3756	5414
5	$\underline{d}_{Lmin}$	—	*97	† 1205	† 2196	† 4912	† 6858
	d_{Hmin}	19	97	725	1351	3467	5053
6	$\underline{d}_{Lmin}$	—	*48	† 965	† 2023	† 4912	† 6858
	d_{Hmin}	14	48	604	1182	3178	4692

* : Also obtained by Theorem 2 or 3.
† : Exceeds the minimum Hamming distance.

References

[1] E. R. Berlekamp, *Algebraic coding theory*, McGraw-Hill, 1968.

[2] N. Mitani, "Error Detecting and Error Correcting Code", *Bulletin of the Electrotechnical Laboratory*, vol.15, no.5, pp.18–22, 1951 (in Japanese).

[3] D. E. Muller, "Application of Boolean algebra to switching circuit design and to error detection", *IRE Trans.*, EC-3, pp.6–12, 1954.

[4] T. Kasami, N. Tokura, Y. Iwadare and Y. Inagaki, *Coding Theory*, Corona Publishing Co., Ltd, 1975 (in Japanese).

[5] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, North-Holland, 1977.

A CLASS OF ERROR MAGNITUDE SUBSET CORRECTING CODES OVER GF(q)

A. Di Porto (*), F. Guida (*), E. Montolivo (*), G.M. Poscetti (**)

(*) *Fondazione Ugo Bordoni*

(**) *Università di Roma "La Sapienza"*

Most non-binary error correcting codes are designed for correcting error patterns regardless of error magnitudes and exhibit the best perfomance when error magnitudes are equally likely. However in non-binary digital transmission links with fairly good SNR only a subset of error magnitudes occurs with not negligible probability. This circumstance suggests to search for codes that are able to correct only error patterns composed of errors belonging to the said magnitude subset. Such codes will be called EMSC codes (Error Magnitude Subset Correcting codes). It is reasonable to conjecture that t-error correcting EMSC codes exist which exhibit a better rate in comparison with all possible conventional t-error correcting codes with the same codeword length.

In this paper a class of Single Error correcting EMSC codes (SE-EMSC codes) is obtained and an efficient decoding procedure is proposed. The above mentioned conjecture is proved for these codes by showing that they have a code rate greater than the value given for conventional codes by the Hamming bound.

In order to evaluate EMSC code performance an extension of the Hamming bound has been worked out as a function of the error magnitude subset cardinality. With respect to this bound *EMSC-perfect codes* have been defined and the existence of EMSC-perfect SE-EMSC codes has been proved.

A SE-EMSC code is defined by building its parity check matrix $\mathbf{H}$. We derive the matrix $\mathbf{H}^T$ of a (n', k') SE-EMSC code over GF(q) ($q=p^c$, p prime) starting from the matrix $\widehat{\mathbf{H}}^T$ of an (n, k) Hamming code over the same field with $n<n'$ and $(n-k)=(n'-k')=m$.

Let β_i, $i=1, 2,..., a$, ($a \le q-1$) be the field elements representing the a error magnitudes that the code can correct and α_j, $j=1, 2,...,h$, be h distinct elements of GF(q) such that

$$\beta_i \alpha_j \neq \beta_r \alpha_s \quad \text{for any } i \neq r \text{ and } j \neq s \tag{1}$$

For any set $\{\beta_i\}$ and $\{\alpha_j\}$ satisfying (1) with $h \ge 2$, a SE-EMSC code exists and it is defined by

$$\mathbf{H}^T = \begin{bmatrix} \alpha_1 & \widehat{\mathbf{H}}^T \\ \alpha_2 & \widehat{\mathbf{H}}^T \\ \cdots & \\ \alpha_h & \widehat{\mathbf{H}}^T \end{bmatrix} \tag{2}$$

and therefore $n'=hn$ and $k'=n'-m$.

The above defined code corrects any single-error pattern with error magnitude belonging to the set $\{\beta_i\}$ (if (1) holds, the $n'a$ syndromes of the above error patterns are distinct).

As a simple example consider the case $q=5$, $a=2$, $\beta_1=1$, $\beta_2=4$. By putting $h=2$, $\alpha_1=1$, $\alpha_2=2$ in (2), a SE-EMSC code can be derived from the (6, 4) Hamming code over GF(5) whose parity check matrix $\widehat{\mathbf{H}}$ is

$$\widehat{\mathbf{H}} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 3 & 4 \end{bmatrix} \tag{3}$$

Decoding a SE-EMSC code is only slightly more complex than decoding the Hamming code used for its construction. As (1) holds, the first non-zero element of the syndrome univocally defines the error magnitude and the n-symbol sub-block (among h sub-blocks) that contains the error. The error position in the sub-block is found in the usual way by dividing the syndrome by its first non-zero element and by looking for the coincidence against the rows of $\widehat{\mathbf{H}}^T$.

As a practical example, we considered the application of the SE-EMSC codes to a q-ary memoryless PAM channel with negligible probability that, when an error occurs, the received amplitude level is not adjacent to the transmitted one. In the said application, for $c=1$, some of these codes are EMSC-perfect, while for $c>1$ they are equivalent to codes with $c=1$. EMSC codes with $c > 1$ result attractive in case of multiple error correction or multi-dimensional signal sets.

Work carried out in the framework of the agreement between the Italian PT Administration and the Fondazione Ugo Bordoni.

A CLASS OF SINGLE ERROR CORRECTING CODES FOR CHANNELS WITH LOCALIZED ERRORS

Per Larsson
Dept. of Electrical Engineering, Linköping University
S-581 83 Linköping, SWEDEN

Introduction

A channel with localized errors is characterized by the property that possible error positions are known to the encoder but not the decoder. Equivalently we can say that the encoder knows the positions that will be error-free after transmission. The other positions are unreliable.

We construct block codes for binary channels with localized errors. Of course ordinary error correcting codes can always be used on a channel with localized errors by simply ignoring the additional information about possible error positions. Some of our codes, however, are better than any possible ordinary codes of the same lengths and error correction capabilities.

Result

Theorem *Given a shortened Hamming code of length* 2^m-1-i *and a Hamming code of length* 2^m-1, *where* $m=3,4,\ldots$ *and* $i=0,1,\ldots,2^m-1$, *then a single error-correcting code for localized errors of length* $2^{m+1}-2-i$ *and size*

$$2^{2^m-1}\left\lfloor\frac{2^{2^m-1-i}}{2^{m+1}-i}\right\rfloor$$

can be formed.

Proof:(Construction) A block of length n is divided into two subblocks of lengths n_0 and n_1, where $n_0=2^m-1-i$ and $n_1=2^m-1$. Notice that $n=2^{m+1}-2-i$. Denote by τ_0 the observed number of possible errors in the first block.

For $\tau_0=1$ we use one of A codewords from the shortened Hamming code in the first block and any vector in the second block. For $\tau_0=0$ we use one of $2^{2^m-1-i}-A(2^m-i)$ vectors in the first block (outside the decoding spheres of the A vectors used in the previous case) and one of the codewords of the Hamming code in the second one. The size of the resulting code is $\min\left\{A2^{2^m-1},\left(2^{2^m-1-i}-A(2^m-i)\right)2^{2^m-m-1}\right\}$. We choose A so that these two quantities are as close as possible. Since A must be an integer we take A equal to $\left\lfloor\frac{2^{2^m-1-i}}{2^{m+1}-i}\right\rfloor$. It is easy to check that A is less than or equal to the size of the code used in the first block as long as $i=0,1,\ldots,2^m-1$. With this choice of A the size equals $A2^{2^m-1}$. □

Summary and Conclusion

Denote by $A(n,t)$ the optimal size of an ordinary error correcting code of length n and error correction capability t. It has been proved [3] that $A(2^{m+1}-3,1)$ equals $2^{2^{m+1}-(m+1)-3}$ and that $A(2^{m+1}-4,1)$ equals $2^{2^{m+1}-(m+1)-4}$.

It is easily verified (with $i=1,2$) that our codes outperform the doubly and triply shortened Hamming codes.

The construction can be generalized in several ways. First we can find other single error correcting codes by using others than Hamming and shortened Hamming codes. Second the ideas can be used for constructing codes correcting more than one error.

References

[1] L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, *Coding for channels with localized errors,* Soviet-Swedish Workshop in Information Theory, Sweden, 1989.

[2] L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, *Coding for partially localized errors,* IEEE Trans. on Information Theory, vol.37, no.2, pp. 880-884, May 1991.

[3] M.R. Best, A.E. Brouwer, *The Triply Shortened Hamming Code is Optimal,* Discrete Mathematics 17, pp. 235-245, 1977.

[4] P. Larsson, *A class of codes correcting localized errors,* Proceedings of the International Workshop on Algebraic and Combinatorial Coding Theory, Voneshta Voda, Bulgaria, 1992.

On Perfectness of Binary Block Codes for Correcting Asymmetric Errors

G. Fang

Department of Mathematics and Computing Science,
Eindhoven University of Technology,
P.O. box 513, 5600 MB Eindhoven,
The Netherlands

Iiro S. Honkala

Department of Mathematics
University of Turku
20500 Turku 50, Finland

Abstract – Binary block codes for correcting asymmetric errors are called binary AsEC block codes. In [1], the definitions of perfect and weakly perfect binary AsEC block codes were introduced, and some properties of such codes were studied. In the present paper, we generalize these concepts and results to a larger class of AsEC codes.

Summary

A binary asymmetric error-correcting code (for short, AsEC code) C of length n and minimum asymmetric distance Δ, denoted by $C_a(n,\Delta)$, is a non-empty proper subset of $\{0,1\}^n$ in which any two distinct vectors are at asymmetric distance at least Δ apart and this distance is realized at least once. With the asymmetric distance metric, the notion of the minimum distance $r(\mathbf{c})$ from a certain codeword $\mathbf{c}$ to all other codewords was defined in [1]. $r(\mathbf{c})$ presents a kind of measurement of error-correcting capability of the codeword $\mathbf{c}$ which is better than that in terms of the minimum asymmetric distance of the code. Also in [1], with the properties of perfect codes for the binary symmetric channel in mind, natural definitions of perfect and weakly perfect binary AsEC block codes were given, which is related to the distance $r(\mathbf{c})$. Some properties of such codes were derived simultaneously there.

Since the packing spheres defined for asymmetric cases with respect to the asymmetric distance metric extend only downwards, namely only towards smaller weights, it follows that for a binary AsEC block code one sometimes could increase the sizes of those packing spheres no matter how they are with radii in terms of the minimum distance of the code or $r(\mathbf{c})$'s, such that all these improved packing spheres still remain disjoint mutually. Therefore, in the sense of error-correcting capabilities of codes, other parameters rather than the minimum distance and $r(\mathbf{c})$'s would be able to be introduced for binary AsEC block codes, and subsequently be used for the study of perfectness of such codes.

On the other hand, for the decoding of a code, one should realize that a received word $\mathbf{y}$ only comes from the codewords covering it. The strategy of a maximum likelihood decoder is of course to decode the received word $\mathbf{y}$ to one of the codewords of lowest weight covering $\mathbf{y}$. In view of the error-correcting capability of codes, one also should be aware of the two following facts. First of all, if $\mathbf{c}$ is the codeword of a $C_a(n,\Delta)$ code C of weight less than Δ, then the error-correcting capability of $\mathbf{c}$ may be referred as any number which is greater than $w(\mathbf{c})$. Hence $r(\mathbf{c})$ does not give an appropriate measure for the error-correcting capability of $\mathbf{c}$. Secondly, sometimes a codeword $\mathbf{c}$ may be able to correct more than $r(\mathbf{c})-1$ errors. Therefore, for the error-correcting capability of codes, other parameters would be able to be introduced instead of the minimum distance and $r(\mathbf{c})$'s. This motivates us to consider perfect and weakly perfect codes capable of correcting asymmetric errors in view of these new parameters, which leads to the present paper.

We will call the weakly perfect codes defined in [1] as r-WP codes. The existence of r-WP $C_a(n,\Delta)$ codes was exampled in [1]. In this paper, we introduce a different parameter $s(\mathbf{c})$ instead of $r(\mathbf{c})$. Generally, $s(\mathbf{c})$ is bigger than $r(\mathbf{c})$. By using the same definition for r-WP codes, the so called s-WP codes are defined in the present paper. We denote by $X_a(n,\Delta)$ the maximum number of codewords in a s-WP $C_a(n,\Delta)$ code, $A_a(n,\Delta)$ the maximum number of codewords in a $C_a(n,\Delta)$ code and $W_a(n,\Delta)$ the maximum number of codewords in a r-WP $C_a(n,\Delta)$ code. It can be readily verified that any r-WP $C_a(n,\Delta)$ code is a s-WP code. Thus, $W_a(n,\Delta) \le X_a(n,\Delta) \le A_a(n,\Delta)$. On the other hand, one can find examples of existence of s-WP codes which are not r-WP codes. Hence the class of s-WP codes is larger than that of r-WP codes. So, any property derived for s-WP codes can be certainly applied to r-WP codes as well. For s-WP $C_a(n,\Delta)$ codes, the following main results have been obtained in this paper: (1) A $C_a(n,\Delta)$ code C is s-perfect if and only if C is the repetition code. (2) If $n \ge 2\Delta$, then $X_a(n,\Delta) < A_a(n,\Delta)$, which also implies that if $n \ge 2\Delta$, then any nontrivial s-WP $C_a(n,\Delta)$ code cannot contain a codeword of weight greater than $n-\Delta$. Therefore, a s-WP $C_a(n,\Delta)$ code with $n \ge 2\Delta$ can always be enlarged with the all-one vector $\mathbf{1}$ to a bigger $C_a(n,\Delta)$ code.

References

[1] G. Fang, H. C. A. van Tilborg, F. W. Sun and I. Honkala, *Some Features of Binary Block Codes for Correcting Asymmetric Errors*, Submitted for publication (July 1992).

[2] T. R. N. Rao and E. Fujiwara, *Error-control coding for computer systems*. Prentice Hall Series in Computer Engineering, Prentice Hall, 1989.

Single Byte Unidirectional Error Locating Codes

Eiji FUJIWARA and Shuxin JIANG
Dept. of Computer Science, Tokyo Institute of Technology
1-12-1, O-okayama, Meguro-Ku, Tokyo 152, JAPAN

1 Introduction

Error control codes such as single-bit error correcting and double-bit error detecting (SEC-DED) codes are popularly used in computer high-speed memories [1]. This paper proposes a new type of error control code [2] which indicates only a location of unidirectional errors clustered in b bits ($b > 2$) length, called *byte*, but does not indicate accurate error bit positions in the byte. This is considered to be cost-effective due to its low redundancy. Also, this is very useful especially for diagnostic purposes in computer systems which give the information to exchange faulty packages or faulty chips. Code construction method of the single b-bit byte unidirectional error locating code, called *SbUEL code*, and its lower bound on redundancy are demonstrated in this paper.

2 Code Construction

Let $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$, where $X_i, Y_i \in \{0,1\}^b, i = 1, 2, \ldots, n$, be two distinct binary codewords each having length of n bytes, included in code C, i.e., $X, Y \in C$. In this case, every byte of $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,b})$ and $Y_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,b})$ has b-bit length, where $x_{i,j}, y_{i,j} \in \{0,1\}, i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, b$.

Definition 1 [3]: The function N is defined as

$$N(X_i, Y_i) = |\{\, j \mid x_{i,j} = 1 \wedge y_{i,j} = 0\}|,$$

where $|A|$ denotes the number of elements in the set A. Then the *unidirectional byte distance* $\mathcal{D}$ is defined as

$$\mathcal{D}(X, Y) = \sum_{i=1}^{n} \mathcal{D}(X_i, Y_i),$$

where

$$\mathcal{D}(X_i, Y_i) = \begin{cases} 2 & \text{if } N(X_i, Y_i) \neq 0 \wedge N(Y_i, X_i) \neq 0. \\ 1 & \text{if } X_i \neq Y_i \wedge (N(X_i, Y_i) = 0 \vee N(Y_i, X_i) = 0). \\ 0 & \text{if } X_i = Y_i. \end{cases}$$

□

Definition 2 [3]: *Unordered byte number* between X and Y is defined as

$$\delta(X, Y) = |\{\, i \mid \mathcal{D}(X_i, Y_i) = 2\}|.$$

□

Theorem 1 : *Code C is an SbUEL code iff any words X and Y included in C satisfy the following relation:*

$$\mathcal{D}(X, Y) \geq 3, \quad \text{or} \quad \delta(X, Y) \geq 1.$$

□

Code construction algorithm :

(1) Let the input word having k bytes be $D = (D_1, D_2, \ldots, D_k)$, where D_i, $i = 1, 2, \ldots, k$, represents the information byte with fixed length of b bits.

(2) Let $w(D)$ be a concatenation of weight of each information byte, that is,

$$w(D) = \{w(D_1), w(D_2), \ldots, w(D_k)\},$$

where $w(D_i)$ represents the weight of the information byte D_i and has value ranging from zero to b. Therefore, bit-length of $w(D_i)$, shown as b_w, is equal to $\lceil \log_2(b+1) \rceil$, where $\lceil \beta \rceil$ represents the smallest integer greater than or equal to β.

(3) The maximal linear code of the single b_w-bit byte error correcting code (Sb_wEC code)[4] is applied to encode the above defined $w(D)$. That is, multiplying $w(D)$ by the encoding parity check matrix of the maximal code of the Sb_wEC code, $\mathbf{H_E}$, i.e., $w(D) \cdot \mathbf{H_E}^T$, generates the check bytes, $CB_1, CB_2, \ldots, CB_{r+1}$, where $CB_i, i = 1, 2, \ldots, r$, is a check byte having length of b_w and CB_{r+1} is the last check byte having length g, $0 \leq g < b_w$. In this case, $\mathbf{H_E}$ of the maximal codes [4] is shown as

$$\mathbf{H_E} = \left[\begin{array}{c|c|c|c|c} & 0_{b_w} \cdots 0_{b_w} & 0_{b_w} \cdots 0_{b_w} & \cdots & 0_{b_w} \cdots 0_{b_w} \\ \hline \mathbf{H}_{R,b_w} & & 0_{b_w} \cdots 0_{b_w} & \cdots & 0_{b_w} \cdots 0_{b_w} \\ \hline & \mathbf{H}_{(R-b_w),b_w} & \mathbf{H}_{(R-2b_w),b_w} & \cdots & \vdots \\ \hline & & & \cdots & \\ \hline & & & \cdots & \mathbf{H}_{(2b_w+g),b_w} \end{array}\right]$$

$$\mathbf{H}_{R,b_w} = \left[\begin{array}{c|c|c|c} \mathbf{I}_{b_w} & \mathbf{I}_{b_w} & \cdots & \mathbf{I}_{b_w} \\ \hline \alpha^0 \cdots \alpha^{b_w-1} & \alpha^i \cdots \alpha^{i+b_w-1} & \cdots & \alpha^{2^{R-b_w}-2} \cdots \alpha^{b-2} \end{array}\right],$$

where $R = r \cdot b_w + g$, $\mathbf{I}_{b_w}$ is a $b_w \times b_w$ identity matrix, 0_{b_w} is a $b_w \times b_w$ zero matrix and α^j expresses a coefficient column vertor of $x^j \bmod g(x)$, where $g(x)$ is a primitive polynomial with degree of $(R - b_w)$.

(4) By appending the check bytes to the original input word D, the codeword of C yields to

$$[\, D \mid CB_1, CB_2, \ldots CB_r, CB_{r+1}\,].$$

□

Theorem 2 : *The set of codewords obtained from the above steps (1) to (4) is an SbUEL code.* □

3 Evaluation

Theorem 3 : *Let k be the number of information bytes with b bits/byte. Then, any code that locates single byte unidirectional errors needs at least $\lceil \log_2(k \cdot b + 1) \rceil$ check bits.* □

Figure 1 shows an example of the relation between the check bit-length and the information bit-length of the SbUEL codes when $b = 4$ bits. In this figure, the dotted line shows the lower bound on the check bit length mentioned in the Theorem 3. The broken line shows the case of a code proposed by Dunning *et al.*[5] which is originally a double byte unidirectional error detecting code for the set of weight symbols of the input word over $GF(p)$, where p is a prime larger than b, and therefore can be regarded as an SbUEL code.

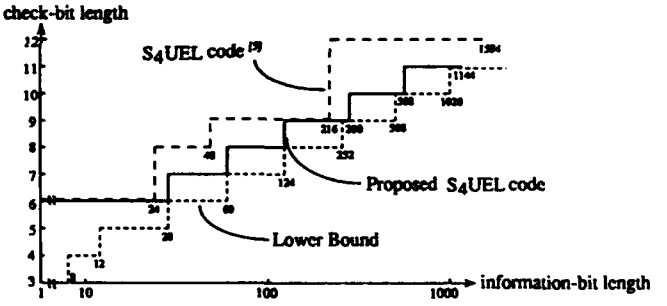


Figure 1: Check-bit length v.s. information-bit length of the S_4UEL Code

4 Conclusion

This paper has proposed the construction method of a new type of unidirectional error control code which indicates the location of single byte unidirectional errors in the received word. It has clarified the necessary and sufficient conditions for this type of code, and the lower bound on the check bit length.

If the linear code having the minimum Hamming distance d_H over $GF(2^{b_w})$ is applied to the proposed code construction method, we can get, in general, the $\frac{d_H-1}{2}$ bytes unidirectional error locating codes for an odd number d_H, and the $\frac{d_H-2}{2}$ bytes unidirectional error locating and $\frac{d_H}{2}$ bytes unidirectional error detecting codes for an even number d_H.

References

[1] T.R.N.Rao and E.Fujiwara, *Error-Control Coding for Computer Systems*, Prentice-Hall, 1989

[2] J.K.Wolf and B.Elspas, "*Error-Locating Codes – A New Concept in Error Control*", IEEE Trans. Inf. Theory, pp.113-117, Apr. 1963

[3] Y.Saitoh and H.Imai, "*All Unidirectional Byte Error Detecting Codes*", The Institute of Electronics, Information and Communication Engineers, Autumn National Convention, A-160, 1990

[4] S.J.Hong and A.M.Patel, "*A General Class of Maximal Codes for Computer Applications*", IEEE Trans. Comput. pp1322-1331, Dec.1972

[5] L.A.Dunning, G.Dial and M.R.Varanasi, "*Unidirectional Byte Error Detecting Codes for Computer Memory Systems*", IEEE Trans. Comput. Vol.39, pp.592-595, April 1990

Efficient Maximum Likelihood Decoding Algorithms for Linear Codes over Z-Channel

Tomohiko UYEMATSU

School of Information Science, Japan Advanced Institute of Science and Technology
Tatsunokuchi, Nomi-gun, Ishikawa 923-12, Japan

Abstract This paper presents three new maximum likelihood decoding (MLD) algorithms for linear codes over Z-channel, which are much more efficient than conventional exhaustive algorithms. In the proposed algorithms, their complexities are reduced by employing the projecting set C_s of the codewords, which is determined by the "projecting" structure of the code. Namely, the complexities of algorithms mainly depend upon the size of C_s which is several times smaller than the total number of codewords. It is shown that the complexities of three decoding algorithms are in proportion to the number of zeros in the received word, Hamming weight of the received word, and the number of parity bits, respectively.

1. Introduction

In the optical communication system or semiconductor memory, it is known that the communication channel can be usually modeled by Z-channel. In Z-channel, symbol '0' does not change to symbol '1', though '1' changes to '0' with probability ϵ. To make the most of the error correcting ability of codes, a maximum likelihood decoding (MLD) algorithm for cyclic codes over Z-channel has been reported[1]. However, in order to improve the performance of communication employing error correcting codes, it is important to develop efficient MLD algorithms for much wider class of codes such as linear codes. This paper presents three new MLD algorithms for linear codes over Z-channel, which are much more efficient than conventional exhaustive algorithms, and clarifies the complexities of these algorithms.

2. Preliminaries

Consider a linear binary (n, k, d) systematic code C, where n, k and d are the code length, the number of information bits and the minimum Hamming distance of C, respectively. Assume that all codewords are equally likely, and an MLD algorithm is defined as follows:

[Definition 1] (MLD algorithm) For a received word $\mathbf{r}$, an MLD algorithm chooses a codeword $\mathbf{c}$ to maximize the the conditional probability $Pr(\mathbf{r}|\mathbf{c})$ that a word $\mathbf{r}$ is received when a codeword $\mathbf{c}$ is sent.

Here, we introduce the concept of "projecting" and "projecting set" as follows[2]:

[Definition 2] (Projecting) If nonzero codewords $\mathbf{c}_1$ and $\mathbf{c}_2$ satisfies $\mathbf{c}_1 \wedge \mathbf{c}_2 = \mathbf{c}_2$, $\mathbf{c}_1$ is *projected* by $\mathbf{c}_2$, and we denote $\mathbf{c}_2 \leq \mathbf{c}_1$, where $\wedge$ denotes a bit-by-bit *and* operator.

[Definition 3] (Projecting set) The projecting set of a linear code C, denoted by C_s, is the smallest subset of nonzero codewords of C such that for any nonzero codeword $\mathbf{c} \in C$ and $\mathbf{c} \notin C_s$, there exists a $\mathbf{c}_s \in C_s$ which projects into $\mathbf{c}$.

It is reported that the actual number of codewords in C_s is at most several times smaller than the total number of codewords[2].

3. MLD Algorithms for Z-channel

By employing the projecting set C_s, we propose the following efficient MLD algorithms for Z-channel. In these algorithms, let $w(\boldsymbol{x})$ denote Hamming weight, let $w_k(\boldsymbol{x})$ denote Hamming weight of the first k bits of $\boldsymbol{x}$, and let $\oplus$ denote the modulo-2 addition.

[MLD Algorithm I] For the received word $\mathbf{r}$:

1. Find a codeword $\mathbf{c}_1 \in C$ such that $\mathbf{r} \leq \mathbf{c}_1$[1].
2. If $w(\mathbf{c}_1 \oplus \mathbf{c}_s) \geq w(\mathbf{c}_1)$ for all $\mathbf{c}_s \in C_s$ satisfying $\mathbf{c}_s \wedge \mathbf{r} = \mathbf{0}$, go to 4.
3. Find $\mathbf{c}_{s_{min}}$ which minimizes $w(\mathbf{c}_1 \oplus \mathbf{c}_s)$ among $\mathbf{c}_s \in C_s$ satisfying $\mathbf{c}_s \wedge \mathbf{r} = \mathbf{0}$. Then $\mathbf{c}_1 \longleftarrow \mathbf{c}_1 \oplus \mathbf{c}_{s_{min}}$ and go to 2.
4. $\mathbf{c}_1$ is the desired codeword. □

[MLD Algorithm II] For the received word $\mathbf{r} = (r_1, r_2, \cdots, r_n)$:

1. Let $i = 1$, $\mathbf{c}_1 = \mathbf{0}$ and $\mathbf{r}_j = (r_1, r_2, \cdots, r_j, 0, \cdots, 0)$ $(j = 1, 2, \cdots, n)$
2. If $r_i \not\leq \mathbf{c}_1$, find $\mathbf{c}_{s_{min}}$ which minimizes $w(\mathbf{c}_1 \oplus \mathbf{c}_s)$ among $\mathbf{c}_s \in C_s$ satisfying $\mathbf{r}_i \leq (\mathbf{c}_1 \oplus \mathbf{c}_s)$. Then $\mathbf{c}_1 \longleftarrow \mathbf{c}_1 \oplus \mathbf{c}_{s_{min}}$.
3. If $i < n$ then $i \longleftarrow i + 1$ and go to 2.
4. $\mathbf{c}_1$ is the desired codeword. □

[MLD Algorithm III] For the received word $\mathbf{r} = (r_1, r_2, \cdots, r_n)$:

1. Let $i = 1$ and $\mathbf{r}_j = (r_1, r_2, \cdots, r_{k+j}, 0, \cdots, 0)$ $(j = 1, 2, \cdots, n-k)$. Let $\mathbf{c}_1$ be a codeword specified by
$$\mathbf{c}_1 = (c_{1,1}, c_{1,2}, \cdots, c_{1,n}) = (r_1, r_2, \cdots, r_k)G,$$
where G is the $k \times n$ generator matrix of the code C.
2. If $c_{1,k+i} \oplus r_{k+i} = 1$, find $\mathbf{c}_{s_{min}}$ which minimizes $w_{k+i}(\mathbf{c}_1 \oplus \mathbf{c}_s)$ among $\mathbf{c}_s \in C_s \cup \{\mathbf{0}\}$ satisfying $\mathbf{r}_i \leq (\mathbf{c}_1 \oplus \mathbf{c}_s)$. Then $\mathbf{c}_1 \longleftarrow \mathbf{c}_1 \oplus \mathbf{c}_{s_{min}}$.
3. If $i < n - k$ then $i \longleftarrow i + 1$ and go to 2.
4. $\mathbf{c}_1$ is the desired codeword. □

In these algorithms, their complexities are mainly depend upon the size of C_s. Since the size of projecting set $|C_s|$ is several times smaller than the total number of codewords (2^k), both computational and space complexities for these algorithms are significantly reduced compared with conventional exhaustive algorithms. The upper-bound of the complexities of the algorithms are shown in the Table 1. It should be noted that the combination of Algorithm I and II yields an MLD algorithm with maximum number of comparisons $\min\{w(\mathbf{r}), n - w(\mathbf{r})\}|C_s|$.

4. Conclusion

This paper proposes some efficient MLD algorithms for linear codes over Z-channel, and clarifies their complexities.

References

[1] H. Inaba, M. Morii, and M. Kasahara: "Notes on Fast Maximum-Likelihood Decoding-Algorithm of Cyclic Code on Z-channel", Trans. IEICE, vol.J74-B-I, no.10, pp.769-777, Oct. 1991 (in Japanese).

[2] T. Y. Hwang, "Decoding Linear Block Codes for Minimizing Word Error Rate", IEEE Trans. on Inform. Theory, vol.IT-25, no.6, pp.733-737, Dec. 1979.

Table 1. Complexities of the proposed algorithms

Algorithm	Number of codewords stored	Maximum number of comparisons
I	$\|C_s\|$	$(n - w(\mathbf{r}))\|C_s\|$
II	$\|C_s\|$	$w(\mathbf{r})\|C_s\|$
III	$\|C_s\| + k$	$(n-k)\|C_s\|$
Exhaustive	2^k or k	2^k

[1] If $(1, 1, \cdots, 1)$ is a codeword of C, this codeword always satisfies the condition.

Reduced State Sequence Detection for Asynchronous Gaussian Multiple-Access Channels

MAHESH K. VARANASI

ECE Department, University of Colorado, Boulder, CO 80309

Abstract: The problem of coherent multiuser detection is considered for the K-user asynchronous Gaussian Code-Division Multiple-Access (CDMA) channel. The maximum likelihood sequence detector (MLSD) is asymptotically optimal in that it achieves the highest error exponent of the bit error probability for each user. However, the MLSD can only be implemented by a dynamic programming algorithm whose complexity depends exponentially on K. In order to mitigate the complexity of this scheme, a class of group detection strategies is derived based on optimal statistical inferential procedures. Each member of this class of detectors corresponds to a L group partition of the K users, and consists of a bank of L group detectors, one for demodulating the information symbols of users in each group. Each group detector is a reduced state sequence detector with the dominant complexity determined by the computation of the solution of a combinatorial optimization problem via a forward dynamic programming algorithm. This algorithm has a complexity that is exponential in the number of users in the corresponding group. The overall complexity is determined by the size of the largest group which is a design parameter that can be chosen to be only as large as complexity considerations allow. The performance analysis of the group detection scheme is obtained by deriving asymptotically tight upper and lower bounds on the bit error probability, thereby characterizing its multiuser asymptotic efficiency.

Summary

The idea of group detection was introduced by the author in [2] in the context of multiuser detection over a QAM synchronous Gaussian CDMA channel. The synchronous channel is memoryless and therefore single-shot decisions can be optimal. However, the more general asynchronous problem is inherently a sequence detection problem and new problems arise in generalizing the results in [2].

It was shown in [2] that for an arbitrary L group partition $\bigcup_{l=1}^{L} G_l = \{1,\ldots,K\}$ of the set of K active users in a M-ary QAM synchronous CDMA channel, a generalized likelihood ratio test-based group detection scheme can be implemented in parallel as a bank of L group detectors, one for each group in the partition. The l^{th} group detector jointly demodulates the users in the group G_l and the time complexity per symbol (TCS) of the l^{th} group detector is $O(M^{|G_l|}/|G_l|)$ for M-ary QAM alphabets. From complexity considerations alone, the trivial single-user partition $\bigcup_{l=1}^{K} \{l\} = \{1,\ldots,K\}$ is the most desirable and yields the decorrelating detection scheme with a complexity that is independent of K. Performance considerations, however, tell a different story. A key result in [2] establishes that a group G detector is optimally group near-far resistant in the sense that, for each user in G, it achieves the highest achievable worst-case asymptotic efficency over the signal amplitudes of users not in G. As a consequence, viewed from the performance viewpoint alone, membership of a given user in a larger group is preferred over that in a smaller group contained by it. A larger group size, however, brings with it a higher complexity.

A vector space interpretation of the group detector for the synchronous channel involves the direct sum decomposition of the space spanned by the signature signals of all the users into two subspaces, one of which is spanned by the signals of the users in the group to be demodulated and the other by the rest of the signals. The complexity of the group detector is due to the computation of orthogonal projections of certain transformations of the outputs of a bank of matched filters (matched to the orthonormal bases of the K dimensional signal space) onto the perp space of the subspace spanned by the users not in the group G_l. In generalizing this approach to the asynchronous channel, the two subspaces in the direct sum decomposition generalize to those spanned by all time-shifted (by integer multiples of symbol durations) versions of signature signals of users belonging to the group under consideration and those that do not belong to this group. The number of orthogonal projections that need to be computed in this case is $M^{N|G_l|}$ (N is the packet length!) and it can be shown that there is no solution with a complexity that is independent of the packet length. A key result of this paper is the derivation of an alternative oblique projections-based group detection strategy where the $M^{N|G_l|}$ oblique projections can be computed by a forward dymanic programming algorithm whose complexity is independent of the packet length and depends exponentially only on the number of users in the group that it demodulates. Since the group size is a design parameter, it can be chosen to be only as large as complexity considerations allow.

It was seen in [2] that the performance analysis of the group detection scheme for the synchronous channel could be deduced from a result on the equivalence of a group-G detector with a maximum likelihood detector in a fictitious $|G|$-user synchronous Gaussian CDMA channel. However, this equivalence doesn't hold for the oblique projections-based group detector for the asynchronous channel. In fact, it is shown that the orthogonal projections-based group detector of [Var92] when generalized to the asynchronous channel, though not practically implementable, has an asymptotic efficiency performance that is an upper bound on the performance of the oblique projections-based group detection scheme. The second key result of this paper is the derivation of asymptotically tight upper and lower bounds on the bit error probability of the proposed group detection scheme for the asynchronous channel thereby characterizing its asymptotic efficiency.

The design and analysis of the reduced state group detection scheme obtained in this work provides a unifying treatment of the multiuser detection problem in the sense that two detectors corresponding to two trivial partitions result in previously proposed schemes. The case of a partition of users into one large group of size K yields the MLSD obtained in [3] with the highest possible asymptotic efficiency for each user, but at the price of an exponential complexity in K. The other extreme case of a partition that consists of K groups, each of size one, results in a group detection scheme which reduces to the decorrelating detector [1]. This detector requires only a K-input K-output digital filter following the bank of matched filters. All other partitions yield new detection schemes and the interplay between the complexity and the performance of these schemes will be presented.

References

[1] R. Lupas and S. Verdu. Near-far resistance of multiuser detectors in asynchronous channels. *IEEE Trans. Commun.*, COM-38:496–508, April 1990.

[2] M. K. Varanasi. Group detection in QAM synchronous CDMA channels. *Proceedings of the 1992 Conference on Information Sciences and Systems*, 2:820–825, 1992.

[3] S. Verdu. Minimum probability of error for asynchronous Gaussian multiple-access channels. *IEEE Trans. on Info. Theory*, IT-32:85–96, January 1986.

ERROR PROBABILITIES FOR FIBER-OPTIC CODE DIVISION MULTIPLE ACCESS SYSTEMS

Narayan B. Mandayam and Behnaam Aazhang
Department of Electrical and Computer Engineering
Rice University
Houston, Texas 77251-1892

Abstract

Performance analyses of fiber-optic code division multiple access (FO-CDMA) systems are intractable and often, Monte Carlo simulations that yield realistic estimates of system performance require a large number of simulation trials for the estimates to be in a reasonable interval of confidence. We develop an Importance Sampling technique to estimate the performance of direct detection FO-CDMA systems, where the "gain" of Importance Sampling over Monte Carlo simulations is shown to increase linearly with the system performance. The quick simulation technique developed extends to avalanche photodetection and is also compatible with a wide variety of coding schemes. Using these efficient simulations, we present a comparitive analysis of systems employing optical-orthogonal-codes and prime-sequences, where only 50-100 trials are required for estimating error probabilities of 10^{-7} and below. Based on an inexact Fourier expansion of the Poisson complimentary probablity distribution function, we derive approximations for the probability of error that require computations increasing linearly with the number of users as opposed to an exponential increase in the case of exact evaluation of the error probability. The inaccuracy of the results are shown to be bounded.

System Description : A FO-CDMA system is considered where the information bit of each user is modulated onto the intensity of the laser transmitted through a single-mode fiber channel. If user k is sending bit i, under hypothesis H_i, then the intensity of the modulated light is given as

$$\lambda_i^{(k)}(t) = \sum_{n=1}^{N} \lambda_i^{(k)}(n)\Pi_{T_c}(t - nT_c),\ i = 0, 1;\ \text{for } t \in [0, T) \qquad (1)$$

and $\Pi_{T_c}(t)$ is a unit rectangular pulse of duration T_c, and $\underline{\lambda}_i^{(k)} = [\lambda_i^{(k)}(1), \cdots, \lambda_i^{(k)}(N)]$ is a signature sequence of length $N = T/T_c$ with each $\lambda_i^{(k)}(n) \in \{0, 1\}$. At the receiving end, this gives rise to the following two hypotheses at receiver of the desired user (taken to be user 1) in the time interval $[0, T)$ as $H_i : \bar{\lambda}_i^{(1)}(t) = \lambda_i^{(1)}(t) + \sum_{k=2}^{K} \lambda_b^{(k)}(t)$, where the symbol b denotes the information of the k^{th} user. The receiver corresponding to user 1 has a replica of the signature sequence assigned to this user, and the light in the channel is correlated with this replicated signature sequence. The correlated intensities are incident on an ideal photodiode and the resulting photoelectron count is compared to a threshold for data recovery. Without loss of generality, we assume that each user is employing on-off keying, and hence at the 1^{st} receiver, the intensities are correlated with $\underline{\lambda}_1^{(1)}$, since $\underline{\lambda}_0^{(1)} = \underline{0}$. For an $\{N, J, \rho_a, \rho_c\}$ optical code sequence (i.e., a sequence of length N, weight J, and, auto and crosscorrelation constraints ρ_a and ρ_c respectively) the probability of r photoelectrons occurring, under H_i, in the sampling interval T is given as

$$p_{R|H_i}(r) = \sum_{k=0}^{\rho_c(K-1)} p_{\mathcal{I}^{(1)}}(k) p_{R|\bar{\nu}_i, H_i}(r \mid \nu_i),\ i = 0, 1, \qquad (2)$$

where $p_{R|\bar{\nu}_i, H_i}(r \mid \nu_i)$ is the conditional photoelectron probability and $p_{\mathcal{I}^{(1)}}(k)$ is the probability that the multiple access interference term $\mathcal{I}^{(1)} (= \sum_{k=2}^{K} I_k^{(1)})$ takes on the value k. The probability of error can be computed for equiprobable hypothesis as

$$P_e = \frac{1}{2}[\sum_{i=0}^{1} \sum_{\Delta_{1-i}} p_{R|H_i}(r)], \qquad (3)$$

where Δ_i is the decision region for H_i, $i = 0, 1$.

Importance Sampling : We obtain the Importance Sampling estimator by rewriting the error rate in (3) as

$$P_e = \frac{1}{2}[\sum_{i=0}^{1} \sum_{\Delta_{1-i}} p_{R^*|H_i}(r \mid H_i) w(r \mid H_i)], \qquad (4)$$

where $w(r \mid H_i) = \frac{p_{R|H_i}}{p_{R^*|H_i}}$ are the weights under H_i. The "gain" of Importance Sampling over Monte Carlo simulations is given by $\Gamma = \frac{M_{MC}}{M_{IS}}$, where M_{MC} and M_{IS} are the number of trials under the respective methods. The sufficient conditions for achieving a realistic "gain" reduce to $p_{R^*|H_i}(r \mid H_i) > p_{R|H_i}(r \mid H_i)$, $\forall r \in \Delta_{1-i}$. Since the optimum solution to maximizing Γ yields a degenerate biasing density [1], we look for a suboptimal solution satisfying the sufficient conditions as given above. To make the problem of determining the suboptimal biasing density tractable, we choose not to bias the multiple access interference parameters [1] and look for biasing densities of the form

$$p_{R^*|H_i}(r) = \sum_{k=0}^{\rho_c(K-1)} p_{\mathcal{I}^{(1)}}(k) p_{R^*|\bar{\nu}_i, H_i}(r \mid \nu_i),\ i = 0, 1. \qquad (5)$$

Further, we show that when an exponential change of measure yields the biasing density (i.e., $p_{R^*}(r) = e^{sr} \frac{p_R(r)}{M_R(s)}$), the sufficient conditions reduce to solving the following minmax problem [1]:

$$\min_{\nu_i^*} \max_{r \in \Delta_{1-i}} \left\{[-(iJ + \nu_d N + k) + \nu_i^* + k] + r \log[\frac{iJ + \nu_d N + k}{\nu_i^* + k}]\right\},$$

where $k \in [0, \rho_c(K-1)]$ and $p_{R^*|H_i}$ is parametrized by the parameter ν_i^*.

Approximations : In the evaluation of the analytical probability of error in equation (3), we need to compute the expectations in (2) over all the information paths of the process. In general, the distributions of the sum of the interfering intensities do not have closed form expressions, and since each $I_k^{(1)} \in \{0, 1, \cdots, \rho_c\}$, roughly $(\rho_c + 1)^{K-1}$ computations are required. If we decompose the photoelectron count at the output of the photodetector as $R = R_s + R_{\mathcal{I}}$, (where R_s and $R_{\mathcal{I}}$ are contributions due to the user intensity and the interference intensities, respectively) then we can write $\sum_{\Delta_i} p_{R|H_0}(r \mid H_0) = \sum_{r_{\mathcal{I}}} p_{R_{\mathcal{I}}}(r_{\mathcal{I}}) Q_{R_s}(\gamma - r_{\mathcal{I}})$, where Q_{R_s} is the complimentary cumulative distribution function of the Poisson random variable with mean ν_d, γ the detector threshold, and $p_{R_{\mathcal{I}}}$ is the probability mass function of the photoelectron count due to interfering intensities. By representing Q_{R_s} in terms of an approximate Fourier series as $Q_{R_s}(x) = \sum_{m=-\infty}^{\infty} c_m e^{jm\omega x} + \epsilon(x)$, where ω is the angular frequency term and $\epsilon(x)$ an error term, we can write $\sum_{\Delta_i} p_{R|H_0}(r \mid H_0) = \sum_{m=-\infty}^{\infty} c_m e^{jm\omega\gamma} (\prod_{k=2}^{K} \Phi_{R_k}(-m\omega)) + \delta$, where $R_{\mathcal{I}}$ has been decomposed into the sum of $K-1$ independent random variables, i.e., $R_{\mathcal{I}} = \sum_{k=2}^{K} R_k$ and $\delta = \sum_{r_{\mathcal{I}}} p_{R_{\mathcal{I}}} \epsilon(\gamma - r_{\mathcal{I}})$. The moment generating functions Φ_{R_k} can be evaluated from knowing the probabilities $P_{I_k^{(1)}}$. If we can truncate the above series to M terms, we see that the computations required are equal to $\rho_c(K-1)M$. Thus we have reduced the number of computations to be linear in K as opposed to being exponential in K.

REFERENCES

[1] N. B. Mandayam and B. Aazhang, "Importance Sampling for Direct-Detection Optical Communication Systems," *IEEE Trans. Commun*, To Appear.

This work is supported by the Advanced Technology Program of the Texas Higher Education Coordinating Board under Grant 003604-018

Performance Analysis of Optimum Demodulation in Optical CDMA *

Laurie B. Nelson and H. Vincent Poor
Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA

ABSTRACT

Performance results for optimum demodulation of optical code division multiple access (CDMA) signals are obtained. Upper and lower bounds on minimum probablity of symbol error for a K-user, symbol-asynchronous optical CDMA system with optical orthogonal codes (OOCs) are derived and evaluated. An asymptotic efficiency defined relative to the performance in known interference is introduced. The results obtained exhibit the asymptotic efficiency of optimum demodulation and suggest the existence of a significant performance gap between the optimum receiver and the conventional correlation receiver even in mild near-far environments.

Overview

In this paper, we consider the performance analysis of optical CDMA communications. Such formats are of interest in several emerging applications, including indoor wireless communications and all-optical processor interconnects. Because of its low complexity, the conventional correlation receiver has been the focus of much work. It is well known that conventional receiver performance suffers when the signals of different users are received with unequal energies. In this situation, the performance of optimum demodulation is of special interest.

In direct sequence optical CDMA, each user k is assigned signature sequences $\underline{a}_k^{(0)}, \underline{a}_k^{(1)} \in \{0,1\}^J$, which effectively divide the symbol interval into J "chips". User k signals a symbol b_k by transmitting an optical pulse in each of the chip intervals corresponding to a "1" in the signature sequence $\underline{a}_k^{(b_k)}$. The K signals are combined non-coherently on the channel, which may be free-space or guided.

Demodulation is based on knowledge of the transmitter delays, energies, and signature sequences and on direct detection of the received signal over each chip interval. The observations may be modelled as conditionally Poisson random variables with rates given by

$$r_{l,j} = \sum_{k:\tau_k<j} \lambda_k a_{k,j-\tau_k}^{(b_k(l))} + \sum_{k:\tau_k\geq j} \lambda_k a_{k,J+j-\tau_k}^{(b_k(l-1))} + \lambda_d,$$

for the observation over the jth chip of the lth symbol of the desired user, user 1. The integer τ_k represents a chip-synchronous transmitter delay relative to τ_1, λ_k corresponds to energy of user k, and λ_d represents photodetection dark current.

The error probability analysis of this paper avoids the unrealistic assumptions of symbol-synchronous transmission and random codes which previous analyses have required [1]. The performance measures considered include the single-user lower bound, the known-interference lower bound (achieved by the likelihood-ratio test when b_k for $k \neq 1$ are known), the Chernoff upper bound, and the performance of a modified conventional detector [2], which ignores observations from those chips during which interfering users are transmitting. The correlation receiver performance in known interference is also evaluated. The use of OOCs [3] with maximum cross-correlations equal to 1 allows the optimum decision for $b_1(l)$ to be made symbol-by-symbol. We also employ saddle-point approximations [4], which expedite numerical analysis and provide *exceptionally* good approximations.

Numerical results are presented in Figure 1 for a 4-user system in which $\lambda_1 = \lambda_2 = \lambda_3$. Curves are plotted versus the near-far ratio (NFR), defined here as λ_4/λ_1. The OOCs utilized have weight equal to 4 and length equal to 73. Even when all users have equal energies (NFR = 0 dB), optimum performance is still more than two orders of magnitude better than conventional receiver performance.

In [2] an *asymptotic multiuser efficiency* for the optical CDMA channel is defined relative to single-user performance. Here we define an efficiency relative to performance in known-interference. (Unlike the analagous situation for radio-frequency channels, *known* interference is not equivalent to *no* interference.) Let $P_{opt}(\lambda_1)$ denote the optimum error probability for user 1 and $P_{ki}(\lambda_1)$ the known-interference lower bound. For a given error probability $P_e = P_{opt}(\lambda_1)$, let $\tilde{\lambda}_1$ represent the energy required for user 1 to achieve P_e in known interference ($P_{ki}(\tilde{\lambda}_1) = P_e$). Then the asymptotic relative efficiency is given by

$$\eta \triangleq \lim_{\lambda_1\to\infty} \frac{\tilde{\lambda}_1}{\lambda_1} = \lim_{\lambda_1\to\infty} \frac{P_{ki}^{(-1)} \circ P_{opt}(\lambda_1)}{\lambda_1},$$

where the energies of the interferers are held proportional to the energy of user 1, i.e. $\lambda_k = \epsilon_k \lambda_1$, $\forall k \geq 2$.

In Figure 2, the asymptotic efficiency of the optimum receiver for a 2-user synchronous system with arbitrary $\{0,1\}$ signature sequences and $\lambda_d = 0$ is plotted versus ϵ_2, the *near-far ratio*, for various values of r, where $1-r$ represents the fraction of symbol energy transmitted during periods of isolated transmission.

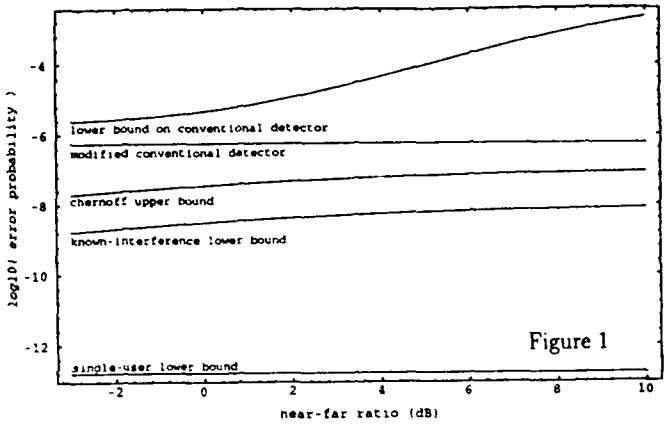


Figure 1

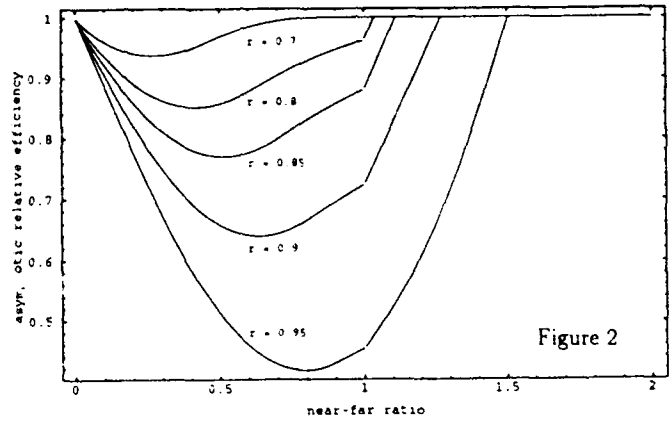


Figure 2

References

[1] M. Brandt-Pearce and B. Aazhang, "Unequal Received Power Effects on Single-User and Multi-User Detection of Optical CDMA," in *Proc. 1992 CISS*, Princeton University, Mar. 1992.

[2] D. Brady, "Asymptotic Multiuser Efficiency for Optical Channels," in *Proc. 1991 CISS*, Johns Hopkins University, Mar. 1991.

[3] J. A. Salehi, "Code Division Multiple-Access Techniques in Optical Fiber Networks—Part I: Fundamental Principles," *IEEE Trans. Comm.*, Vol. 37, No. 8, pp. 824–833, Aug. 1989.

[4] C. W. Helstrom, "Computing the Performance of Optical Receivers with Avalanche Diode Detectors," *IEEE Trans. Comm.*, Vol. 36, No. 1, pp. 61–66, Jan. 1988.

*This work was supported by the U.S. National Science Foundation under Grants NCR-90-02767 and EID-90-19951.

A LINEAR ADAPTIVE FRACTIONALLY SPACED SINGLE USER RECEIVER FOR ASYNCHRONOUS CDMA SYSTEMS

Predrag B. Rapajic and Branka S. Vucetic

Department of Electrical Engineering
The University of Sydney, Sydney, NSW 2006, Australia

Abstract

A fully asynchronous single user receiver in a code-division multiple-access (CDMA) system is considered. It is assumed that the receiver has no knowledge of the signature waveforms or timing information of other users. The receiver is trained by a known training sequence prior to data transmission, and continuously adjusted by an adaptive algorithm during data transmission. An adaptive, fractionally spaced least mean square (LMS) filter is employed for each user separately, instead of matched filters with constant coefficients. The proposed receiver is as simple as a standard single user detector receiver but it achieves essential advantages with respect to timing recovery, multiple-access interference elimination, narrowband interference suppression and user privacy. In comparison to the centralized linear multi-user detectior it has the same bit error performance while the computation complexity is substantially lower and independent of the number of users. The receiver structure is investigated and tested by simulation using a set of Gold sequences of length 31. Experimental results show that a considerable improvement in bit error rate is achieved with respect to the conventional single-user receiver.

1 Introduction

Several approaches to the CDMA demodulation problem have been considered so far. The conventional approach consists in demodulating each signal using a single user detector with a matched filter, thereby ignoring the multiple access interference (MAI) caused by cross-correlation between signals of different users [3]. This approach has two major shortcomings: (1) high sensitivity to the near-far effect, and (2) the channel capacity being interference limited, instead of being limited by the AWGN level. On the other hand this approach has the advantage of being very simple to implement.

An alternative receiver structure is a maximum likelihood multi-user demodulator for synchronous and asynchronous transmission. The maximum likelihood multi user receiver consists of a bank of matched filters followed by a Viterbi maximum likelihood detector [1]. The computational complexity of the optimum demodulator increases exponentially with the number of users.

In a number of papers a less complex class of suboptimal centralised linear multi user detectors (CLMD) is proposed where the computational complexity of the receiver increases linearly with the number of users [2]. The receiver is "near-far resistant" eliminating the need for strict power control. The drawback of this approach is that the parameters of all users including signatures, timings and carrier phases have to be known. The accuracy in estimation of these parameters strongly influences the single user detection process and instability can spread to other users making whole system unstable. The CLMD is considerably complex relative to the conventional single user detector. The CLMD proved that the capacity limitation of the CDMA system by MAI is consequence of the conventional single user approach rather then the inherent property of the CDMA system. Moreover, it proved that this limitation can be overcome by a linear receiver.

In this paper we consider a single user detector approach. A single adaptive minimum mean square error (MMSE) filter assigned to each user eliminates interference from other users to the same extent as it does linear multi user detector. However, timing, signatures or carrier phase information from other users are not needed. Receivers perform independently making the system more stable and suitable for adaptive implementation. An adaptive filter is necessary to handle time varying system parameters. It is important to note that in contrast to the centralised multi user receiver the observation vector is not the output from the bank of matched filters, but the sampled signal itself. Another important feature of the proposed receiver is the use of a fractionally spaced filter which is insensitive to the time differences in the signal arrival times of different users. Thus, the receiver timing recovery is extremely simplified (if necessary at all) [4].

References

[1] S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," IEEE Trans. Inform. Theory, vol. IT-32, pp. 85-96, Jan. 1986.

[2] R. Lupas and S. Verdu, "Near-far resistance of multi-user detectors in asynchronous channels," IEEE Trans. Commun., vol. COM-38, pp. 496-508, Apr. 1990.

[3] M. B. Pursley, D. V. Sarwate and W. E. Stark, "Error probability for direct-sequence spread-spectrum multiple-acces communications, Part I: Upper and lower bounds," IEEE Trans. Commun., vol. COM-30, pp. 975-984, May 1982.

[4] R. D. Gitlin and H. C. Meadors, Jr. "Center-tap tracking algorithms for timing recovery, " Bell Syst. Tech. J., vol. 66, no. 6, pp. 73-78, Nov. 1987.

FADING RESISTANT MULTIUSER DETECTION FOR CDMA COMMUNICATIONS

Subramanian Vasudevan AND Mahesh K. Varanasi

Department of Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309

Coherent detection of asynchronous Code-Division Multiple-Access (CDMA) data transmissions over a Rician fading channel is considered in the context of a multipoint-to-point communication system, where a centralised receiver that is assumed to have knowledge of the signature signals of the system users, including the arrival times of the former, observes a superposition of the specular and faded signal components of each of the users in additive noise. The channel itself is assumed to be non-dispersive, and with no fading memory, i.e., the random attenuations and phase shifts experienced by the users' transmissions over different bit intervals are assumed to be independent of each other, rendering them inestimable.

It turns out that this channel is equivalent to a fictitious CDMA-AWGN (Additive White Gaussian Noise) channel from the point of view of optimal detection [1]; it may be shown that the optimum decision rules over the two channels as well as the statistical characterization of the sufficient statistics in each case parallel each other. Unlike the optimum AWGN multiuser detector however, the optimum faded detector is unimplementable using a Viterbi algorithm. This motivates the derivation of the polynomial-complexity faded decorrelator for this channel.

Detector performance in a multiuser faded environment may be adversely affected by both the interferer specular (or known) and faded (or unestimatable) signal components. The performance limiting effect of the former on conventional detection as well as the ability of fading channel strategies to withstand such interference has been studied earlier [1]. The issue of the limitations on detector performance, if any, due to interfering fading is addressed here. To this end, we introduce the fading susceptance and fading resistance measures; the former as a measure of whether degradations in detector performance due to interfering fading are so great so as to prevent them from being competetive with optimum detection strategies over single-user channels, and the latter as a measure that captures the ability of detectors to withstand such interference. These asymptotic measures characterize detector performance in regions where the fading of the interfering users as opposed to their specular energies, is the dominant impediment to detection.

An analysis of the conventional detector's performance in the multiuser faded environment reveals that fading interference is capable of limiting detector performance in a manner similar to specular component interference bringing out the hitherto unrecognized performance limiting effect of fading interference in Rician fading CDMA channels. The AWGN multiuser detectors (optimal and decorrelating), which are designed for the CDMA-AWGN channel, while sub-optimally resistant to specular interference, pay a penalty for ignoring the presence of fading in that they are also found to be susceptible to fading interference. Thus we demonstrate that, even in environments where specular interference is marginal, both of the above detectors are incapable of competing with detectors of isolated transmissions, if the fading of an intefering user dominates the background noise. The faded detectors, both optimal and decorrelating, are however found to withstand such interference. In light of their previously demonstrated resistance to specular interference, and the additional computational considerations, our results make the case for the use of the faded decorrelators for multiuser detection over asynchronous Rician faded CDMA channels.

The plots of Figure 1 are an illustration of the implications of our results for detector Bit-Error Rates (BERs) in realistic as opposed to asymptotic environments. We observe that, over a two-user channel, even with a weak interferer specular signal, the first user of the faded decorrelator alone exhibits an exponential decay in BER with increasing Signal-to-Noise Ratio (SNR) with both the conventional detector and the AWGN decorrelator forced by the non-zero, fixed interferer fading, to approach error floors.

References

[1] S. Vasudevan & M. Varanasi, *Multiuser Detectors for Asynchronous CDMA Communication over Rician-Fading Channels*, To appear in the Proceedings of the 1992 IEEE Global Telecommunications Conference (GLOBECOM), December 1992.

[2] S. Vasudevan & M. Varanasi, *Multiuser Detectors for Asynchronous CDMA Communication over Rician Fading Channels : Parts 1 and 2*, DSP Technical Reports 508 and 509, Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309.

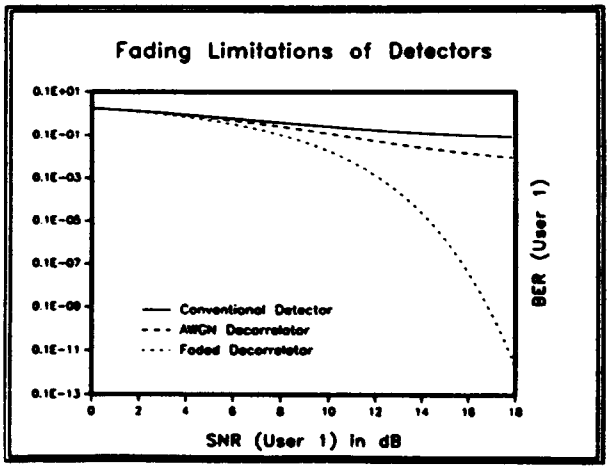


Figure 1. BERs of the conventional detector, the AWGN and faded decorrelators versus SNR.

Equalization Techniques for Direct Sequence Code-Division Multiple Access Systems in Multipath Channels

Sarah Kate Wilson *and John M. Cioffi
Information Systems Laboratory
Stanford University
Stanford, California 94305-4055
email: wilson@isl.stanford.edu

Abstract

We address the problem of Code-Division Multiple Access (CDMA) systems in a multipath environment where a high data rate introduces intersymbol interference as well as inter-user interference. We discuss a linear equalizer as well as a reduced-state-sequence estimation (RSSE) scheme for CDMA channels with intersymbol interference. The linear equalizer is a modification of the decorrelating detector proposed by Lupas and Verdu. The RSSE collapses 2^K states into 2 states where K is the number of users. Both techniques are superior to a traditional matched filter detector (RAKE) when the interuser interference and/or the intersymbol interference is relatively strong. In a severe near-far environment, the RSSE can show significant (2 - 3 dB) improvement over the linear equalizer in terms of receiver signal-to-noise ratio.

Direct Sequence Code-Division Multiple Access (CDMA) has been proposed for commercial data networks. Its strength is it can increase the capacity of a system due to the absence of a guardband requirement[5]. Its weakness is that it can suffer from the near-far problem: a user experiencing strong interference from other users (near) while its own signal is relatively weak (far). Techniques such as power control have been proposed for combating the near-far problem. We propose two additional signal processing methods for when power control is not feasible. The first technique is a linear equalizer, a modification of the decorrelating detector proposed by Lupas and Verdu [2], [3]. The second is a reduced-state sequence estimation (RSSE), an implementation of the Maximum-Likelihood Sequence Detector (MLSD) for CDMA channels with intersymbol interference (ISI) as well as interuser-interference [1].

For modeling the CDMA channels with multi-path, we assume a maximum delay spread of 250 nanoseconds, with 40 chips per bit and a bit rate of 5 Mbs. We modulate the chips with a square-root raised-cosine pulse. We experience ISI from the adjacent data bit, as well as interuser interference. In simulation we have restricted ourselves to 2 and 3 users, but the techniques can be extended to multiple users. At the data and chip rates and delay-spreads we have assumed, our single-user ISI channel is analogous to a $1 + aD$ channel. We also assume that the chip sequence is repeated each information bit and the multipath channels share the same group delay. In both the Linear Detector and the RSSE we assume a good knowledge of the effective multi-dimensional channel.

In the Linear Detector, we match the received signal with each user's multipath channel and chip sequence. The effective channel response after matching has the form of an invertible matrix. We can invert or decorrelate the channel with either a zero-forcing or minimum-mean-square-error solution. It can be shown using the matrix inversion lemma that this solution is equivalent to a linear equalizer at the chip rate.

Implementation at the chip rate has the advantage of not explicitly needing to evaluate the interfering users' channel responses. For the chip-rate equalizer we can treat the interfering users as noise, and estimate the covariance matrix via a training sequence [4].

The RSSE algorithm is an implementation of the MLSD that reduces the number of states from 2^K to 2. In each state the elements differ by at least 2 user values. We use a modified Viterbi algorithm where first we search among possible pairs of inputs within one state, then we find the minimum cost into each of the 2 states of our trellis. The new minimum distance for the RSSE is the minimum of the trellis distance and the distance between the pulse responses to the elements within the states. The trellis minimum distance is determined by the smallest user received energy. By forcing the elements within a state to differ by at least two values, the system must mistake the output from two different users before deciding incorrectly. In a severe near-far environment, the minimum distance of the trellis will often be the smaller of the two distances, guaranteeing optimum detection in the presence of additive white Gaussian noise.

We have simulated the performance of the linear equalizer and the RSSE and compared them to the output of a matched filter (RAKE) detector in multi-channel environments where one user experiences a near-far problem. We have found that both techniques have a SNR 15-20 dB greater than the traditional RAKE, with the RSSE performing 2-3 dB better than the linear equalizer in most severe near-far environments with intersymbol interference.

References

[1] M. V. Eyoboğlu and S. U. Qureshi. Reduced-state sequence estimation with set partitioning and decision feedback. *IEEE Transactions on Communications*, 36:13–20, January 1988.

[2] R. Lupas and S. Verdu. Linear multiuser detectors for synchronous code-division multiple-access channels. *IEEE Transactions on Information Theory*, 35(1):123–136, January 1989.

[3] R. Lupas and S. Verdu. Near-far resistance of multiuser detectors in asynchronous channels. *IEEE Transactions on Communications*, 38(4):496–508, April 1990.

[4] D.D. Falconer M. Abdulrahman and A. U. H. Sheikh. Equalization for interference cancellation in spread spectrum multiple acess systems. In *Vehicular Technology Conference VTC '92*, Denver, Colorado, April 1992.

[5] K. Gilhousen I. Jacobs R. Padovani and L. Weaver. Increased capacity using cdma for mobile satellite communication. *IEEE Journal on Selected Areas in Communications*, 8(4):503–514, May 1990.

This work was supported by CASIS project number NAGW-419-S10 and JSEP project number DAAL03-88-C-0011

A Comparison of Differentially Coherent and Coherent Multiuser Detection With Imperfect Phase Estimates in a Rayleigh Fading Channel

Zoran Zvonar and David Brady
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115

Summary

Multiuser detectors have superior performance over their single-user counterparts in a multiple-access channel, assuming perfect knowledge of system parameters [1-3]. In this paper we extend the analysis of multiuser detectors in fading channels by incorporating the effects of imperfect parameter estimates on symbol error probability. This type of analysis should be useful in designing multiuser receivers, showing the error rate sensitivity to channel parameter mismatch.

We focus on a synchronous CDMA channel shared by K users where the signal of each user arrives at the central receiver through an independent flat Rayleigh fading channel. The central receiver has the knowledge of the signature waveforms of all users, and the outputs of a matched filter bank provide the sufficient statistics. We also assume that a state space description of the fading distortion is available.

In a single-user situation an optimum receiver structure in the flat fading channel [4] consists of an adaptive estimator of fading distortion, and a detector which utilizes these estimates. The estimation of the complex channel distortion performs the task of the carrier recovery. Given the state-space model of the channel distortion, the Kalman filter is the optimal, minimum variance state estimator. A suboptimum, realizable receiver can be implemented using the decision-directed approach where the data dependence is removed from the matched filter output.

Multiuser carrier recovery can be accomplished in two ways. The first approach is proposed in [5] in which the matched filter outputs are decorrelated and each user employs phase estimators which assume isolated transmission. In this case the multiple-access interference is removed from the matched filter outputs at the expense of noise enhancement and correlation, which affects the performance of the carrier recovery circuit. We also consider the vector generalization of the receiver proposed for the single-user channel [4]. Due to synchronism among the users, the data dependency can be removed in a decision-directed manner and the joint phase estimates are obtained by using a multi-input multi-output Kalman filter.

We focus our analysis on two low-complexity suboptimum multiuser detectors, the coherent and differentially coherent decorrelating detector. The coherent decorrelating detector utilizes phase estimates obtained by the aforementioned carrier recovery techniques. In this case, the probability of error can be calculated using Stein's unified analysis [7]. Assuming perfect symbol phase elimination, the lower bound on the error probability is given by

$$P_{k,coh.} = \frac{1}{2}\left[1 - \sqrt{\frac{1 - G_{kk}}{1 + \frac{[\mathbf{R}^{-1}]_{kk}}{\gamma_k}}}\right] \qquad (1)$$

where G_{kk} is the error variance of the phase estimate, $[\mathbf{R}^{-1}]_{kk}$ is the element of the cross-correlation matrix inverse and γ_k is the average signal to noise ratio, all corresponding to the k^{th} user. Note that the error probability of the coherent decorrelating detector does not depend on interfering signal amplitudes, although it depends on the cross-correlations of normalized signature waveforms and the estimation error. In contrast to the case of perfect estimation, the error probability floor is observed, which depends on the error variance related to the phase tracking inaccuracies.

Considering the carrier recovery as the estimation of the fading distortion we reveal a means for comparing coherent and differentially coherent detectors [6]. In the case when we are not able to estimate complex channel coefficients, both signal energies and phases of all users are unknown at the central receiver, and we resort to differentially coherent decorrelating detector, applying the differential decision logic after the decorrelating filter [2]. Taking into consideration the performance degradation due to channel phase changes over two consecutive signaling intervals, the error probability expression for the k^{th} user is

$$P_{k,diff.coh.} = \frac{1}{2}\left[1 - \frac{r_z(1)}{1 + \frac{[\mathbf{R}^{-1}]_{kk}}{\gamma_k}}\right] \qquad (2)$$

where

$$r_z(1) = \frac{E\{z_k(i)z_k^*(i-1)\}}{E_b E\{|c_k|^2\}} \qquad (3)$$

and E_b is energy per bit.

Several numerical examples will be provided for the comparison of these multiuser detectors. For the coherent decorrelating detector the comparison of two analyzed carrier recovery strategies indicate that the joint phase detection results in smaller variance of the phase estimate, resulting in better performance of the detector. This is to be expected since the decorrelating filter enhances the noise prior to carrier recovery. Although an error probability floor is observed for both analyzed multiuser detectors, the coherent detector outperforms the differentially coherent one. However, this is true comparing the lower bound, when perfect elimination of the symbol phase has been assumed, to the exact expression for the error probability of the differentially coherent scheme.

References

[1] R.Lupas, S.Verdu, "Linear Multiuser Detectors for Synchronous Code-Division Multiple-Access Channels", IEEE Trans. on Information Theory, Vol IT-35, No 1, pp 123-136, January 1989.

[2] M.Varanasi, B.Aazhang, "Optimally Near-Far Resistant Multiuser Detection in Differentially Coherent Synchronous Channels", IEEE Trans. on Information Theory, Vol IT-37, No 4, pp 1006-1018, July 1991.

[3] Z.Zvonar, D.Brady, "On Multiuser Detection in Asynchronous CDMA Flat Rayleigh Fading Channels", Proceedings of The Third International Symposium on Personal, Indoor and Mobile Radio Communications, Boston, Massachusetts, October 1992, pp 123-127.

[4] R.Haeb, H.Meyr, "A Systematic Approach to Carrier Recovery and Detection of Digitally Phase Modulated Signals on Fading Channels", IEEE Trans. on Commun., Vol. 37, No 7, pp 748-754, July 1989.

[5] S.Miller, "Detection and Estimation in Multiple-Access Channels", PhD Thesis, Princeton University, 1989.

[6] R.Haeb, "A Comparison of Coherent and Differentially Coherent Detection Schemes for Fading Channels", Proceedings of VTC 1988, pp 364-370.

[7] M.Fitz, "Further Results in the Unified Analysis of Digital Communication Systems", IEEE Trans. on Commun., Vol. 40, No 3, pp 521-532, March 1992.

MMSE DETECTION OF CDMA SIGNALS: ANALYSIS FOR RANDOM SIGNATURE SEQUENCES

Upamanyu Madhow and Michael L. Honig
Bellcore
445 South Street, Morristown, NJ 07960

Abstract: The performance of a finite complexity Minimum Mean Squared Error (MMSE) linear detector for demodulating Direct Sequence Spread-Spectrum (DS/SS) Code Division Multiple Access (CDMA) signals is studied. The MMSE detector is near-far resistant, and can be implemented adaptively when no explicit knowledge of the interferers' signature sequences is available. We assume that users are assigned random binary signature sequences and derive upper and lower bounds on the average near-far resistance of the MMSE detector. For synchronous CDMA, the MMSE detector considered has the same near-far resistance as the maximum likelihood and decorrelating detectors, so that the bounds derived apply to these detectors as well. Approximate expressions for average error probability and signal-to-interference ratio are also presented, and are compared with the analogous results for the matched filter receiver with random signature sequences.

I. INTRODUCTION

Minimum Mean Squared Error (MMSE) linear detection for direct-sequence spread-spectrum (DS/SS) signals has recently been considered in [1], [4]-[5]. In [4] MMSE linear detectors of varying complexity were proposed, and in [5] the near-far resistance and error probability of these detectors were evaluated for a specific assignment of signature sequences. Assuming that the complexity of the detector is matched to the number of strong interferers, these detectors do not suffer from the near-far problem. Furthermore, the MMSE criterion leads to adaptive implementations in which the interference parameters are not explicitly known *a priori*. These schemes are decentralized in the sense that they are designed to demodulate a single user in the presence of multiple-access interference, as opposed to the centralized demodulation of all active users described in [2]-[3] and the references therein.

Here we analyze the performance of the N-tap MMSE detector (N is the processing gain), introduced in [4], assuming that the signature sequences assigned to different users are independent random binary sequences. The performance measures are averaged over the signature sequences of all the users. Although deterministic sequences are used in practice, the assumption of random signature sequences yields a rough characterization of system performance in terms of a few key system parameters (the processing gain N and the number of active users K in this case). This approach has been extensively employed to analyze the performance of the matched filter receiver. It is shown that the N-tap MMSE detector achieves significant performance gains relative to the matched filter.

The N-tap MMSE detector consists of an N-tap linear filter followed by a threshold detector. The tap spacing is assumed to be the chip interval, and the taps are selected to minimize the Mean Squared Error (MSE) between the detected and transmitted symbols. We derive upper and lower bounds on the average near-far resistance of this detector, together with approximations for the signal-to-interference ratio and the error probability. For a synchronous system, it is interesting to note that the MMSE detector has the same near-far resistance as centralized detection schemes such as the maximum likelihood detector and the decorrelating detector (see [2]-[3]), so that the bounds on near-far resistance given here apply to these latter detectors as well.

For the purpose of exposition, we consider a system in which the transmissions are both chip- and symbol-synchronous. Results for asynchronous systems have also been obtained, but are omitted from this summary.

II. SYSTEM MODEL AND RESULTS

Consider the equivalent discrete-time system obtained by sampling the output of a filter matched to the chip waveform at the chip rate. There are then N samples per bit interval, which form a received vector $\mathbf{r} \in \mathbb{R}^N$. Denoting the kth bit of user one (the desired user) as $b_1[k] \in \{-1,1\}$, the N-tap MMSE detector forms the estimate $\hat{b}_1[k] = \mathrm{sgn}(\mathbf{c}^T \mathbf{r})$, where $\mathbf{c}$ is selected to minimize $MSE = E\{(\mathbf{c}^T \mathbf{r} - b_1[k])^2\}$. For a synchronous system, the received vector $\mathbf{r} \in \mathbb{R}^N$ corresponding to the kth bit is given by

$$\mathbf{r}[k] = \sum_{j=1}^{K} b_j[k] A_j \mathbf{a}_j + \mathbf{n}[k],$$

where the vector $\mathbf{a}_j$ is the signature sequence of the jth user, A_j is the received amplitude of user j, and the noise vector $\mathbf{n}$ is Gaussian with mean zero and covariance matrix $\sigma^2 I_N$, where I_N denotes the $N \times N$ identity matrix. The signature sequences $\mathbf{a}_j = (a_j[0], a_j[1], \ldots, a_j[N-1])^T$, $j = 1, \ldots, K$, are assumed to be random. That is, $a_j[l]$, $1 \le j \le K$, $0 \le l \le N-1$, are independent random variables each taking value $+1$ or -1 with equal probability.

The *near-far resistance* of the detector is a measure of the robustness of the detector with respect to variations in the received interference power (see [2]-[3] for a technical definition). If the users' signature sequences are not orthogonal, then the near-far resistance of the matched filter detector is zero. For the MMSE detector considered, the near-far resistance is evaluated by letting the interference amplitudes $A_j \to \infty$. In this case the MMSE solution for $\mathbf{c}$ is the orthogonal projection of $\mathbf{a}_1$ onto the space spanned by the interfering vectors $\mathbf{a}_2, \ldots, \mathbf{a}_K$. That is, the MMSE solution becomes the *zero-forcing* solution in the sense that the interference is completely suppressed (at the expense of enhancing the noise). Denoting the preceding orthogonal projection as $\mathbf{o}_1$, the near-far resistance of the MMSE detector is given by $\eta = |\mathbf{o}_1|^2 / |\mathbf{a}_1|^2$.

Let $\mathbf{R}$ denote the normalized crosscorrelation matrix of the interfering users' signature sequences. That is, $\mathbf{R}_{ij} = (\mathbf{a}_i^T \mathbf{a}_j)/N$ for $2 \le i, j \le K$. Also define the normalized crosscorrelation of the desired vector with the ith interference vector as $\rho_i = (\mathbf{a}_1^T \mathbf{a}_i)/N$. Then the near-far resistance of the MMSE detector considered can be written as $\eta = 1 - \rho^T \mathbf{R}^\dagger \rho$, where $\mathbf{R}^\dagger$ is a pseudo-inverse of $\mathbf{R}$. Our main results are bounds on the expected value of η, where expectation is with respect to the users' signature sequences. Specifically, we first show that $E[\eta] = 1 - E[d_I]/N$, where d_I is the (random) dimension of the subspace of $\mathbb{R}^N$ spanned by the interference vectors $\mathbf{a}_2, \ldots, \mathbf{a}_K$. We then obtain upper and lower bounds for $E[d_I]/N$, and thereby obtain the following upper and lower bounds for the average near-far resistance,

$$1 - (K-1)/N \le E[\eta] \le 1 - f(K-1)[(K-1)/N],$$

where $f(n) = \prod_{i=1}^{n} [1 - 2^{-(N-i)}]$. Note that the upper and lower bounds are tight for $K \ll N$. Further, the upper bound can be tightened by applying a stochastic domination argument.

We also consider two other performance measures, the signal-to-interference ratio and the error probability. Approximations assuming large N are derived for the expected values of these quantities. Numerical results contrasting the different performance of the MMSE and matched filter receivers for random signature sequences will be presented at the conference. In addition, analogous results for asynchronous systems will be mentioned.

REFERENCES

[1] M. Abdulrahman, D. D. Falconer, and A. U. H. Sheikh, "Equalization for Interference Cancellation in Spread Spectrum Multiple Access Systems," *Proc. VTC '92*, May 1992.

[2] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-35, no. 1, pp. 123-136, January 1989.

[3] R. Lupas and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. COM-38, no. 4, pp. 496-508, April 1990.

[4] U. Madhow and M. L. Honig, "Minimum mean squared error interference suppression for direct-sequence spread-spectrum code-division multiple-access," *Proc. 1st Int. Conf. Universal Personal Commun.*, Dallas, TX, Sept. 28-Oct. 1, 1992.

[5] U. Madhow and M. L. Honig, "Error probability and near-far resistance of minimum mean squared error interference suppression schemes for CDMA," to appear, *Proc. Globecom '92*, Orlando, FL, Dec. 6-9, 1992.

ASYMPTOTIC MULTIUSER EFFICIENCY FOR 2-STAGE DETECTORS IN AWGN CHANNELS

David Brady, ECE Dept., Northeastern University, Boston, MA 02115

Abstract

In the AWGN multiple-access channel with binary phase-shift keying modulation, the k^{th} user error probability for a given demodulator vanishes exponentially with the noise level as $-\eta_k\ SNR_k/2$, where η_k is the asymptotic multiuser efficiency (AME), and SNR_k is the received signal-to-background-noise ratio. Thus, the asymptotic multiuser efficiency is an attenuation of the error rate exponent for isolated transmission and maximum a posteriori demodulation, and provides a simple yet precise means of comparing multiuser receivers for sufficiently low noise levels. To date, this parameter is only known for the following receivers in the 2-user, asynchronous AWGN channel: the maximum likelihood sequence detector, the decorrelating detector, the linear MMSE detector, and the conventional detector. In this talk the asymptotic multiuser efficiencies for a class of detectors for the 2-user, asynchronous AWGN channel will be presented. This class may be loosely described as receivers which estimate and subtract multiple-access interference (MAI) by using tentative data decisions, and includes the two-stage detectors with both conventional or decorrelated tentative decisions. The asymptotic multiuser efficiencies for this class of detectors clearly indicate regions for which a given user should avoid updating tentative decisions and suggest combinations of the above receivers to improve single-user performance. This technique applies to the AME of soft tentative decision strategies as well, and we demonstrate that the near-far resistance of two-stage detectors may be markedly improved using soft decision nonlinearities. Below we present an outline of the approach for conventional tentative decisions.

System Model

The matched-filter output for user 1 at time 0 may be written as

$$y_1(0) = b_1(0)w_1 + b_2(-1)\rho_{21} + b_2(0)\rho_{12} + n_1(0),$$

where $b_k(i)$ is the binary antipodal data of user k during time $[iT,(i+1)T)$, w_k is the energy of the waveform $s_k(t)$, the symbol waveform for user k, and ρ_{jk} describes the (2) partial cross-correlation among the asynchronous waveforms $s_j(t)$ and $s_k(t)$. In general, we define $n_k(j)$ as the Gaussian noise component in $y_k(j)$, the matched-filter output for user k at the end of the j^{th} symbol period. A general two-stage detector forms a final decision for $b_1(0)$ via sign detection

$$\widehat{b_1(0)} = sgn\left[y_1(0) - b_2\widetilde{(-1)}\rho_{21} - b_2\widetilde{(0)}\rho_{12}\right]$$

where $\tilde{b}$ denotes a tentative decision for the symbol b. If conventional tentative decisions are to be employed, then

$$\widetilde{b_2(i)} = sgn[y_2(i)].$$

It has been shown that the error probability for user 1 may be expressed as a finite number of terms, each one is proportional to

$$E\left[\ 1_{\{l(-1,i)n_2(-1)>l(-1,i)\alpha(-1,i)\}}\ 1_{\{n_1(0)>w_1-2(i\rho_{21}+j\rho_{12})\}}\ 1_{\{l(0,j)n_2(0)>l(0,j)\alpha(0,j)\}}\right]$$

where $l(\ell,j) = \pm 1$, $i,j \in \{0,\pm 1\}$, and $\alpha(\ell,j)$ are constants.

An exact form for the exponential rate of this probability is crucial to the solution of the asymptotic multiuser efficiency for the two-stage detector, and is found via the following lemma. Let the noise vector $[n_2(-1),n_1(0),n_2(0)]^T$ be Gaussian with zero mean vector and autocovariance matrix $\sigma^2\mathcal{K} = \sigma^2 SS^T$, and let the i^{th} row of S be denoted by S_i^T. Let $\mathbf{g} = [g_1,g_2,g_3]^T$ denote a Gaussian vector with zero mean and autocovariance matrix $\sigma^2 I$, and let $\chi_{i,j} = \{\mathbf{g} : l(-1,i)S_1\cdot\mathbf{g} > l(-1,i)\alpha(-1,i)\}\cap\{\mathbf{g} : S_2\cdot\mathbf{g} > w_1 - 2(i\rho_{21}+j\rho_{12})\}\cap\{\mathbf{g} : l(0,j)S_3\cdot\mathbf{g} > l(0,j)\alpha(0,j)\}$, where $S_i\cdot\mathbf{g}$ denotes a vector inner product.

Lemma

$$\lim_{\sigma\to 0}\frac{\log\mathcal{P}\left[\chi_{i,j}\right]}{-\frac{\|\mathbf{q}\|^2}{2\sigma^2}} = 1$$

where $\mathbf{q}$ is the vector which satisfies

$$\mathbf{q} = \text{argmin}_{\mathbf{x}\in\chi_{i,j}}\ \|\mathbf{x}\|$$

Numerical Example

Figure 1 shows the AME for user 1 in the asynchronous, 2-user AWGN channel with significant correlation among the normalized waveforms. As usual, the AME is displayed as a function of the relative energy of the interferer. We have shown the AME for the maximum likelihood sequence detector (MLS), the decorrelator, and the conventional detector, and the two-stage detector with both conventional and decorrelated tentative decisions. Note that the AME of the two-stage detector with decorrelated tentative decisions dominates that using conventional tentative decisions. Also of interest is to note that the AME of the two-stage detector with conventional tentative decisions is dominated by that of the conventional detector for sufficiently weak interference, and that the near-far resistance of the former detector is zero. Both two-stage detectors exhibit similar error rate exponents to their tentative decision counterparts when the energies of the users are roughly the same.

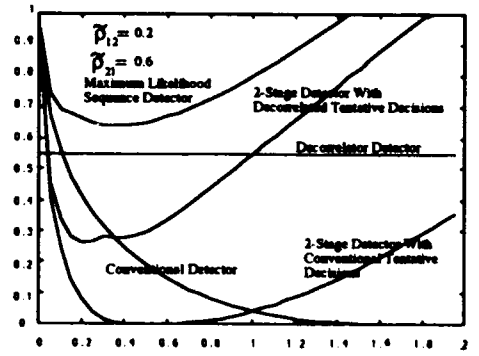


Figure 1. AME for User 1 in the 2-User Asynchronous AWGN Channel for Various Detectors

Universal coding of non-discrete sources based on distribution estimation consistent in expected information divergence

Andrew R. Barron*, László Györfi†, and Edward C. van der Meulen‡

We show how a certain distribution estimator μ_n^* which is consistent in expected information divergence leads to a universal code for the class of all probability measures μ on a non-discrete space $\mathcal{X}$ which are dominated in I-divergence by a known probability measure ν.

Let μ be an unknown probability measure on $\mathcal{X} = \mathbb{R}^d$ and consider the problem of estimating μ based on i.i.d. observations $X_1, ..., X_n$ from μ. As a priori information we assume that there exists a known probability measure ν on $\mathcal{X}$ such that $I(\mu, \nu) < \infty$. Define integers $m_n, 0 < m_n < n$, and real numbers $h_n > 0$. Let

$$\mathcal{P}_n' = \{A_{n,1}', ..., A_{n,m_n}'\}$$

be a sequence of partitions of $\mathcal{X}$ such that each $A_{n,i}'$ is a cube of width h_n and $\nu(A_{n,i}') \geq h_n^d$. Let $0 < a_n < 1$ be a given sequence with

$$\lim_{n\to\infty} a_n = 0.$$

Let μ_n denote the standard empirical measure for $X_1, ..., X_n$. In [1] we introduced the distribution estimator μ_n^* defined by

$$\mu_n^*(A) = (1-a_n)\sum_{i=1}^{m_n} \mu_n(A_{n,i}')\frac{\nu(A \cap A_{n,i}')}{\nu(A_{n,i}')} + a_n\nu(A)$$

and proved the following theorem.

Theorem 1 If $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} \frac{m_n}{n} = 0$, and $\limsup_{n\to\infty} \frac{1}{na_nh_n} \leq 1$, then

$$\lim_{n\to\infty} E(I(\mu, \mu_n^*)) = 0$$

for all μ such that $I(\mu, \nu) < \infty$.

Hence our distribution estimator μ_n^* is consistent in expected information divergence for all μ for which $I(\mu, \nu) < \infty$. This consistent estimation leads naturally to a universal source code (in the sense of [2]) for the same class of distributions for arbitrarily fine quantizations of the data in the following way.

Since $\mathcal{X}$ is not a discrete space, data sequences $X_1, ..., X_n$ cannot be represented exactly by a noiseless source code. Nevertheless, for any given partition $\mathcal{P}_n = \{A_{n,i}\}$ of $\mathcal{X}^n$, no matter how fine, we can code the element of the partition that includes the data in a uniquely decodable way, using a Shannon code, which assigns a codeword $\phi(A_{n,i})$ of length $\lceil \log 1/\eta_n(A_{n,i}) \rceil$ to $A_{n,i}$ for any given probability distribution η_n on $\mathcal{X}^n$. The redundancy $R_n(\mathcal{P}_n)$ of this code equals $\frac{1}{n} I_n(\mu^n, \eta_n)$, where $I_n(\mu^n, \eta_n)$ denotes the information divergence between μ^n and η_n restricted to $\mathcal{P}_n$. The least upper bound R_n^* (over all partitions $\mathcal{P}_n$) on the redundancy is provided by $\frac{1}{n} I(\mu^n, \eta_n)$. Based on our distribution estimator μ_n^* we can construct a distribution η_n^* such that

$$\frac{1}{n} I(\mu^n, \eta_n^*) = \frac{1}{n}\sum_{k=1}^{n} E(I(\mu, \mu_{k-1}^*)).$$

The latter term, being a Cesàro average, will tend to zero as $E(I(\mu, \mu_n^*)) \to 0$. Hence, using η_n^* to encode elements of a partition $\mathcal{P}_n$, we obtain

Theorem 2 For all discrete memoryless sources with marginal distribution μ for which $I(\mu, \nu) < \infty$, there exists a universal uniquely decodable code for any partition $\mathcal{P}_n$ such that

$$\lim_{n\to\infty} R_n^* \to 0.$$

We conclude that the estimator μ_n^*, being consistent in expected information divergence, provides a universal code for arbitrarily fine quantizations of the data for the class of all distributions μ on $\mathbb{R}^d$ with $I(\mu, \nu) < \infty$.

References

[1] A. R. Barron, L. Györfi, and E. C. van der Meulen, "Distribution estimation consistent in total variation and in two types of information divergence," *IEEE Trans. Inform. Theory*, vol. IT-38, no. 5, pp. 1437-1454, Sept. 1992.

[2] L. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, pp. 783-795, Nov. 1973.

*Department of Statistics, Yale University, New Haven, CT 06520, USA.

†Department of Mathematics, Technical University of Budapest, Budapest, Hungary.

‡Department of Mathematics, Katholieke Universiteit Leuven, Leuven, Belgium.

Sequential Model Estimation for Universal Coding and the Predictive Stochastic Complexity of Finite-State Sources

Marcelo J. Weinberger * Meir Feder † Jorma Rissanen ‡

In this work we consider sequential universal coding of Finite-State-Machine (FSM) probabilistic sources. A unifilar FSM source $\mathcal{X}$ over a discrete alphabet A with α letters is defined by an FSM F and a set of parameters $\underline{\theta} = \{p(x|s), s \in S, x \in A\}$, where the number of states $S = |\mathcal{S}|$ is finite, and F has an initial state s_0 and its progress is determined by a next-state function

$$s_i = f(s_{i-1}, x_i).$$

In the sequel we refer to F as the machine supporting the model, and it is assumed irreducible and aperiodic. The probability that $\mathcal{X}$ assigns to a string $x_1^n = x_1, \ldots, x_n$ is

$$P(x_1^n; F, \underline{\theta}) = \prod_{i=1}^{n} p(x_i|s_{i-1}). \tag{1}$$

The per-symbol entropy of blocks of length n emitted by $\mathcal{X}$ is denoted $H_n(\mathcal{X}; F)$.

It was shown in [1] and [2] that the average codelength $E\{L(x_1^n)\}$ of any encoder in compressing the outcome of every FSM source $\mathcal{X}$, except, possibly, a set of FSM sources whose volume vanishes as n increases, satisfies

$$\frac{1}{n} E\{L(x_1^n)\} \geq H_n(\mathcal{X}; F) + \frac{S(\alpha-1)\log n}{2n}(1-\epsilon), \tag{2}$$

for every $\epsilon > 0$ and n sufficiently large.

The lower bound in (2) can be achieved up to $O(n^{-1})$ by a simple batch universal encoder which sends as a header the empirical counts in each state of the FSM, and then assigns to the data a code matched to these counts. If F is unknown, it is estimated from the data and then sent in the header as well at a cost $O(n^{-1})$. The model is estimated by minimizing

$$-\frac{1}{n}\log P(X_1^n; F, \underline{\theta}) + \frac{S(\alpha-1)\log(n+1)}{2n} \tag{3}$$

over $\underline{\theta}$ and F. Note that

$$\min_{\underline{\theta}} [-\log P(x_1^n; \underline{\theta}, F)] = n\hat{H}(x_1^n; F) \tag{4}$$

i.e., at the optimal choice of parameters, the string probability becomes the empirical entropy with respect to F. The model selection rule of (3) minimizes the codelength of this batch universal procedure since the first term represents the cost of encoding the data given the model and its parameters, the second term represents the cost of encoding the $S(\alpha-1)$ empirical counts, and the cost of encoding the description of F is independent of n. The minimum description length of a sequence with respect to a class of models, using a possibly batch procedure, has been termed [3] the *non-predictive stochastic complexity* of that sequence with respect to the class.

It was also shown [2] that a similar codelength is obtained if instead of sending the empirical counts explicitly, they are estimated sequentially and used at each time instance to encode the next symbol. F is still estimated from the entire data in a batch procedure by minimizing (3), and is sent as a header. The minimal codelength attained by this universal coding procedure was termed *semi-predictive complexity*.

The most interesting case is a fully predictive one where, in addition to the parameters, F is also estimated sequentially. Thus, such an algorithm assigns to each symbol a probability that depends only on past outcomes and hence can be used to define a *universal* process. It was conjectured in [2], [3] that the fully predictive stochastic complexity is also asymptotically equivalent to the non-predictive stochastic complexity. The main result of this work is a proof of this conjecture in a *probabilistic* setting, where it is assumed that the data is generated by some FSM source. Specifically, if at each time t, F is estimated by

$$\hat{F}(t) = \arg\min_{F} \left[\hat{H}(x_1^n; F) + \frac{2C\alpha S\log(t+1)}{t}\right], \tag{5}$$

where the minimum is taken over all FSM models F and $C > 1 + 1/2\alpha$, then the resulting expected codelength approaches the entropy at the optimal rate of $(k/2)(\log n/n)$ where $k = S(\alpha-1)$ is the number of parameters. Note that the criterion used in (5) is slightly different from a sequential MDL criterion in which the model is estimated sequentially by minimizing an expression similar to (3). The difference is only in the constant of the "penalty term", and not in its functional behavior. Nevertheless the resulting fully predictive complexity is the optimal one, since its expected value is

$$H_n(\mathcal{X}; F) + \frac{S(\alpha-1)\log n}{2n}, \tag{6}$$

up to $O(n^{-1})$ term. This result can be viewed as proving the existence of *universal* FSM sources.

To prove this main result we use the observation [4] that a sufficient condition for achieving the optimal coding rate is that the estimator of F satisfies

$$\sum_{t=1}^{\infty} P_e(t)\log t < \infty \tag{7}$$

where $P_e(t)$ is the probability of error in estimating the model at time t. We show that the estimator (5) is strongly consistent and furthermore satisfies (7). The detailed proof is given in [5].

The fully sequential compression algorithm presented here is not efficient. In a related work [6] we have considered the effective *context* algorithm and have shown that for the restricted class of tree sources its average codelength also approaches the entropy at the optimal rate.

References

[1] J. Rissanen. "Universal Coding, Information, Prediction and Estimation", *IEEE Trans. Information Theory*, IT-30:629–636, 1984.

[2] J. Rissanen. "Complexity of Strings in the Class of Markov Sources", *IEEE Trans. Information Theory*, IT-32:526–532, 1986.

[3] J. Rissanen. "Stochastic Complexity and Modelling" *The Annals of Statistics*, 14:1080–1100, 1986.

[4] M. J. Weinberger, A. Lempel, and J. Ziv. "A Sequential Algorithm for the Universal Coding of Finite-Memory Sources", *IEEE Trans. Information Theory*, IT-38, May 1992.

[5] M. J. Weinberger, M. Feder, and J. Rissanen. "Sequential Model Estimation for Universal Coding and the Predictive Stochastic Complexity of Finite-State Sources", in preparation.

[6] M. J. Weinberger, J. Rissanen, and M. Feder. "A Universal Finite-Memory Source", submitted for publication, *IEEE Trans. Information Theory*.

*Marcelo J. Weinberger was with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, 32000, ISRAEL. He is now with IBM Almaden Research Center, San Jose, 95120, CA, U.S.A

†Meir Feder is with the Department of Electrical Engineering - Systems, Tel-Aviv University, Tel-Aviv, 69978, ISRAEL

‡Jorma Rissanen is with IBM Almaden Research Center, San Jose, 95120, CA, U.S.A

Rate and Distortion Redundancies for Universal Source Coding with Respect to a Fidelity Criterion

Philip A. Chou† and Michelle Effros‡
† *Xerox Palo Alto Research Center, 3333 Coyote Road, Palo Alto, CA 94304*
‡ *Information Systems Laboratory, Stanford University, Stanford, CA 94305-4055*

Let $\{X_i\} \sim P_\theta$, $\theta \in \Lambda \subseteq \mathsf{R}^K$. Rissanen has shown that there exist universal noiseless codes for $\{X_i\}$ with per-letter rate redundancy as low as $\frac{K}{2}\frac{\log N}{N}$, where N is the blocklength and K is the number of source parameters. We derive an analogous result for universal source coding with respect to the squared error fidelity criterion: there exist codes with per-letter rate redundancy as low as $\frac{K}{2}\frac{\log N}{N}$ and per-letter distortion (averaged over X^N and θ) at most $D(R)[1+\frac{K}{N}]$, where $D(R)$ is an average distortion-rate function and K is now the number of parameters in the code.

Let $\{X_i\}$ be a random process over alphabet $\mathcal{X}$ with process measure P_θ, $\theta \in \Lambda$, and let $q^N = \beta^N \circ \alpha^N$ be a length-N block code with encoder $\alpha^N : \mathcal{X}^N \to \mathcal{C}$ and decoder $\beta^N : \mathcal{C} \to \mathcal{Y}^N$, where $\mathcal{C} = \{c_1, \ldots, c_M\} \subseteq \{0,1\}^*$ is some binary prefix code and $\mathcal{Y}$ is the reproduction alphabet, typically equal to $\mathcal{X}$. A universal source code with respect to a fidelity criterion $d(X^N, Y^N) = \sum d(X_i, Y_i)$ is a sequence of block codes $\{q^N\}$ such that for each $\theta \in \Lambda$ there exists a corresponding sequence of points $\{(R_{N,\theta}, D_{N,\theta})\}$ on the graph of the Nth order operational distortion-rate function for P_θ for which the per-letter "rate" redundancy

$$\frac{1}{N}\rho_\theta(q^N) = \frac{1}{N}E_\theta|\alpha^N(X^N)| - R_{N,\theta} \tag{1}$$

and the per-letter "distortion" redundancy

$$\frac{1}{N}\delta_\theta(q^N) = \frac{1}{N}E_\theta d(X^N, q^N(X^N)) - D_{N,\theta} \tag{2}$$

each go to zero uniformly in θ (in which case the code is *strongly minimax* universal), pointwise in θ (in which case the code is *weakly minimax* universal), or in expectation with respect to a probability measure on θ (in which case the code is *weighted* universal). In the noiseless case, where $D_{N,\theta} = 0$ and $R_{N,\theta} = H_\theta(X^N)$, Rissanen [1] has shown that when $\Lambda \subset \mathsf{R}^K$ is compact with a non-empty interior, and $\{P_\theta\}$ satisfies certain regularity conditions, there exists a universal code $\{q^N\}$ with per-letter rate redundancy

$$\frac{1}{N}\rho_\theta(q^N) \le \frac{K}{2}\frac{\log N}{N} + o\left(\frac{\log N}{N}\right) \tag{3}$$

for each $\theta \in \Lambda$. (Hence the code is weakly minimax universal.) Rissanen goes on to show that this is also the minimum redundancy achievable by any universal code $\{q^N\}$, for almost all θ (with respect to Lebesgue measure).

We derive a result analogous to (3) for weighted universal source coding with respect to the squared error criterion by analyzing a two-part fixed-rate coding scheme [2], first analyzed by Zeger and Bist [3]. In that scheme, each block $X^N = X^{nk} = (X_1^k, \ldots, X_n^k)$ is encoded in two parts. In the first part, a k-dimensional, M-codeword vector quantizer $q^k = \beta^k \circ \alpha^k$ is optimized for the ample distribution of $X_1^k, \ldots, X_n^k$, and then the M reproduction codewords $\{\beta^k(c)\}$ are themselves encoded using a fixed "universal" vector quantizer optimized for the distribution of the $\beta^k(c)$s (averaged over all c, X^N, and θ). In the second part of the encoding, $X_1^k, \ldots, X_n^k$ are encoded using the quantized code $\hat{q}^k = \hat{\beta}^k \circ \hat{\alpha}^k$.

Let $R = \log M/k$ be the rate in bits per letter of q^k, and let $D_k(R) = ED_{k,\Theta}(R)$ be the average kth order operational distortion-rate function $D_{k,\theta}(R)$. We use the high resolution approximation $D_{k,\theta}(R) = C_{k,\theta}e^{-2R}$ and a Lagrangian formulation to determine the optimal bit allocation between q^k and the "universal" quantizer. For this optimal allocation, we show that (with $R_{N,\theta} = R$ in (1) and $D_{N,\theta} = D_{N,\theta}(R)$ in (2)) the per-letter rate and distortion redundancies for the overall code q^{nk} are

$$\frac{1}{N}\rho_\theta(q^{nk}) = \frac{K}{2}\frac{\log N}{N} + o\left(\frac{\log N}{N}\right) \tag{4}$$

and

$$\frac{1}{N}E\delta_\Theta(q^{nk}) \lesssim D_k(R)\left[1+\frac{K}{N}\right] - D_N(R), \tag{5}$$

where $K = Mk$ is the total number of parameters in the code q^{nk}. The same results hold in the case where the codewords $\beta(c)$ are scalar quantized. The recent work of Zeger, Bist, and Linder [4] supports these results.

While our rate redundancy result (4) for universal source coding with respect to a fidelity criterion is consistent with Rissanen's result (3) for universal noiseless coding, our distortion redundancy result (5) is consistent with Akaike's result on the expected decrease in log likelihood for empirical maximum likelihood on N samples, with Davisson's result on the expected increase in squared error for empirical linear prediction on N Gaussian samples, and with Pollard's result that the codewords in a quantizer follow a central limit theorem (which implies that the expected increase in squared error is inversely proportional to the number of samples N).

References

[1] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Trans. Information Theory*, 30(4):629–636, July 1984.

[2] S. Panchanathan and M. Goldberg. Algorithms and architecture for image adaptive vector quantization. In *Proc. Visual Communications and Image Processing*, Cambridge, MA, November 1988. SPIE.

[3] K. Zeger and A. Bist. Universal adaptive vector quantization using codebook quantization. In *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing*, pages III.381–384, San Francisco, March 1992. IEEE.

[4] K. Zeger, A. Bist, and T. Linder. Universal source coding with codebook transmission. Preprint, May 1992.

INFORMATION BOUNDS FOR THE RISK OF BAYESIAN PREDICTIONS AND THE REDUNDANCY OF UNIVERSAL CODES

Andrew Barron, Bertrand Clarke, and David Haussler
Yale Univ., Univ. British Columbia, and Univ. California at Santa Cruz

ABSTRACT: Several diverse problems have solutions in terms of an information-theoretic quantity for which we examine the asymptotics. Let $Y_1, Y_2, \ldots, Y_N$ be a sample of random variables with distribution depending on a (possibly infinite-dimensional) parameter θ. The maximum of the mutual information $I_N = I(\theta; Y_1, Y_2, \ldots, Y_N)$ over choices of the prior distribution of θ provides a bound on the cummulative Bayes risk of prediction of the sequence of random variables for several choices of loss function. This same quantity is the minimax redundancy of universal data compression and the capacity of certain channels. General bounds for this mutual information are given. A special case concerns the estimation of binary-valued functions with Vapnik-Chervonenkis dimension d_{vc}, for which the information is bounded by $d_{vc} \log N$. For smooth families of probability densities with a Euclidean parameter of dimension d, the information bound is $(d/2) \log N$ plus a constant. The prior density proportional to the square root of the Fisher information determinant is the unique continuous density that achieves a mutual information within $o(1)$ of the capacity for large N. The Bayesian procedure with this prior is asymptotically minimax for the cumulative relative entropy risk.

SUMMARY: A parameterized family of distributions $P_{Y^N|\theta}$ is used to model a sequence of random variables $Y^N = (Y_1, Y_2, \ldots, Y_N)$. For problems of data compression and on-line prediction we compare the performance that can be achieved when θ is unknown to the performance that would be achieved if it were known. Entropy and probability of error, repectively, can used to measure the performance. The relative entropy is used to bound the additional risk due to lack of knowledge of the parameter. If θ were known, the best on-line prediction and compression of the sequence of variables Y_k would be available from the conditional distribution $P_{Y_k|Y^{k-1},\theta}$. If θ is unknown, these actions may be based on an estimate of the conditional distribution using the observed past. When a prior distribution is assigned to the parameter, Bayesian proceedures use the distribution $P_{Y_k|Y^{k-1}}$ obtained by averaging out the parameter. We examine the cumulative relative entropy distance between these predictive distibutions. By the chain rule this quantity reduces to the relative entropy $D_{N,\theta} = D(P_{Y^N|\theta} || P_{Y^N})$ between the joint distributions of Y^N, with and without conditioning on θ. In statistical terminology, $D_{N,\theta}$ is the cumulative risk, when relative entropy is used as the loss function. Averaging with respect to the prior distribution of θ yields the mutual information I_N as the (cumulative) Bayesian risk. Maximizing the Bayes risk I_N with respect to the choice of the prior for θ, yields the information capacity C_N and determines the sequence of Bayes estimators of the conditional distribution that are minimax, i.e., that minimizes the maximum value of $D_{N,\theta}$. In situations where determination of the exact asymptotics of I_N is not possible, bounds on I_N may be used to provide bounds on the minimax cumulative risk.

In universal noiseless coding of discrete random variables, the redundancy $R_{N,\theta}$ of a code is the increase in the expected total codelength due to the lack of knowledge of the parameter value. For the code based on P_{Y^N}, the relative entropy $D_{N,\theta}$ is the redundancy; the information I_N is the average redundancy; the information capacity C_N is the minimax redundancy; and the choice of the prior that achieves the capacity provides the minimax code (Davisson 1973, Davisson and Leon-Garcia 1980).

In the online prediction problem, we let the regret $r_{N,\theta}$ be defined as the increase in the expected frequency of mistakes in predicting the values of the sequence, due to the lack of knowledge of the parameter value. The regret of the sequence of Bayesian predictions is bounded by

$$r_{N,\theta} \leq (2D_{N,\theta}/N)^{1/2}$$

Thus the regret converges to zero if the relative entropy is of smaller order than N. A tighter bound between $r_{N,\theta}$ and $D_{N,\theta}$ is possible if the sequence of conditional distributions satisfy an $\alpha-$separation property, that is, for some $\alpha > 0$, the difference between the first and second largest values of $P(Y_k = y | Y^{k-1}, \theta)$ is never less than α. In this case, the regret of the Bayesian predictions is shown to be bounded by $r_{N,\theta} \leq (2/\alpha) D_{N,\theta}/N$. Averaging with respect to the prior yields Bayes average regret

$$r_N \leq (2/\alpha) I_N/N.$$

A basic role in the analysis of the asymptotics of the mutual information is played by the relative entropy $D(P_{Y^N|\theta} || P_{Y^N|\theta'})$ between the distributions at neighboring parameter points θ and θ'. It is shown that the mutual information is bounded by

$$I_N \leq \inf_{\Pi} \{ D_N(\Pi) + H(\Pi) \}$$

where the infimum is over partitions Π of the parameter space. Here $D_N(\Pi)$ is the average diameter of the cells of the partition as measured by the relative entropy distance and $H(\Pi)$ is the entropy of the discrete random variable induced by the partition. This bound may be used to show that for certain "nonparametric" cases I_N is of order N^ρ with $0 < \rho < 1$. We also give finite and infinite dimensional cases where I_N is of order $\log N$. So the price for lack of knowledge of the parameter is small compared to the total entropy.

In these bounds, we are permitted to have a sequence of exogenous input variables $X_1, X_2, \ldots, X_N$ on which the distributions are conditioned. For example the Y_k may equal a function $f_\theta(X_k)$ corrupted by noise. Of particular interest is the case that the Y_k variables are binary-valued and equal $f_\theta(X_k)$ plus independent Bernoulli (λ) noise (modulo 2), where $f_\theta(x)$ is a given family of binary-valued functions of Vapnik-Chervonenkis dimension d_{vc}, and the noise rate satisfies $0 < \lambda < 1/2$. Then for any prior distribution on θ,

$$I_N \leq d_{vc} \log(eN/d_{vc}).$$

It follows that for the on-line Bayesian prediction of $Y_1, Y_2, \ldots, Y_N$ the relative frequency of errors has average that exceeds the noise level λ by not more than a multiple of $(d_{vc}/N) \log(N/d_{vc})$. Likewise for universal data compression, the length of the Shannon code based on the Bayesian model for $Y_1, Y_2, \ldots, Y_N$, divided by the sample size N, has average that exceeds the noise entropy $h(\lambda)$ by not more than $(d_{vc}/N) \log(eN/d_{vc})$.

Refined results are possible in the case of smooth parametric families of densities $p(y|\theta)$ indexed by a finite-dimensional parameter vector θ. Here $Y_1, Y_2, \ldots, Y_N$ are assumed to be independent and identically distributed when conditioned on the parameter. An asymptotic expression for the mutual information I_N of the form $(d/2) \log N + c(p) + o(1)$ has been determined by Ibragimov and Hasminskii (1973), in which the constant $c(p)$ is precisely determined as a function of the prior density $p(\theta)$. (Somewhat stringent conditions are required for their result; see Efroimovich 1980, Clarke 1989 for other formulations of conditions). Here d is the Euclidean dimension. A related asymptotic expression for $D_{N,\theta}$ is given in Clarke and Barron (1990). This leads us to examine the asymptotics of the capacity C_N and the choices of prior distributions for θ that asymptotically achieve this capacity. For each finite N the optimizing prior distribution is generally discrete (Berger and Bernardo 1989, Zhang and Hartigan 1992). Nevertheless, we show under general smoothness conditions that a unique continuous density $p(\theta)$ achieves a value I_N within $o(1)$ of the capacity C_N. As conjectured by Bernardo (1979), it is Jeffrey's prior, i.e., the prior proportional to the square root of the determinant of the Fisher information matrix. No other prior (continuous or discrete) achieves asymptotically larger value of the mutual information.

We give a further asymptotic decision-theoretic property of the optimal prior. Jeffrey's prior is shown to be asymptotically least favorable, that is, the minimax statistical risk $\inf_p \max_\theta D_{N,\theta}$ (which also equals the capacity C_N) is achieved asymptotically by the Bayesian procedure with Jeffrey's prior, uniquely among continuous priors. Moreover, with this choice of prior, $D_{N,\theta}$ is asymptotically independent of the parameter θ, so that, in this case, the relative entropy D_N, the mutual information I_N, and the capacity C_N are asymptotically the same.

REFERENCES

J. O. Berger and J. M. Bernardo, "Ordered group reference priors with applications to multinomial and variance component problems." Purdue University, Department of Statistics Technical Report, 1989.

J. M. Bernardo, "Reference posterior distributions for Bayesian inference." *Journal Royal Statistics Society, Ser. B* vol. 41, pp. 113-147, 1979.

B. S. Clarke and A. R. Barron, "Information theoretic asymptotics of Bayes methods." *IEEE Transactions on Information Theory* vol. 36, no. 3, pp. 453-471, 1990.

B. S. Clarke, "Asymptotic cumulative risk and Bayes risk under entropy loss, with applications." Ph. D. Thesis, Department of Statistics, University of Illinois, 1989.

L. D. Davisson, "Universal noiseless coding." *IEEE Transactions on Information Theory* vol 19, pp. 783-795, 1973.

L. D. Davisson and A. Leon-Garcia, "A source matching approach to finding minimax codes." *IEEE Transactions on Information Theory* vol. 26, pp. 166-174, 1980.

S. Yu. Efroimovich, "Information contained in a sequence of observations." *Problems in Information Transmission* vol.15, pp. 178-189, 1980.

D. Haussler and A. R. Barron, "How well do Bayes Methods work for on-line prediction of ± 1 values?" To appear in *Proc. Third NEC Symposium on Computation and Cognition*, 1992.

I. A. Ibragimov and R. Z. Hasminskii. "On the information in a sample about a parameter." *Second International Symposium on Information Theory* pp. 295-309, Akademiai, Kiado, Budapest, 1972.

Z. Zhang and J. Hartigan, Department of Statistics, Yale University, personal correspondence, January, 1992.

There is no Universal Source Code for Infinite Alphabet

László Györfi*, István Páli*, and Edward C. van der Meulen†

The vast majority of results in information theory is on situations where the actual probability law is known. Applying information theory in real life problems, there is an obvious question whether the probability law can be learned from data as far as information theory is concerned. In noiseless source coding, for example, if the source alphabet is finite, then the answer to this question is yes, since there are good universal source coding procedures (see e.g. [2]). This paper is on coding for a discrete infinite source alphabet showing that there is no universal source code over the class of discrete memoryless sources with infinite source alphabet and finite entropy.

Let X be a random variable taking values in $\mathcal{X} = \{1,2,3,\ldots\}$ with distribution μ and entropy $H(X) < \infty$. A discrete memoryless source $\{X_i\}$ with the marginal distribution μ is considered.

For a discrete memoryless source let f_n be a variable length uniquely decodable code with source block length n. Let the average codeword length of f_n be denoted by $\bar{l}_n$. The redundancy per letter of f_n is defined by $R_n = \frac{1}{n}(\bar{l}_n - H(X_1,\ldots,X_n))$.

There is a well-known duality between universal coding and distribution estimation consistent in information divergence, namely, there is a universal source code over a subset of the set of all discrete memoryless sources with finite entropy if and only if there is a distribution estimate consistent in information divergence for all sources within this subset. Concerning the aim of this paper the important direction of this equivalence is as follows:

Theorem 1 For any uniquely decodable code f_n we can construct a distribution estimate $\hat{\mu}_n$ such that

$$R_n \geq \mathrm{E}\{I(\mu,\hat{\mu}_n)\}.$$

Theorem 2 If $\{\hat{\mu}_n\}$ is an arbitrary sequence of estimates of μ then there is a μ with $H(X) < \infty$ such that we have

$$I(\mu,\hat{\mu}_n) = \infty \text{ for all } n \geq 1 \quad \text{a.s.}$$

As in Davisson [1], a sequence of uniquely decodable codes $f_1, f_2, \ldots$ is called weakly universal for a class of sources if

$$\lim_{n\to\infty} R_n = 0$$

for all sources in this class.

The following theorem implies that there is no universal code for the class of discrete memoryless sources with finite entropy.

Theorem 3 For any sequence of source codes $\{f_n\}$ there is a memoryless source with finite entropy such that

$$R_n = \bar{l}_n = \infty \quad \text{for all } n.$$

References

[1] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 6, pp. 783–795, Nov. 1973.

[2] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 5, pp. 530–536, Sept. 1978.

*Department of Mathematics, Technical University of Budapest, Budapest, Hungary.

†Department of Mathematics, Katholieke Universiteit Leuven, Leuven, Belgium.

NOISELESS UNIVERSAL ENCODING OF NON-UNIFILAR SOURCES

Yuri M.Shtarkov
Institute for Problems of Information Transmission
101447,Moscow, GSP-4, Ermolovoy str.,19, Russia

SUMMARY

Let s be a discrete m-ary source over alphabet A, f_n be a block- to-variable encoding method for blocks of length n, $r_n(f_n, s)$ and $\rho_n(f_n, s)$ be an "average" redundancy and "maximal" (over all blocks of length n) "individual" redundancy of encoding of source s with code f_n correspondingly. For given set S of sources s the efficiency of encoding f_n is estimated with $r_n(f_n, S) = sup\{r_n(f_n, s), s \in S\}$ or $\rho_n(f_n, S) = sup\{\rho_n(f_n, s), s \in S\}$.

For all the considered sets of unifilar sources the maximal probabilities codes or MP-codes $f_n^* = f_n^*(S)$ [1] satisfy inequality

$$r_n(f_n^*, S) \le \rho_n(f_n^*, S) \le \frac{1}{2n}[\alpha(S)\log n + \beta(S)] \quad (1)$$

where $\alpha(S)$ is a number of independent parameters of distributions in S except apriori distributions (for sources with memory) and $\beta(S)$ is independent on n. For most cases this results are asymptotically optimal.

For many reasons we need to widen the considering sets of sources. The finite-state m-ary source s is described with conditional probabilities $\theta(a, u|u')$, where a and u corresponds to arbitrary moment and u' is a preceding state. But the set $S(m, w)$ of all finite state sources with given alphabet A and set U of w states is very large: it describes with $\alpha(S(m, w)) = (mw - 1)w$ independent parameters of conditional probabilities. So we need to define and to consider the reasonable subsets of $S(m, w)$.

Switching source $s = (s_0, s_1, \ldots, s_M)$ (see also [2]) consists of M subsources $s_1, \ldots, s_M$ which generate letters of alphabet A independently one from another one, and of control source s_0, which realises sequential commutation of subsource's outputs with output of source s. After subsource s_i is switched off it continue to generate "blind" letters during $l_i \ge 0$ steps and then stops (if during this l_i steps it is not switched on again). The blind letters influent the probabilities of the next letters at the output of s_i but not the output of s. And the statistical properties of subsources for "switched on" and "switched off" modes can be different. Let $s_0, s_1, \ldots, s_M$ are chosen independently from sets $S_0, S_1, \ldots, S_M$ correspondingly. The different sets $S = S_0 \times S_1 \times \cdots \times S_M$ were considered. And the main results are those.

Theorem 1. If $S_0, S_1, \ldots, S_M$ are sets of finite-state sources then inequality (1) is true for corresponding set S of switching sources,where

$$\alpha(S) = \sum_{i=0}^{M} \alpha(S_i) \quad (2)$$

is a number of independent parameters.

$S(m, w)$ is just the particular case with $M = w^2$ memoryless (and stable) components, and I.Csiszar proved inequality (2) for it. But in general case $S_1, \ldots, S_M$ can contain both stable and unstable components.

Theorem 2. The sequential universal encoding for set S of switching sources with finite-state subsources and finite-state control source let us satisfy (1) and (2) and needs not more than

$$K = O(n^{\gamma(S)+1}), n \to \infty \quad (3)$$

arithmetic computations and memory sells of fixed size, where $\gamma(S)$ is not less than $\alpha(S)$ and not more than general number of parameters in S.

REFERENCES

[1] Yu.M.Shtarkov, "Sequential Universal Encoding of Single Messages", Problemy Peredachi Informatsii, vol.23, no 3, 1987, pp.3-17.

[2] T.Berger, Rate-Distortion Theory. A mathematical Basis for Data Compression. New Jersey, Prentice-Hall, 1971.2

FAST CODING OF SOURCES WITH UNKNOWN STATISTICS

B.Ya.Ryabko

Department of Appl. Math. and Cybernetics, Novosibirsk Telecommunication Institute, Kirov Street 86, Novosibirsk-125, Russia

Abstract The problem of encoding of the sources with unknow statistics is considered. The efficiency of the codes is estimated by three characteristics: i) the redundancy, defined as the difference between the average codeword length and the Shannon entropy; ii) memory size (in bits) of the coder and decoder program when it is realized on a computer (S) and iii) the average time of encoding and decoding a single letter (T). The time is measured by the number of operation with single-bit words. All of the known methods may be divided in two classes. The Ziv-Lempel's codes and their variants [1] fall under the first class and the arithmetic code [2] with the Lynch-Devisson's code [3] fall under the second one. The codes from the first class need exponential memory size $S = O(\exp(1/r))$ for the achievement of the redundancy r, when r turns to 0. The methods from the second class have small memory size as well as low rate of encoding: $S = O(1/r)^{const}$, $T = O(1/r\ (\log(1/r)^{const})$. In this report we present the code, that combines the merits of both methods: the memory size is small and the rate is high: $S = O(1/r^{const})$, $T = O((\log 1/r)^3)$. We called this method FAST code. We consider the encoding of the Bernoulli sourses only, but it is obvious how to carry the results over to the Markov sourses.

The FAST code. We consider the main idea of the FAST code for the case of encoding of source with known probabilities. Let $A=\{a_1,a_2,\dots a_n\}$ is an alphabet of a source, A^m is a set of words with the length m, $m\geq 1$. Let's assign lexicographic order on A^m. Let p(a) is the probability of the letter $a\in A$. We suppose that all p(a), $a\in A$ have the form of binary fraction with t digits.($t\geq[\log n]$). For every word $U \in A^m$ we'll determine $P(U) = p(u_1)\dots p(u_m)$;

$Q(a_1a_1\dots a_1)=0$; $Q(U) = \sum_{V<U;\,U\in A^m} P(V)$;

$R(U) = Q(U) + P(U)/2$. The code of the word u consists of $\lceil -\log P(U)\rceil + 1$ of binary letters of the word R(U). This is the known Gilbert - Moore alphabetical code . It being deciphered and it's redundancy is equal 2/m. FAST method is based on this code. The main problem is to compute P(U) and R(U) rather rapidly.Let $m = 2^{\delta}$. Let's define

$\Pi_1^1 = P(u_1),\dots,\Pi_m^1 = P(u_m)$, $\lambda_1^1 = Q(u_1),\dots\lambda_m^1 = Q(u_m)$, $\Pi_k^i = \Pi_{2k-1}^{i-1}\,\Pi_{2k}^{i-1}$, $\lambda_k^i = \lambda_{2k-1}^{i-1} + \Pi_{2k-1}^{i-1}\lambda_{2k}^{i-1}$

$i = 2,\dots,\delta+1$; $k = 1,\dots,m/2^{i-1}$.

It's easy to see that

$R(U) = \lambda_1^{\delta+1} + \Pi_1^{\delta+1}/2$; $P(U) = \Pi_1^{\delta+1}$.

Let's consider the example.Let $A = \{a_1, a_2\}$, $p(a_1) = .11$, $p(a_2) = .01$, $t=2$, $m=4$, $U= a_1a_2a_2a_1$. Then $\Pi_1^1=.01$, $\Pi_2^1=.11$, $\Pi_3^1=.11$, $\Pi_4^1=.01$, $\Pi_1^2=.01\cdot .11=.0011$, $\Pi_2^2=.11\cdot .01=.0011$, $\Pi_1^3=.0011\cdot .0011=.00001001$, $\lambda_1^1 = \lambda_4^1=.11$, $\lambda_2^1 = \lambda_3^1=.00$, $\lambda_1^2 =.11$, $\lambda_2^2=.1001$, $\lambda_1^3=.11011011$, $R(u)=\lambda_1^3+ \Pi_1^3/2 =.110111111$, $\lceil -\log\Pi_1^3\rceil + 1 = 6$ Consequently, code(u) =110111. As is obvious from this example, the main part of calculation is carried on the short words. So the general time of calculation is rather small. The decoding also based on using of $\{\lambda_k^i\}$and $\{\Pi_k^i\}$. The complexity of decoding and coding is similar. The universal FAST code is based on this algorithm and on the author's method of fast estimation of probabilities $p(a_1),\dots,p(a_n)$ [4].It should be noted that the complexity of the code for source with known statistic is less: $T(r)=O((\log(1/r))^2\log\log(1/r))$.

References

[1] Bell T.C.,Cleary J.G.,Witten I.H. Text Compression. Prentice Hall,New Jersey,1990.

[2] Rissanen J., Langdon G.G. "Universal modeling and coding." IEEE Trans. Inform. Theory, v.27, N 1, 1981.

[3] Krichevsky R. E., Trofimov V. K. "The performance universal encoding". IEEE Trans. Inform. Theory, v.27, N 2, 1981.

[4] Ryabko B.Ya. "A fast on-line adaptiv code". IEEE Trans.Inform.Theory,v.38,N 4,1992.

Minimax Redundancy for Sources with an Unknown Model

Joe SUZUKI

College of Science and Engineering, Aoyama Gakuin University
Chitosedai 6-16-1 Setagaya-ku Tokyo 157, Japan

Abstract: This paper describes the construction of a universal code for minimizing L. D. Davisson's minimax redundancy in a range where the true model and stochastic parameters are unknown.

Universal coding can be generally described as a compression method for sources with an unknown or incompletely specified probability distribution [1]. The specific problem investigated here is the development of a universal code that minimizes the redundancy, i.e., the difference $r_n(l,\theta)$ between the expected per-symbol length $\frac{1}{n}E_\theta[l]$ of the codeword generated by the code's length function l [2], and the per-symbol entropy $H(\theta)$ of the source θ, for each source in a predetermined range Λ (not for a specific source θ [1,3,4]), where $E_\theta[\cdot]$, n, and l, are respectively the expectation on the source θ, the length of the data to be compressed $x_1^n = x_1x_2\ldots x_n$, and the length function determining the codeword length. In Davisson [1], coding scheme universality is strictly defined as the property that the redundancy $r_n(l,\theta)$ converges to zero uniformly over all sources in the range Λ, by taking n sufficiently large (strong universality). This property requires redundancy minimization for finite sequences as well as asymptotical optimality (weak universality), and is assured by minimizing minimax redundancy $\sup_{\theta\in\Lambda} r_n(l,\theta)$ for each n in the range Λ [1].

The primary goal of the present paper is to determine the length function l which minimizes the minimax redundancy $\sup_{\theta\in\Lambda} r_n(l,\theta)$ in the range Λ where both the model and stochastic parameters of each source θ in the range Λ are unknown. Furthermore, it is assumed that the model is included in the set of Markov models, with the stochastic parameters being the probabilities that each symbol occurs based on each state in each model.

First, it is shown that universal coding is reduced to determining the weight function $w(\theta)$ which generates the length function $l(x_1^n)$ [1] as

$$l(x_1^n) = -\log[\sum_{\theta\in\Lambda} w(\theta)P(x_1^n|\theta)], \qquad (1)$$

where $P(x_1^n|\theta)$ is the probability that the data to be compressed is x_1^n based on the source θ.

Secondly, the weight function for the framework of state decomposition [5] with a known model is presented as[1]

$$w(\theta) = w[g](p^{k(g)}) = \prod_{s=1}^{S(g)}\{\frac{\Gamma(\alpha a)}{[\Gamma(a)]^\alpha}\prod_{q=0}^{\alpha-1} p[q,s,g]^{a-1}\}, \quad (2)$$

[1] $\Gamma(x)$ is the gamma function of x.

where $S(g)$ and α are respectively the number of the states $s = 1,2,\ldots,S(g)$ and the number of the symbols, one parameter $a > 0$ is selected so that the minimax redundancy $\max_{\theta\in\Lambda} r_n(l,\theta)$ is minimized, and the occurrence probability of each symbols are $p[q,s,g]$ $q = 0,1,\ldots,\alpha-1$ in the same state s. This result is an extension of previous results [3] for composite sources with a known model. A general form of the weight function with an unknown model is then presented as

$$w(\theta) = h(g)w[g](p^{k(g)}), \qquad (3)$$

where $\sum_g h(g) \le 1$ and the function h is seleted so that the minimax redundancy $\max_{\theta\in\Lambda} r_n(l,\theta)$ is minimized, in order to formulate a universal coding method when the model is unknown.

Finally, it is shown that the minimax redundancy achieved with the presented coding method (available for sequential, or adaptive coding [6]) is upper-bounded by the minimax redundancy achieved of J. Rissanen's semi-predictive coding method [7].

Topics for a future study include developing a method to determine the value a for the weight function with a known model, and further investigation into determining the function h for models which minimizes the minimax redundancy.

References

[1] L.D. Davisson. Universal noiseless coding. *IEEE Trans. Inform. Theory*, IT-19(6):783–795, Nov. 1973.

[2] J. Rissanen. A universal prior integer and estimation by minimum description length. *The Annals of Statistics*, 11(2):416–431, 1983.

[3] L.D. Davisson, R.J. Mceliece, M.B. Pursley, and M.S. Wallace. Efficient universal noiseless source codes. *IEEE Trans. Inform. Theory*, IT-27(3):269–279, May 1981.

[4] L.D. Davisson. Minimax noiseless universal coding for Markov sources. *IEEE Trans. Inform. Theory*, IT-29(2):211–215, March 1983.

[5] J. Suzuki. Generalization of the learning method for classifying rules with consistency irrespective of the representation form and the number of the classified patterns. In *ISITA 90*, Waikiki, Hawaii, Nov. 1990.

[6] J.G. Cleary and I.H. Witten. A comparison of enumerative and adaptive codes. *IEEE Trans. Inform. Theory*, IT-30(2):306–315, March 1984.

[7] J. Rissanen. Stochastic complexity and modeling. *The Annals of Statistics*, 14(3):1080–1100, 1986.

Context Tree Weighting : A Sequential Universal Source Coding Procedure for FSMX Sources

F.M.J. Willems, Y.M. Shtarkov and Tj.J. Tjalkens

EE Dept., Eindhoven Univ. and I.P.P.I., Moscow

1 Introduction

A binary *FSMX source* (see [3]) generates a sequence $\{x_t\}_{t=-\infty}^{\infty}$ of digits from $\{0,1\}$, whose statistical behaviour can be described using a *postfix set* $\mathcal{S}$. This postfix set is a collection of binary strings which is *proper* and *complete*. We can now define the *state function* $f(\cdot)$ which maps semi-infinite source sequences $x_{-\infty}^{t-1} = \cdots, x_{t-2}, x_{t-1}$ onto their unique postfix in $\mathcal{S}$. This $s_t = f(x_{-\infty}^{t-1})$ is the state of the source, hence $Pr\{X_t = x | x_{-\infty}^{t-1}\} = P(x|f(x_{-\infty}^{t-1}))$, for $x \in \{0,1\}$. All $P(\cdot|s)$ for $s \in \mathcal{S}$ are probability distributions on $\{0,1\}$. FSMX sources with the same postfix set are said to have the same *model*. In the binary case we need $|\mathcal{S}|$ *free parameters* to specify all its distributions $P(\cdot|s), s \in \mathcal{S}$. The number of free parameters of the actual FSMX source determines the asymptotic redundancy for an optimal code. Sequential source coding procedures for FSMX sources often use a *context tree* (see fig.1). The standard approach (see [4], [2], [3], and [5]) is that one uses this context tree to estimate the current context, i.e. the actual state of the source. This context is used to estimate the distribution of the next source digit. Arithmetic coding procedures can then be used to encode this next symbol with negligible coding-redundancy.

However, instead of estimating the actual state we should try to find a good encoding distribution. This first principle immediately suggests *model weighting*. Model weighting increases the (block-)model-redundancy by at most $-\log(W(\sigma))$ where σ is the actual model, and $W(\sigma)$ its weighting probability. To weight an infinite number of models we introduce a second principle which says that the model-redundancy has to be *proportional* to the number of free parameters of the model. It gives us a weighting distribution over all models. The next section describes an efficient method that weights the block-probabilities of all models according to this distribution.

2 The Context Tree Weighting Procedure

We assume that the maximal depth d of the context tree is finite. A node in the context tree corresponding to context s contains $n_0(s)$ and $n_1(s)$, i.e. the numbers of zeros and ones in the source sequence $x_1 x_2 \cdots x_{t-1}$ that were preceded by s. We assign a block probability

$$P(n_0, n_1) := \frac{\frac{1}{2} \cdot (1+\frac{1}{2}) \cdot \ldots \cdot (n_0 - \frac{1}{2}) \cdot \frac{1}{2} \cdot (1+\frac{1}{2}) \cdot \ldots \cdot (n_1 - \frac{1}{2})}{1 \cdot 2 \cdot \ldots \cdot (n_0 + n_1)} \quad (1)$$

to a (sub-)sequence with $n_0 > 0$ zeros and $n_1 > 0$ ones, etc. This estimator guarantees uniform convergence of the redundancy (see [1]). Using the context tree we can now *recursively* define the weighted probability corresponding to a given context as

$$Q(s) := \begin{cases} P(n_0(s), n_1(s))/2 + Q(0s) \cdot Q(1s)/2 & \text{if depth}(s) < d \\ P(n_0(s), n_1(s)) & \text{if depth}(s) = d. \end{cases} \quad (2)$$

If we apply this method to the context tree in figure 1 we obtain $Q(\lambda) = \frac{45}{16384}$. $Q(\lambda)$ corresponding to the sequence $x_1 x_2 \cdots x_{t-1}$ is our weighted block probability $Pr\{x_1, x_2, \cdots, x_t | x_{-\infty}^0\}$. To process the next x_t we first increment n_0 for contexts $\lambda, x_{t-1}, x_{t-2}x_{t-1}, \cdots$, and $x_{t-d}x_{t-d+1}\cdots x_{t-1}$. Then we update $Q(\lambda), Q(x_{t-1}), Q(x_{t-2}x_{t-1}), \cdots$, and $Q(x_{t-d}x_{t-d+1}\cdots x_{t-1})$ in reverse order. $Q(\lambda)$ is now equal to $Pr\{x_1^{t-1}, X_t = 0 | x_{-\infty}^0\}$. Analogously, by incrementing n_1, we can find $Pr\{x_1^{t-1}, X_t = 1 | x_{-\infty}^0\}$. Division yields the distribution for x_t. The number of operations needed is proportional to the maximal depth d.

3 Upper Bound on the Redundancy

Suppose we know the model $\mathcal{S}$ of the FSMX source but not the values of its free parameters. Then we can use $P(\lambda) := \prod_{s \in \mathcal{S}} P(n_0(s), n_1(s))$ as block estimator. This gives us an optimal (see [2]) parameter-redundancy which can be upperbounded by $|\mathcal{S}|(\frac{1}{2}\log(\frac{t}{|\mathcal{S}|}) + \frac{1}{6}\log(\pi^3 e))$. Inspection of the procedure in the previous section shows that the weighted block probability $Q(\lambda) \geq 2^{-2|\mathcal{S}|+1} \cdot P(\lambda)$. Therefore the model-redundancy of our procedure is at most $(2|\mathcal{S}| - 1)\log 2$.

Note that the above holds for *individual* redundancy relative to the actual source, but also for individual redundancies corresponding to any other FSMX source. Therefore for each source sequence, the context-tree weighting algorithm produces at most $2|\mathcal{S}| - 1$ codebits more than an estimator matched to model $\mathcal{S}$, for any $\mathcal{S}$.

The Eindhovens Hogeschoolfonds supported the second author when he visited Eindhoven's Information Theory Group. Thanks.

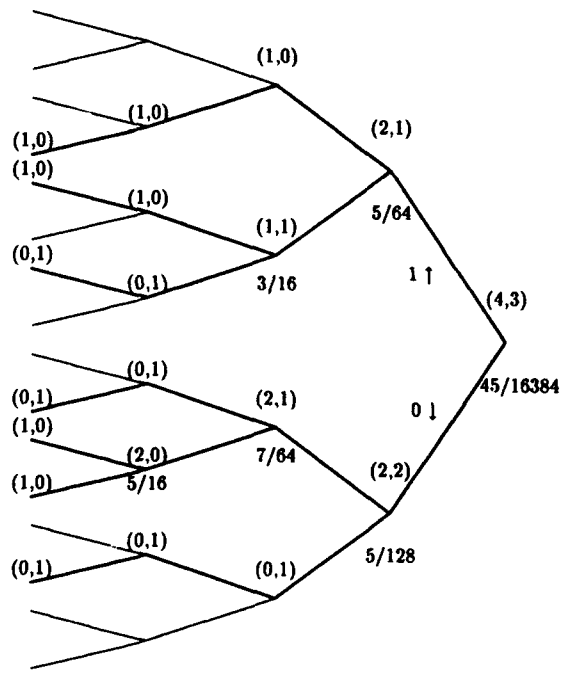


Figure 1: Context tree for $x_1, x_2, \cdots, x_7 = 0110100$, $d = 4$, and $\cdots, x_{-1}, x_0 = \cdots 0010$. Non-trivial Q-values are listed.

References

[1] R.E. Krichevsky and V.K. Trofimov, "The Performance of Universal Encoding," *IEEE Trans. Inform. Theory*, pp. 199-207, March 1981.

[2] J. Rissanen, "A Universal Data Compression System," *IEEE Inform. Theory*, pp.656-664, Sept. 1983.

[3] J. Rissanen, "Complexity of Strings in the Class of Markov Sources," *IEEE Trans. Inform. Theory*, pp. 526-532, July 1986.

[4] J. Rissanen and G.G. Langdon, Jr., "Universal Modeling and Coding," *IEEE Trans. Inform. Theory*, pp. 12-23, Jan. 1981.

[5] M.J. Weinberger, A. Lempel, and J. Ziv, "A Sequential Algorithm for the Universal Coding of Finite Memory Sources," *IEEE Trans. Inform. Theory*, pp. 1002-1014, May 1992.

NEW SPHERICAL 4-DESIGNS

R. H. Hardin
N. J. A. Sloane
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

A spherical code is a collection of points on the surface of a sphere in d-dimensional Eudidean space. A spherical t-design is a spherical code consisting of N points $X_1,\ldots,X_N$ such that the integral of any polynomial of degree at most t over the surface of the sphere is equal to its average value at these N points. Given d and t, one wishes to minimize the value of N.

We have made considerable progress recently on the case $t = 4$ (and informally think we have completely solved the problem). We will give very strong numerical evidence that spherical 4-designs containing N points in d-dimensional space with $d \le 8$ exist precisely for the following values of N and d: N even and ≥ 2 for $d = 1$; $N \ge 5$ for $d = 2$; $N = 12, 14, = 16$ for $d = 3$; $N \ge 20$ for $d = 4$; $N \ge 29$ for $d = 5$; $N = 27, 36, \ge 39$ for $d = 6$; $N \ge 53$ for $d = 7$; and $N \ge 69$ for $d = 8$. These spherical codes also provide optimal (and rotatable) experimental designs for quadratic modelling in the ball.

The full text may be found in our paper "New Spherical 4-Designs", Discrete Math., Vol. 106/107, 1992, pp. 255-264. Details of the methods used and the application to optimal experimental design are in our paper "A New Approach to the Construction of Optimal Designs", J. Stat. Planning and Inference, 1993, in press.

Reduced Complexity Bounded-Distance Decoding of the Leech Lattice

Ofer Amrani and **Yair Be'ery** , Tel-Aviv University Department of Electrical Engineering.
Alexander Vardy , IBM Research Division Almaden Research Center.

***Abstract*-A new efficient algorithm for bounded-distance decoding of the Leech lattice is presented. The algorithm decodes correctly at least up to the guaranteed error-correction radius of the Leech lattice. The proposed decoder is based on projecting the points of the Leech lattice onto the codewords of the (6,3,4) quarternary code, — the *hexacode* H_6. Projection on the hexacode induces partition of the Leech lattice into four cosets of Q_{24}, beyond the conventional partition into two H_{24} cosets. This enables bounded-distance decoding of the Leech lattice with only 1127 real operations in the worst case, as compared to about 3600 operations for the maximum-likelihood decoding of [9]. The proposed algorithm is at least 30% more efficient than Forney's algorithm [5] in terms of computational complexity, while the coding gain loss is no more than 0.05 dB (over BER ranging from 10^{-1} to 10^{-6}).**

The *Leech lattice* Λ_{24} is one of the most interesting and well studied lattices [3]. Maximum-likelihood decoding of Λ_{24} was intensively investigated during the last few years. Conway and Sloane [2], Forney [4], Lang and Longstaff [6], Be'ery, Shahar, and Snyders [1], and Vardy and Be'ery [9] have devised various decoding algorithms with complexities ranging from 55968 down to 3595 operations. While the problem of maximum-likelihood decoding of Λ_{24} is interesting in its own right, in practice it may be rewarding to use a slightly suboptimal but more efficient bounded-distance decoding algorithm. One such algorithm was developed by Forney [5]. The computational complexity of the original Forney's algorithm is somewhat less than 2000 operations. However, since Forney's decoder is based on soft-decision decoding of the Golay code, utilizing the Golay decoder of [8] in Forney's bounded-distance algorithm yields a computational complexity of less than 1500 operations. In this paper we propose a more efficient bounded-distance decoding algorithm which requires only 1127 operations. The proposed algorithm is shown to decode correctly at least up to the guaranteed error-correction radius of the Leech lattice. Simulation results, which compare the coding gain obtained using the new algorithm with the coding gain of the Forney's algorithm, are also provided.

Our construction of the Leech-lattice involves the two-dimensional checkerboard lattice D_2. Partition D_2 into 16 subsets and arrange the labels of the 16 subsets in the following configuration:

$$\begin{array}{cccc} A_{000} & B_{000} & A_{110} & B_{110} \\ B_{101} & A_{101} & B_{010} & A_{101} \\ A_{111} & B_{111} & A_{001} & B_{001} \\ B_{011} & A_{100} & B_{100} & A_{011} \end{array}$$

Tiling the entire space with nonoverlapping copies of scaled and rotated version of this 16-point configuration establishes a correspondence between the labels of the 16 subsets and the points of D_2. Let us represent the points of Λ_{24} by 2×6 arrays of D_2 points, such as :

$$\begin{bmatrix} A_{i_1 j_1 k_1} & A_{i_3 j_3 k_3} & \cdots & A_{i_{11} j_{11} k_{11}} \\ A_{i_2 j_2 k_2} & A_{i_4 j_4 k_4} & \cdots & A_{i_{12} j_{12} k_{12}} \end{bmatrix} \qquad (1)$$

The array in (1) is called type-A since it contains only A_{ijk} points. Similarly, type-B array will consist of only B_{ijk} points. Let $(A_{i_1 j_1 k_1}, A_{i_2 j_2 k_2})^t$ be a column of a type-A array and let $\underline{u} = (i_1, j_1, i_2, j_2)$ be the corresponding binary 4-tuple. If $\underline{u}$ contains an even number of nonzeros then the column is said to be *even*, otherwise the column is said to be *odd*. The index i_1 is called the *h-parity* of the column. The *overall h-parity* of the array is defined as the modulo-2 sum of the h-parities of the six columns. The *overall k-parity* of the array is the modulo-2 sum of the k subscripts of the 12 points. As in [8] any binary 4-tuple $\underline{u}$ is regarded as an *interpretation* of a character $x \in \{0, 1, \omega, \bar{\omega}\} = GF(4)$. Conversely any $x \in \{0, 1, \omega, \bar{\omega}\}$ may be regarded as the *projection* of four different binary 4-tuples $\underline{u}$, such that $x = (0, 1, \omega, \bar{\omega}) \cdot \underline{u}$. The *projection of the array* is a vector $\underline{x} \in GF(4)^6$ consisting of the projections of the six columns. Using the above notation the Leech lattice may be defined as follows [9]:

Definition 1. *The Leech lattice is the set of all the 2×6 arrays whose entries are points of D_2, such that each array satisfies the following conditions:*

i. *It is either type-A or type-B.*

ii. *It consists either of only even columns or only odd columns.*

iii. *The overall k-parity is even if the array is type-A, and odd otherwise.*

iv. *The overall h-parity is even if the columns are even, and odd otherwise.*

v. *The projection of the array is a codeword of H_6.*

Note that by restricting condition i of the foregoing definition, that is taking only the type-A arrays, the *Leech half-lattice* H_{24} is obtained. Further restricting condition ii, that is taking only even columns, produces the *Leech quarter-lattice* Q_{24}, as defined in [9]. The proposed decoding algorithm consists of four separate Q_{24} decoders operating concurrently. Basically the four decoders are identical. We therefore describe only the decoder for Q_{24}. This decoder operates on type-A arrays consisting of only even columns.

Precomputation: Let us assume an AWGN channel model, and let a sequence of 12 two-dimensional symbols $\{r(n)\}_{n=1}^{12}$ be observed at the channel output. For $n = 1, 2, \ldots 12$ find in each A_{ijk}-subset of D_2 a point $\hat{A}_{ijk}(n)$ which is closest to $r(n)$, and set this point as a representative of the entire subset.

Computation: For each $x \in GF(4)$ and for each of the six array locations, the decoder first finds the *preferable representation*, which is the column with the minimum squared Euclidean distance (SED) from the appropriate pair of received symbols. This SED is taken to be the metric of x. Using the acquired information the decoder finds the codeword of H_6 with the minimum metric. A type-A array with even columns is then reconstructed from this codeword of H_6. We show that this array is the closest to the received sequence of symbols. Next conditions iv and iii are checked, in this order. If either of these conditions is violated, correction is performed for condition iv and independently for condition iii using the "Wagner decoding rule" of [7].

The output of the Q_{24} decoder is a Leech quarter-lattice point accompanied by a corresponding metric. This point is not necessarily the closest to the received sequence of symbols due to the independent Wagner decoding. Finally we choose among the outputs of the four Q_{24} decoders the point with the minimal metric, and select this point as the output of our Leech lattice decoder.

Now let d_0 be the minimum distance between points in the checkerboard lattice D_2. The corresponding minimum SED between points in Λ_{24} is given by $d_{min}^2 = 16 d_0^2$. We have the following theorem.

Theorem 1. *Given a received vector of 12 two-dimensional symbols $\mathbf{r} = \{r(n)\}_{n=1}^{12}$, if there is a point $\lambda \in \Lambda_{24}$ such that $\|\mathbf{r} - \lambda\|^2 < 4 d_0^2$, the proposed algorithm decodes $\mathbf{r}$ to λ.*

Theorem 1 implies that the proposed algorithm decodes correctly at least up to the guaranteed error correction radius of the Leech lattice $d_{min}/2 = 2d_0$. This correction capability is the same as that of the bounded-distance decoder of Forney [5]. A comprehensive computer simulation has been performed for both the proposed algorithm and the algorithm of Forney [5]. The simulation assumed a 64-QAM square constellation transmitted over an AWGN channel. Results show no more than 0.05 dB loss in the coding gain for our algorithm versus that of Forney, over the whole range of BER from 10^{-1} to 10^{-6}. This gain loss is due to an increase in the *effective error coefficient*, or the number of nearest neighbors, for the proposed algorithm. This issue will be further elaborated in the paper.

References

[1] Y. Be'ery, B. Shahar, and J. Snyders, "Fast decoding of the Leech lattice," *IEEE J. Select. Areas Comm.*, vol. SAC-7, pp. 959–967, 1989.

[2] J.H. Conway and N.J.A. Sloane, "Soft decoding techniques for codes and lattices, including the Golay code and the Leech lattice," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 41–50, 1986.

[3] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Veralg, New York, 1988.

[4] G.D. Forney, Jr., "Coset Codes II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1152–1187, 1988.

[5] G.D. Forney, Jr., "A bounded distance decoding algorithm for the Leech lattice, with generalizations," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 906–909, 1989.

[6] G.R. Lang and F.M. Longstaff, "A Leech lattice modem," *IEEE J. Select. Areas Comm.*, vol. SAC-7, pp. 968–973, 1989.

[7] R.A. Silverman and M. Balser, "Coding for a constant data rate source," *IRE Trans. Inform. Theory*, vol. 4, pp. 50–63, 1954.

[8] A. Vardy and Y. Be'ery, "More efficient soft-decision decoding of the Golay codes," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 667–672, 1991.

[9] A. Vardy and Y. Be'ery, "More efficient maximum-likelihood decoding of the Leech lattice," to appear in *IEEE Trans. Inform. Theory*.

AN UPPER BOUND ON THE PROBABILITY OF DECODING ERROR FOR M–ARY PSK BLOCK CODED MODULATION STRUCTURES

Hanan Herzberg and Gregory Poltyrev

Department of Electrical Engineering – Systems, Tel–Aviv University,
Ramat–Aviv 69978, Israel.

SUMMARY

Coded modulation, which is an efficient way of combining error correction coding with modulation, is considered here for the case of M–ary Phase Shift Keying (M–PSK). In this paper we are interested in the probability of decoding error for an additive white Gaussian noise (AWGN) channel, in which the well known union bound is useful only when the desired probability of error is rather small. However, the coded modulation structure can be implemented as an inner code concatenated with a Reed – Solomon outer code [1]. Therefore, for a low signal to noise ratio, a tighter upper bound on the probability of decoding error must be derived.

The tangential union bound, which is tighter than the union bound, is presented by Berlekamp in [2] for binary codes. Since each transmitted codeword of an M–PSK coded modulation structure has the same energy, a tangential union bound can be derived for this structure as well. On the other hand, a sphere upper bound, which is derived in [3] for any block coded modulation scheme and is also tighter than the union bound, is applicable also for the M–PSK constellation. However, in the derivation of this bound, the important fact that each transmitted codeword for M–PSK constellation has the same energy, was not taken into account. In our paper an upper bound, named tangential sphere bound (for the case of binary codes see [4]), is derived. It is also proven that our bound is tighter then Berlekamp's tangential bound. In Example 1 of this summary, it is shown that for a particular scheme, which is practically important, our bound is much tighter than the tangential bound for high levels of noise.

Consider a code with fixed length n, M codewords, a set of Euclidean distances $\{\delta_j\}$ (j=1,2,···,N) and a set of average coefficients $\{A_j\}$ (j=1,2,···,N), where A_j is the average number of pairs of codewords with the Euclidean distance δ_j. Let the additive white Gaussian noise at the input to the decoder be a 2n–dimensional vector of a random variable denoted by $\underline{z} = (z_1, z_2 \cdots, z_{2n})$, and let the event of error at the output of the decoder be denoted by E. The probability of decoding error, P_e, can be written as follows:

$$P_e = \Pr(E/\underline{z}\in Cn(\theta))\Pr(\underline{z}\in Cn(\theta)) + \Pr(E/\underline{z}\notin Cn(\theta))\Pr(\underline{z}\notin Cn(\theta)), \quad (1)$$

where $Cn(\theta)$ is a 2n–dimensional cone with half angle θ and a center at the point of 0 energy. Clearly, $\Pr(E/\underline{z}\notin Cn(\theta)) \le 1$. Assuming that the energy of each signal in the M–PSK constellation equals one, from (1) the following tight tangential sphere upper bound is derived

$$P_e \le \int_{-\infty}^{\infty} f(z_1)\, \min_r \Big\{ \sum_{j:\delta_j/2<\alpha} A_j \int_{\Delta_j(z_1)}^{\infty} f(z_2) \int_0^{r_z^2 - z_2^2} f(y_1)\, dz_2\, dy_1 + \int_{r_z^2}^{\infty} f(y)\, dy \Big\}\, dz_1, \quad (2)$$

where f(y) and $f(y_1)$ are the chi–square density functions, $f(z_1)$ and $f(z_2)$ are the normal density functions, r is a real positive parameter (a radius of an 2n–1 dimensional sphere), $r_z = \frac{r}{\sqrt{n}}(\sqrt{n} - z_1)$, $\Delta_j(z_1) = \delta_j/2\ (\sqrt{n} - z_1)$ and $\alpha = r\sqrt{1 - \delta_j^2/4n}$.

It is also proven in the paper that the optimal value of r, denoted by r_o, is the root of the following equation.

$$\sum_{j:\delta_j/2<\alpha} A_j \frac{\Gamma(n-1/2)}{\sqrt{\pi}\ \Gamma(n-1)} \int_0^{\theta_j} \sin^{2n-3}(u)du = 1 \quad (3)$$

where $\cos(\theta_j)=\delta_j/2r_o$ and $\Gamma(\cdot)$ is the Gamma function.

Example 1: Let the structure be a multilevel code, based on the following component codes: at the first partition level the binary Golay(24,12,8) code is employed, at the second level of partition we have the single parity check (24,23,2) code, and the remaining bits are uncoded (these codes are also employed by the well known Leech half lattice). The resulting tangential sphere bound, tangential bound and union bound are presented in Fig. 1. From Fig. 1 we deduce that for the structure of Example 1 our upper bound is much tighter than the union and the tangential bounds for high levels of noise.

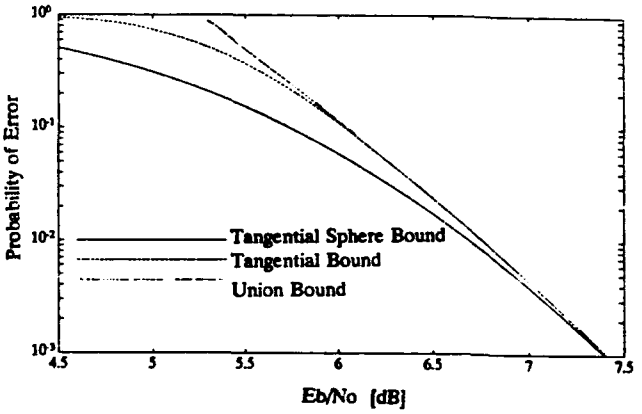


Fig. 1. Upper bounds on the probability of decoding error for the structure of Example 1.

REFERENCES

[1] H. Hersberg Y. Be'ery and J. Snyders, "Concatenated multilevel block coded modulation," accepted for publication in *IEEE Trans. Commun.*

[2] E. R. Berlekamp, "The technology of error–correction codes," *Proceedings of the IEEE*, vol. 68, pp. 564–593, May 1980.

[3] H. Hersberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," Submitted for publication.

[4] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectrum," Submitted.

A BOUNDED-DISTANCE DECODING ALGORITHM FOR LATTICES OBTAINED FROM A GENERALIZED CODE FORMULA

[1]Mauro A.O. da Costa e Silva, and [2]Reginaldo Palazzo Jr.*

[1]EED-EESC-USP, P.O. Box 359, 13560 São Carlos-SP-Brazil

[2]DECOM-FEE-UNICAMP, P.O. Box 6101, 13081 Campinas-SP-Brazil

A multistage decoding algorithm is given for lattices obtained from a generalized code formula. The corresponding multilevel construction is based on a chain of two-way lattice partitions and a family of binary linear block codes, whereas the multistage algorithm involves the use of a maximum-likelihood (ML) decoding algorithm for each two-way lattice partitions as well as a soft-decision ML decoding algorithm for each binary linear block code. The algorithm is shown to have the same effective error-correcting radius as ML decoding. Several known lattices and two new ones were then constructed by the generalized code formula. The trade-off between complexity reduction and performance loss achieved by the proposed algorithm is presented.

INTRODUCTION

The use of multidimensional lattices in block or trellis codes for band limited channels has focused the attention of many researchers on the problem of complexity reduction in lattice decoding. For lattices expressible in terms of code formulas based on the chain of lattice partitions $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}/\ldots$, Forney [1] has proposed a suboptimum algorithm with a better trade-off between complexity reduction and performance loss when the number of levels in the code formula is greater than one. However the decoding of lattices with single-level code formulas like H_{16}, X_{24} and X_{32} can not benefit directly from this algorithm. The present work extends the previous approach by generalizing its multistage algorithm to more general code formulas.

GENERALIZED BINARY CODE FORMULAS (GBCF)

The lattice construction used in the generalized multistage decoding algorithm is based on a chain of two-way n-dimensional lattice partitions $\Gamma_0/\Gamma_1/\ldots/\Gamma_{b-1}/\Gamma_b$ with selected sets of coset representatives and a family of binary linear block codes with length m, dimension k_i and minimum Hamming distance $d_H(C_i)$, $0 \le i \le b-1$.

Let Γ/Λ be an elementary binary partition of order 2^b. Let $\{\alpha_0, \alpha_1, \ldots, \alpha_{b-1}\}$ be a set of vectors forming a binary basis to the partition Γ/Λ, i.e., be a system of coset representatives of Λ in Γ/Λ.

Definition 1 (GBCF) : Let $\Gamma_0/\Gamma_1/\ldots/\Gamma_{b-1}/\Gamma_b$ be a lattice partition chain, and $C_0, C_1, \ldots, C_{b-1}$ be linear binary block codes with blocklength m, then a periodic array L can be defined as follows:

$$L \triangleq \left\{ \bar{c}_0 \bullet \alpha_0 + \bar{c}_1 \bullet \alpha_1 + \ldots + \bar{c}_{b-1} \bullet \alpha_{b-1} : \bar{c}_i \in C_i,\ 0 \le i \le b-1 \right\} + \Gamma_b^m$$

Let us denote the Generalize Binary Code Formula by

$$L = C_0 \bullet [\Gamma_0/\Gamma_1] + C_1 \bullet [\Gamma_1/\Gamma_2] + \ldots + C_{b-1} \bullet [\Gamma_{b-1}/\Gamma_b] + \Gamma_b^m$$

Then the following statements are proved: 1) L is a lattice; 2) exact expressions for the minimum distance and the error coefficient of L.

We provide some examples of lattices L that can be obtained by the GBCF as well as new lattices.

GENERALIZED MULTISTAGE DECODING FOR THE GBCF

Given a lattice L expressed by the GBCF, the following algorithm can be employed.

Algorithm: Given $r \in \mathbb{R}^{m.n}$:

0) Set $r_0 = r$ and decode r_0 to the closest point $\hat{\bar{c}}_0 \bullet \alpha_0 + \hat{\gamma}_1$ in the lattice $\Lambda_0 = C_0 \bullet [\Gamma_0/\Gamma_1] + \Gamma_1^m$

1) Set $r_1 = r_0 - \hat{\bar{c}}_0 \bullet \alpha_0$ and decode r_1 to the closest point $\hat{\bar{c}}_1 \bullet \alpha_1 + \hat{\gamma}_2$ in the lattice $\Lambda_1 = C_1 \bullet [\Gamma_1/\Gamma_2] + \Gamma_2^m$

$\vdots$

b-1) Set $r_{b-1} = r_{b-2} - \hat{\bar{c}}_{b-2} \bullet \alpha_{b-2}$ and decode r_{b-1} to the closest point $\hat{\bar{c}}_{b-1} \bullet \alpha_{b-1} + \hat{\gamma}_b$ in the lattice $\Lambda_{b-1} = C_{b-1} \bullet [\Gamma_{b-1}/\Gamma_b] + \Gamma_b^m$.

The received point r is decoded as the lattice point $\hat{\ell} = \hat{\bar{c}}_0 \bullet \alpha_0 + \ldots + \hat{\bar{c}}_{b-1} \bullet \alpha_{b-1} + \hat{\gamma}_b$ of the lattice $L = C_0 \bullet [\Gamma_0/\Gamma_1] + \ldots + C_{b-1} \bullet [\Gamma_{b-1}/\Gamma_b] + \Gamma_b^m$.

Then the following statements are proved: 1) The generalized multistage decoding algorithm is invariant by translation; 2) For each $\ell \in L$, any point $r \in \mathbb{R}^{m.n}$ such that $|r - \ell|^2 < d_{min}^2(L)/4$, is decoded to ℓ by the generalized multistage decoding algorithm; 3) For nested linear binary block codes, $C_0 \subseteq C_1 \subseteq \ldots \subseteq C_{b-1}$, an exact expression for the effective error coefficient is given; 4) Given $r_i \in \mathbb{R}^{m.n}$, the codewords $\hat{c}_i \in C_i$ in step i of the generalized multistage decoding algorithm is the one which minimizes:

$$\sum_{j=1}^{m} (-1)^{\hat{c}_{i,j}} \cdot \mu_j^{(i)} \qquad \text{where} \quad \mu_j^{(i)} = d_{j_0}^{(i)} - d_{j_1}^{(i)}$$

We show that the number of computations necessary to decode Λ_i, $0 \le i \le b-1$ is

$$N_D(\Lambda_i) = m \cdot \left[N_D(\Gamma_i/\Gamma_{i+1}) + 1 \right] + N_D(C_i)$$

and that the overall complexity of the generalized multistage decoding algorithm can then be estimated by

$$N_D^A(L) = \sum_{i=0}^{b-1} N_D(\Lambda_i) + b.m.n$$

Finally, we determine the performance and the complexity of the proposed algorithm for several lattices L.

REFERENCE

[1] G.D. Forney Jr., "A bounded-distance decoding algorithm for the Leech lattice, with generalizations", *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 906-909, 1989.

* This research has been supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq Grant No. 301416/95-0, and Centro de Pesquisa e Desenvolvimento da Telebrás - CPqD-Telebrás under Grant No. 367/90.

A NEW BLOCK CODED MODULATION SCHEME AND ITS SOFT DECISION DECODING

Kazuhiko YAMAGUCHI†
† Department of Computer Science and Information Mathematics, The University of Electro-Communications

Hideki IMAI‡
‡ Division of Electrical and Computer Engineering, Faculty of Engineering, Yokohama National University

Introduction

Recently various schemes of coding and modulation have been proposed as efficient methods to improve the performance of digital communication systems. One of interesting approach was presented by Imai and Hirakawa in the early stage of researches on the coded modulation. Their scheme (IH-scheme) utilize component codes having different error protection capabilities are employed with step wise decoding or multistage decoding. It is noted that the component codes are designed on the basis of Hamming distance, but nonuniformity is introduced by letting their respective error protection capabilities different.

We propose a multilevel coded modulation scheme using an unequal error protection code. The basis of the scheme can be considered as an extension of the IH-scheme. Instead of using several error-correcting codes in the IH-scheme, we use a block or trellis code which has unequal error protection capability.

To obtain large coding gain from the UEP code, codeword is mapped into channel symbols by using finite memories in the scheme. Figire 1 shows the encoding and decoding block diagram of 3-level coding (i.e. 8-PSK etc.). In the figure, each of '1 'is memory unit, that is shift register of length $n/3$, where n is the code length of block code C. BCM coding based on the same structure for the 2-levels has been proposed by M. C. Lin[1]. he calls it ***block coded modulation with inter block memory***. We studied the structure from the view point of application of unequal error protection code.

We describe the minimum squared Euclidian distances of UEP-BCM and UEP-TCM by using the ***separations***; that are the measurements of error protection capability of an unequal error protection(UEP) code. Our scheme can be considered as a generalization of his scheme.

Although the error performance of UEP-BCM is described from the viewpoint of UEP, ordinary error correction codes having uniform error protection capability can be applied to UEP-BCM. UEP-BCM provides attractive coded scheme when the scheme is easy implemented by using algebraic decoding. This study deals with UEP-BCM based on RS code and BCH code.

UEP-BCM

The unequal error protection capability of an UEP code over Hamming space is described by using 'separations '. Let C be a linear (n, k) block code over a finite field, where n is the code length and k is the number of information symbols. If we define separation[1] S_i of UEP code as

$$S_i = \min_{v \in C, v_i \neq 0} (W_h(v)) \quad (i = 0, 1, \cdots, n-1), \tag{1}$$

where v_i is the i-th symbol in codeword v, and $W_h(x)$ is Hamming weight of x. It is easily proved that the squared minimum Euclidian distance of UEP-BCM based on (n, k) block code C is bounded by the following inequality:

$$D^2_{ED} \geq \min\{(S_0 \cdot \Delta^2_{i_0}), (S_1 \cdot \Delta^2_{i_1}), \cdots, (S_{n-1} \cdot \Delta^2_{i_{n-1}})\}. \tag{2}$$

Δ_i is the subset distance which is given by the set partitioning method. Define $S^2_{ED}(i)$ as

$$S^2_{ED}(i) = \min_{v \in C, v_i \neq 0} \{\sum_{j=0}^{n-1} \delta(v_j, 0) \cdot \Delta^2_j\} \tag{3}$$

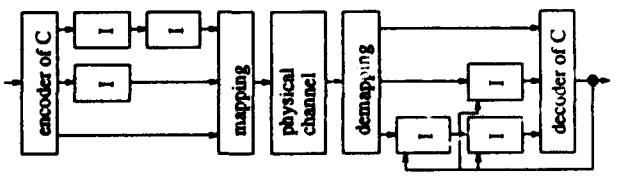


Figure 1: Block diagram of Proposed Scheme

We call $S^2_{ED}(i)$ a ***squared Euclidian separation***. Then the squared minimum Euclidian distance of the UEP-BCM is given by

$$D^2_{ED} = \min\{(S^2_{ED}(0)), (S^2_{ED}(1)), \cdots, (S^2_{ED}(n-1))\} \tag{4}$$

and that D^2_{ED} is not smaller than the bound given by 3.

Performance of UEP-BCM

Lots of BCM schemes use Viterbi decoding have been proposed. However, one of the advantages of block coding is in algebraic decoding. This viewpoint is the same as the original IH-scheme.

New soft decision decoding algorithm for proposed BCM are studied. The algorithm can performed by ordinary erasures and errors decoding after the erasure locations are determined by soft output of demodulator and decoding results of former blocks.

The decoding is repeatedly performed the erasures and errors decoding with changing the erasure locations, We can obtain better performance. The repeated case can be said generalized minimum distance decoding algorithm for UEP-BCM.

Figure 2. demonstrates the performance of proposed coded scheme and decoding algorithm obtained by numerical calculation. We evaluated the block error probabilities of (255, 170, 86) RS coded 8-PSK and (255, 171, 23) BCH coded 8-PSK, where the RS code is defined over GF(256). In the figure, these are indicated by RS and BCH respectively. The evaluation are obtained ordinary hard decision decoding as well as proposed soft decision decoding. Those are indicated by "h" and "s". For comparison, we show event error provability of Ungerboeck's trellis coded 8-PSK ($\nu = 9$) and uncoded bit error rate of QPSK.

The performances are compared with different length block size. The RS code shows extremely good performance, even it has large codelength. The erasures and errors decoding of distance 86 RS code may not be so simple, and number of transmission data more than 1Kbits/block is not suitable for applications required small delay (i.e. voice). However the result shows a solution of mass data transmission with ultimate reliability.

References

[1] M.-C. Lin : "A coded modulation scheme with interblock memory," IEEE ISIT Jan. (1990)(submitted to Trans. IT)

[2] K. Yamaguchi, R. Kohno and H. Imai : "Coded modulation based on nonuniformity in encoding and mapping," the 1991 Tirrenia International Workshop on Digital Communications, Sep. (1991).

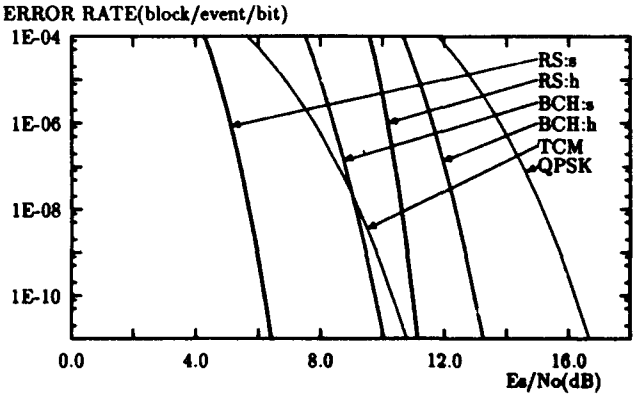


Figure 2: Error Performance of Coded 8-PSK

Correction and Interpretation of de Buda's Theorem

Tamás Linder
Department of Telecommunications
Technical University of Budapest
1521 Stoczek u. 2, Budapest, Hungary

Christian Schlegel
Digital Communications Group
University of South Australia
The Levels, SA 5095, Australia

Kenneth Zeger
Coordinated Science Laboratory
University of Illinois
1308 W. Main St., Urbana, IL 61801

ABSTRACT

De Buda's theorem states that, for asymptotically large numbers of dimensions, there exist "structured" codes which are optimal for the AWGN channel. First, we point out an error in de Buda's proof and then we correct the proof using a slightly different approach. The original erroneous proof uses thick shells of sphere bounded lattices for its optimal codes whereas we use thin annulus lattice codes for the corrected proof. We discuss the algorithmic structure of these codes as well as the implications obtained through a coding-shaping gain argument.

SUMMARY

We correct, clarify, and interpret a recent paper [1] by R. de Buda, in which he states that there exist lattice based channel codes which meet Shannon's bound for optimal codes [2]. Unfortunately, there appears to be an error with the clever proof presented by de Buda. Here, we carefully examine de Buda's proof and discuss the problems. We show that de Buda's proof can be mended, but the resulting optimal lattice code is degenerate in the sense that its "structure" appears to be lost. More precisely, the result in [1] is valid only for lattice codes whose code points lie within a thin spherical shell. Such a code resembles more a random spherical code than a lattice code.

Shannon in [2] developed tight upper and lower bounds on the error probability of optimal codes for the AWGN channel. His random coding argument used n-dimensional codes whose M_n codewords are drawn from a uniform distribution on the surface of a sphere of radius $\sqrt{nS}$ centered at the origin. Such codes have transmission rate $R = \frac{1}{n} \log M_n$.

In [1] de Buda aimed at showing that there exist structured (namely lattice based) codes for the AWGN channel that have the same near-optimal error probability properties as Shannon's "random" codes. To this end, de Buda considers an n-dimensional lattice Λ, which is translated by a vector $\bar{s}$. The bounding region of the code is a "thick" shell (or annulus), i.e., the region T between an outer sphere and an inner sphere both centered at the origin.

De Buda's main result claims that for each dimension n, there exists a lattice code of the above type with at least 2^{nR} codepoints such that its error probability $P_e(n)$ satisfies

$$P_e(n) \leq 4F_n(\theta_b, R, S/N), \tag{1}$$

where the right side is defined in [1]. This implies that essentially the same upper bounds are valid on the decrease of the error probability for rates below the channel's capacity as the ones Shannon derived for random codes.

There seems to be a technical error in [1] in the proof of (1), with important consequences and changes in the scope of the result. To correct the error we use a bounding region T which is more appropriately described as a *thin shell*.

Fortunately there is a way to modify de Buda's proof so that essentially all his steps remain valid. The conclusion, however will be somewhat different. The idea is to consider the code that results from the radial projection of the lattice code onto the *inner* sphere. In this way we get a code whose error probability is *larger* then that of the lattice code. Thus choosing the inner radius for each dimension n as $R_n = \sqrt{nS_n}$, where $S_n \uparrow S$ as $n \to \infty$, the above argument and de Buda's corrected result show that there exists a sequence of n-dimensional lattice codes with error probability $P_e(n)$ for which $P_e(n) \leq e^{-n[E(R,S/N)-o(1)]}$ holds, where $E(R, S/N) = \lim_n - \log F_n(\theta_b, R, S/N)$. This means that for rates satisfying $R_c < R < C$, de Buda's lattice codes have the same reliability exponent as that of optimal codes, and for rates below the critical rate R_c the error probability of these lattice codes has essentially the same exponential upper bond as Shannon's code.

The shell that contains the codepoints can no longer be be called a "thick shell"; the more appropriate description is "thin shell". It is worth noting, that since the function $F_n(\theta_b, R, S'/N)$ is continuous in S', by choosing the inner radius $S' < S$ close enough to S, de Buda's result guarantees the existence of an n-dimensional lattice code whose error probability is upper bounded by a quantity arbitrarily close to the upper bound for Shannon's code. However, the better this approximation is the less the thin shell bounded lattice code resembles a lattice code in the usual sense, and the more it looks like a "random" spherical code, for which Shannon originally proved the error bounds.

Were de Buda's original proof to be correct, one might argue that the class of sphere bounded lattice codes or even lattice bounded lattice codes are asymptotically optimal as the dimension of the signal constellation grows. However, this conclusion initially appears not to follow from our corrected version of the proof since the codepoints derived from the lattice are those which lie in a thin spherical shell, and specifically exclude the lattice points interior to the inner sphere. Adding these points to the code would invalidate our presented proof.

In effect, the radius of the thin spherical shell is made to be large enough that enough lattice points fall within the sphere as needed. The main advantages of structured codes such as those derived from lattices are generally that: (i) its points can be easily enumerated thus avoiding an exhaustive storage of the points, and (ii) signal decoding can be computed efficiently, using algorithms that exploit the lattice's structure. These advantages appear to be lost for the codes we used to correct de Buda's result.

However, an argument sustaining the asymptotic optimality of structured codes can be given using a coding/shaping gain approach. We give a discussion of this implication.

ACKNOWLEDGEMENTS The research was supported in part by Hewlett-Packard Co., and the National Science Foundation under Grants No. NCR-90-09766 and NCR-91-57770.

References

[1] R. de Buda, "Some optimal codes have structure," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, pp. 893-899, August 1989.

[2] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611-656, May 1959.

Decoding Lattice Partitions with application to Decoding Coset Codes

F.-W. Sun and Henk .C.A.van Tilborg

Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O.Box 513, 5600 MB Eindhoven
The Netherlands

Abstract—Several new algorithms for decoding lattice partitions are presented. They apply to Viterbi decoding of multidimensional trellis codes based on these partitions. In [1, 2], trellis-based algorithms were presented for decoding the lattice partitions. The new algorithms can achieve about 50% reduction of the complexity of decoding the lattice partitions in terms of real additions/comparisons compared with the algorithms of [1, 2]. The complexity of the resulting overall Viterbi decoding algorithms still shows a modest improvement. An algorithm for soft decision decoding the first-order Reed-Muller code $(8, 4, 4)$ or the Gosset lattice is also presented. It involves at most 17 real operations, thus, improving the best known algorithm.

Summary

A typical *multidimensional trellis coded modulation(MTCM)* scheme can be simply described by two basic ingredients: one is the cosets of a lattice partition Λ/Λ', where Λ is a lattice and Λ' is a sublattice such that the order of the partition is finite; the other is a conventional binary *convolutional code*. The output of the binary encoder chooses the coset, and some other information bits specify an element in the coset [1, Fig.1].

The trellis diagram of the resulting multidimensional trellis code is essentially the same as that of the convolutional code. The difference is that the labels on the branches of the trellis disgram of the convolutional code now correspond to cosets. Thus, a trellis-searched decoding algorithm such as the Viterbi algorithm can be used to decode a multidimensional trellis code.

In a soft-decision Viterbi-decoding algorithm, the first step of the decoding requires computing the branch metrics. This step is called *decoding the branches*. For an MTCM based on a lattice partition, decoding the branches turns out to be equivalent to *decoding the lattice partition* in use. This means that the closest points in each of the cosets to the received point has to determined and the associated metrics need to be calculated.

In [1, 2], Forney gave *trellis-based algorithms* for decoding lattice partitions. His algorithms are optimal trellis decoding for given coordinate order and alphabet among all the trellis decoding in the sense that it uses smallest number of trellis states [3]. Therefore, the expression *trellis-based algorithm* will simply stand for the kind of trellis decoding algorithms described in [2].

Certainly decoding the branches is only part of the overall decoding procedure. However, for a code whose number of states is small relative to its dimension, a considerable portion of the overall decoding work is due to decoding the branches. Furthermore, in most practical implementations, the number of trellis code states used has been very low(typically 4 or 8, occasionally 16 but rarely more than 16) [4]. Therefore, by reducing the complexity of decoding the branches, it is possible to achieve a considerable amount of reduction of the overall decoding complexity.

In this work, we present several new algorithms for decoding the lattice partitions. Most of the algorithms can achieve about 50% reduction of the complexity of decoding the branches. They result in modest improvement of the overall decoding complexity.

Most previous known efficient algorithms for soft-decision decoding block codes and lattices rely on decoding each coset of a subcode of certain type in combination with Wagner decoding rule. The Wagner decoding rule applies to binary codes whose check matrix consists of a single all-one row. It states that an entry-by-entry hard detection of the received word is to be followed, unless the number of 1 bits is already even, by inversion of the least reliable bit. Our algorithms also fall into this category.

The complexity of the decoding is measured by the total number of real additions and comparisons. Certainly the actually running time always depend to some extent on implementation technology in use. We did try, however, to evaluate all the algorithms in a uniform way. In compliance with the convention estabilished in the literature, we ignore such operations as memory addressing, negation, taking the absolute value, and multiplication by 2, as well as the checking of logical conditions and modulo 2 additions.

References

[1] G.D.Forney Jr.,"Coset codes-Part I: Introduction and geometrical classification," *IEEE Trans. on Inform. Theory* vol. IT-34,pp. 1123-1151, Sept. 1988.

[2] G.D.Forney Jr.. "Coset codes-Part II:Binary lattices and related codes," *IEEE Trans. on Inform. Theory* vol. 34,pp.1152-1187, Sept. 1988.

[3] D.J. Muder, " Minimal Trellises for Block Codes," *IEEE Trans. on Inform. Theory* vol.34, NO.5, pp1049-1083 Sept. 1988.

[4] A.J.Viterbi, J.K. Wolf,E. Aehavi, R. Padovani, " A Pragmatic Approach to Trellis-Coded Modulation," *IEEE Com. Mag.* pp. 11-19, July 1989.

Code optimisation for finite error rate

A. G. Burr and T. J. Lunn+*

Abstract

We present a new construction based on Blokh and Zyablov's generalised concatenated codes for codes and coded modulation schemes with coding gain optimised for a given decoded word error rate, rather than for asymptotic coding gain. It is shown that this may be achieved by a geometric structure for the codes in which not all neighbouring codewords are at equal distances, which implies also that the minimum distance of the code is no longer maximised. The technique may be applied to coded modulation schemes where the "inner code" is a multilevel signalling constellation, or to concatenated binary codes or codes over GF(q). The outer code may be a block or a trellis code. The technique is illustrated with reference to a block coded modulation scheme, block coded 8-PSK.

Introduction

Error correcting codes and coded modulation schemes are conventionally designed for maximum asymptotic coding gain (ACG), i.e. for maximum coding gain as signal/noise ratio $\Rightarrow \infty$ and error rate $\Rightarrow 0$. However, in a practical communication system the asymptotic coding gain is often not the prime consideration, since there is a finite error rate that may be tolerated: 10^{-8} or 10^{-9} in a telecommunications system, or 10^{-3} or poorer in a speech system. Codes designed for good ACG, and especially block or lattice codes, may well perform poorly at such error rates [1].

It is well-known [2] that ACG may be optimised by maximising the minimum distance between any pair of codewords. In the case of binary block codes the distance metric is generally the minimum Hamming distance, while for coded modulation schemes it is the minimum Euclidean distance. However, it is also clear that this technique may not necessarily yield the best coding gain at finite error rates. This paper presents a technique based on Blokh & Zyblov's generalised concatenated codes [3,4] to design codes with optimum coding gain at finite error rates.

A (block) code with maximised minimum distance corresponds to an optimally dense sphere packing in n- dimensional space [2]. Here each codeword may be represented as the centre of a sphere in the signal space surrounded by and in contact with a number n_n of other identical spheres. Hence the minimum Euclidean distance d_{min} between codewords is twice the radius of the spheres, and all n_n neighbouring codewords lie at this distance. The word error rate may be approximated using the union bound as:

$$P_e \le n_n Q\left(\frac{d_{min}}{2\sigma^2}\right)$$

where σ is the standard deviation due to noise.

Geometric structure for finite error rates

For equal minimum distance codes the asymptotic coding gain is determined by the minimum distance, and it is known that it is maximised for given rate and dimensionality. However, at finite error rates their performance is significantly affected by the number of neighbours n_n, which can be extremely large in block coded modulation schemes. Hence block coded modulation schemes with good asymptotic coding gain frequently achieve a much poorer result at practical error rates. A geometric structure that maximises coding gain at finite error rate is not known.

We therefore propose a geometric structure for optimum performance at finite error rate in which all neighbours no longer lie at the same distance. In the sphere packing model, we allow some spheres to shrink, while others grow, so that the overall number of spheres remains the same. In the codes presented, this results in a series of shells of neighbours containing different numbers of codewords at different minimum distances. The word error rate then becomes:

$$P_e = \sum_i n_i Q\left(\frac{d_i}{2\sigma^2}\right)$$

where n_i, d_i are the number of neighbours and the distance, respectively, for the i^{th} shell. The distances and number of neighbours in each of the shells may then be chosen to minimise their contribution at a particular error rate, while allowing it to increase elsewhere, where it will exceed that error rate.

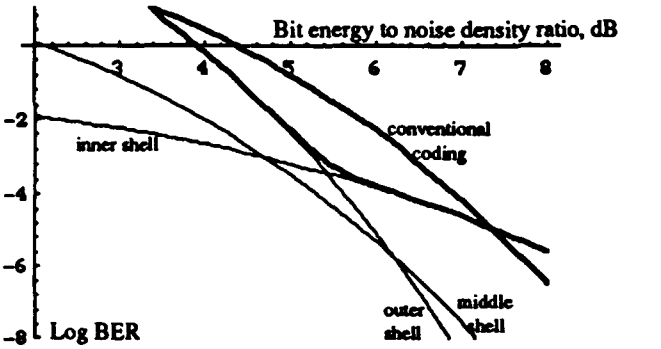


Fig. 1. BER curves (upper bound), showing contributions from shells at different distances, and comparison with a conventional scheme

Design of codes using BCM construction

The Blokh and Zyablov/Cusack construction for GCC/BCM [3,4] provides a means of constructing codes and coded modulation schemes with these characteristics. The code is defined by a matrix of codes defining the choice of inner code or constellation points from a series of partitions. The top row of the matrix chooses between partitions having the smallest minimum Euclidean distance, and therefore requires the code with the largest Hamming distance. The next row corresponds to a partition with a larger Euclidean distance, and therefore the code has smaller Hamming distance.

Conventionally the set of Hamming distances are chosen so that the effective minimum squared distance, given by the product of the Hamming and the squared Euclidean distance, is equal for each row. However, if they are chosen to give different effective distances, then the required code structure for finite error rate is created.

The diagram (Fig. 1.) shows the BER curves, using the union upper bound, for an example with three shells, optimised for coding gain at approximately 10^{-3}, showing the curves for the separate shells and the overall result. This example uses coded 8-PSK length 31, with (31,1,31) (repetition code), (31,20,6) (expurgated BCH), and (31,31,1) (uncoded) codes on the top, middle and bottom rows respectively. It can be seen that the top row contribution dominates at lower signal/noise ratio, the bottom row at higher signal/noise, while at the design error rate the three contributions are quite similar, resulting in a "bite" in the curve at this point which optimises the coding gain.

The BER curve for a conventionally designed BCM scheme, using the codes (32,7,14) (extended Goppa), (32,26,4) (extended Hamming) and (32,31,2) (even parity), is also shown in Fig. 1. for comparison. It can be seen that while this gives a significantly higher asymptotic coding gain, at BER of 10^{-3} the coding gain is poorer by more than 1 dB.

References

1. Burr, A. G, Sheppard, J. A. and Lunn, T. J. "Comparison of block and trellis coded modulation schemes" Int.Symp. on Communication Theory and Applications, Crieff, Scotland, 9-13th September 1991
2. Conway, J. H, and Sloane, N. J. A. "Sphere packings, Lattices and Groups" Springer-Verlag, New York, 1988
3. Blokh, E. L. and Zyablov, V. V "Coding of generalised concatenated codes" Prob. Peredach. Inform. vol. 10, no. 3, pp 218-22, 1974
4. Cusack, E "Error control codes for QAM signalling". Electron. Lett., 1984, vol 20, no. 2, pp 62-63

* *University of York, U. K.*
\+ *BT Laboratories, U. K.*

Evaluation of the Block Error Probability of Block Modulation Codes by the Maximum-likelihood Decoding for an AWGN Channel*

Tadao Kasami†, Toru Fujiwara††, Toyoo Takata†† and Shu Lin†††

†Graduate School of Information Science
Advanced Institute of Science and Technology, Nara
Ikoma, Nara, 630-01 Japan

††Dept. of Information and Computer Science
Osaka University
Toyonaka, Osaka, 560 Japan

†††Dept. of Electrical Engineering
University of Hawaii at Manoa
Honolulu, Hawaii, 96822 U.S.A.

Abstract

This paper is concerned with the evaluation of the block error probability P_e of a block modulation code by optimum or suboptimum soft-decision decoding for an AWGN channel. In the range D_L of low signal-to-noise ratios, it is feasible to evaluate P_e by simulation. In the range D_H of very high signal-to-noise ratios, tight upper bounds on P_e are available. However, in most cases, there remains a gap between D_L and D_H, and the value P_e for signal-to-noise ratios in this gap may be of practical interest. In general, very small values of the block error probability are required. It is infeasible to evaluate a very small P_e simply by simulation.

In this paper, the maximum-likelihood soft-decision decoding of a block modulation code for an AWGN channel is considered, and a new method of evaluating the block error probability P_e for a wide range of signal-to-noise ratios is presented. The evaluation of P_e is reduced to numerical computations and simulations on statistics with not very small mean value. Computation results for the block error probabilities of some block modulation codes are given.

Summary

Let C be a block code of length n over a finite set L of symbols. Let h be a positive integer. For $u \in L$, let $s(u)$ denote the signal point in $\mathcal{R}^h$ represented by u, where $\mathcal{R}^h$ denotes the set of all h-tuples of real numbers, and for an n-tuple $\mathbf{u} = (u_1, u_2, \cdots, u_n)$ over L, let $s(\mathbf{u})$ denote the hn-tuple $(s(u_1), s(u_2), \cdots, s(u_n))$. For $\mathbf{z}$ and $\mathbf{z}'$ in $\mathcal{R}^{hn}$, let $||\mathbf{z} - \mathbf{z}'||$ denote the Euclidean distance between $\mathbf{z}$ and $\mathbf{z}'$.

Assume that the channel is an AWGN channel and that the maximum likelihood decoding is used and every codeword is equally likely to be transmitted. For simplicity, we consider the case where $s(\mathbf{v_0})$ is transmitted for a fixed codeword $\mathbf{v_0}$. Let P_e denote the probability of incorrect decoding when $s(\mathbf{v_0})$ is transmitted. For a codeword $\mathbf{v}$ of $C - \{\mathbf{v_0}\}$, let D_e be defined as

$$D_e \triangleq \bigcup_{\mathbf{v} \in C - \{\mathbf{v_0}\}} \{\mathbf{z} \in \mathcal{R}^{hn} : ||\mathbf{z} - s(\mathbf{v})|| \leq ||\mathbf{z} - s(\mathbf{v_0})||\}. \quad (1)$$

For a nonnegative real number r, let $S(r)$ be the surface of the hn-dimensional sphere of radius r. Let $D_e(r)$ be defined by

$$D_e(r) \triangleq D_e \cap S(r). \quad (2)$$

For a surface S, area$[S]$ denotes the area of S. Let $f(r)$ denote the probability of incorrect decoding under the condition that the Euclidean distance between transmitted $s(\mathbf{v_0})$ and received hn-tuple $\mathbf{z}$ is r, and let $g(r)$ denote the density function of probability that the Euclidean distance between transmitted $s(\mathbf{v_0})$ and received $\mathbf{z}$ is r. It follows from the definitions that

$$f(r) = \text{area}[D_e(r)]/\text{area}[S(r)] \leq 1, \quad (3)$$

$$P_e = \int_0^\infty f(r)g(r)dr, \quad (4)$$

where

$$g(r) = \frac{2}{2^{\frac{hn}{2}}\sigma^{hn}\Gamma(\frac{hn}{2})} r^{hn-1} e^{-\frac{r^2}{2\sigma^2}} \quad (5)$$

and $\sigma^2 = \frac{nN_0}{2E_b \log_2 |C|}$. For a codeword $\mathbf{v}$ in $C - \{\mathbf{v_0}\}$ and a positive real number r, let $D_e(r, \mathbf{v})$ be defined as

$$D_e(r, \mathbf{v}) \triangleq \{\mathbf{z} \in \mathcal{R}^{hn} : ||\mathbf{z}|| = r \quad \text{and} \quad ||\mathbf{z} - s(\mathbf{v})|| \leq ||\mathbf{z} - s(\mathbf{v_0})||\}. \quad (6)$$

Then, the ratio area$[D_e(r, \mathbf{v})]$/area$[S(r)]$ depends only on r/δ, where $\delta \triangleq ||s(\mathbf{v}) - s(\mathbf{v_0})||$. Let $\rho(r/\delta)$ denote the ratio. Then, $\rho(r/\delta)$ is given as follows: 1) For $r/\delta \leq 1/2$,

$$\rho(r/\delta) = 0 \quad (7)$$

and 2) for $r/\delta > 1/2$ and $hn > 2$,

$$\rho(r/\delta) = \frac{\Gamma(\frac{hn}{2})}{\sqrt{\pi}\Gamma(\frac{hn-1}{2})} \int_0^\theta (\sin\varphi)^{hn-2} d\varphi \quad (8)$$

where $\theta = \cos^{-1}\frac{\delta}{2r}$, and a formula for $\int_0^\theta (\sin\varphi)^{hn-2}d\varphi$ is known.

*This research is supported by NSF Grant NCR-9115400, NASA Grant NAG 5-931 and the Ministry of Education, Japan, Grant No.(C)04650287. A preliminary version [3] was presented at the 15th Symposium on Information Theory and Its Applications.

Arrange $\{||s(\mathbf{v}) - s(\mathbf{v_0})|| : \mathbf{v} \in C - \{\mathbf{v_0}\}\}$ into the increasing sequence of real numbers without repetition, denoted $d^{(1)} < d^{(2)} < \cdots < d^{(m)}$. For $1 \leq i \leq m$, let $C^{(i)}$ be defined as

$$C^{(i)} \triangleq \{\mathbf{v} \in C : ||s(\mathbf{v}) - s(\mathbf{v_0})|| = d^{(i)}\}, \quad (9)$$

and let $A^{(i)}$ denote the number of codewords in $C^{(i)}$. Let $A^{(i)}(r)$ be the number of codewords in the following subset $C^{(i)}(r)$ of $C^{(i)}$:

$$C^{(i)}(r) \triangleq C^{(i)} - \{\mathbf{v} : \text{there exists a codeword } \mathbf{u} \text{ of } C \text{ such that (1) } ||s(\mathbf{v}) - s(\mathbf{v_0})||^2 \geq ||s(\mathbf{u}) - s(\mathbf{v_0})||^2 + ||s(\mathbf{u}) - s(\mathbf{v})||^2 \text{ and (2) the radius of the circumscribed circle of } \textit{triangles}(\mathbf{v_0})s(\mathbf{v})s(\mathbf{u}) \text{ is not smaller than } r\}.$$

Then, the following theorem [3] holds for $f(r)$.

Theorem 1: Suppose that

$$d^{(1)} = \min_{\substack{\mathbf{u},\mathbf{v} \in C \\ \mathbf{u} \neq \mathbf{v}}} ||s(\mathbf{u}) - s(\mathbf{v})||. \quad (10)$$

Then (i) for $0 < r \leq d^{(1)}/2$, $f(r) = 0$,
(ii) for $d^{(1)}/2 \leq r$, $f(r)$ increases monotonously as r increases,
(iii) for $d^{(1)}/2 < r \leq d^{(1)}/\sqrt{3}$,

$$f(r) = \sum_{i=1}^{j} A^{(i)}\rho(r/d^{(i)}), \quad (11)$$

where j is the largest index such that $d^{(j)} < (2d^{(1)})/\sqrt{3}$,
and (iv) for $d^{(1)}/\sqrt{3} < r$

$$f(r) \leq \min\{1, \sum_{i=1}^{\ell} A^{(i)}(r)\rho(r/d^{(i)})\} \leq \min\{1, \sum_{i=1}^{\ell} A^{(i)}\rho(r/d^{(i)})\}, \quad (12)$$

where ℓ is the largest index such that $d^{(\ell)} < 2r$. $\Delta\Delta$

In the range of $r > d^{(1)}/\sqrt{3}$ such that $f(r)$ is very small, the upper bound given by (12) may be used, and in the range of r such that $f(r)$ is not very small, $f(r)$ may be evaluated by simulation. Several examples show that Theorem 1 is useful. Let $\overline{f}(r)$ denote the right-hand side of (12). Then upper bound $\int_0^\infty \overline{f}(r)g(r)dr$ on P_e is better than the conventional union bound at low signal-to-noise ratios.

Let $\text{RM}_{m,j}$ be the jth order Reed-Muller code of length 2^m. The block error probability of $\text{RM}_{6,2}$ as a BPSK code is evaluated. The code has a 4-section trellis diagram with 1024 states and the value of $d^{(1)}$ is 8. Let $P_{e,s}^{(1)}$ denote the value given by (4) in which $f(r)$ is evaluated by simulation for $7 \leq r \leq 15$ and by using the right-hand side of (12) for other values of r. Direct simulations on the block error probabilities for $-1.36 \leq E_b/N_0 \leq 3.64$ were made. The simulation results are almost the same as the values of $P_{e,s}^{(1)}$ in the range of signal-to-noise ratios. The values of $P_{e,s}^{(1)}$ approach to the conventional union bounds for $4.6 \leq E_b/N_0$.

The block error probability of a basic multilevel 8-PSK code [1] of length 32is also evaluated. The component codes are $\text{RM}_{5,1}$, $\text{RM}_{5,3}$ and P_{32}, where P_{32} denotes the all even weight code of length 32. The value of $d^{(1)}$ is 2.828. Let $P_{e,s}^{(2)}$ denote the value given by (4) in which $f(r)$ is evaluated by simulation for $2.3 \leq r \leq 4$ and by using the right-hand side of (12) for other values of r. The values of $P_{e,s}^{(2)}$ are almost the same as the direct simulation results for $2.06 \leq E_b/N_0 \leq 6.06$ and approach to the conventional union bounds for $6.06 \leq E_b/N_0$.

The above method may be extended to suboptimum soft-decision multistage decoding.

References

[1] T.Kasami, T.Takata, T.Fujiwara and S.Lin, "On Multilevel Block Modulation Codes," *IEEE Trans. Inf. Theory*, Vol.37, pp.965-975, July 1991.

[2] E.R.Berlekamp, "The Technology of Error-Correcting Codes," *Proc. of the IEEE*, Vol. 68, No. 5, pp. 564-593, May 1980.

[3] T.Kasami, T.Fujiwara, T.Takata, Kenichi Tomita and S.Lin, "Evaluation of the Block Error Probability of Block Modulation Codes by the Maximum-likelihood Decoding for an AWGN Channel," *Proc. of the 15th Symposium on Information Theory and Its Applications*, pp. 161-164, Minakami, Japan, September 1992.

Common Information of Two Correlated Random Variables

Hirosuke YAMAMOTO

Department of Communications and Systems, University of Electro-Communications,
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182 Japan (E-mail: yamamoto@cas.uec.ac.jp)

Abstract Two kinds of common information are defined for two correlated random variables, and they are represented by single letter characterizations.

Summary

Let X and Y be independent, identically distributed, but mutually correlated random variables. Assume that we want to encode $(\mathbf{X}^K, \mathbf{Y}^K)$ to three codewords (V_X, V_Y, V_C), where V_X and V_Y represent each private information of $\mathbf{X}^K$ and $\mathbf{Y}^K$, respectively, while V_C represents their common information.

In the special case where we can write $X = (S_X, S_C)$ and $Y = (S_Y, S_C)$ where S_X, S_Y, S_C are mutually independent, we can easily show by encoding $\mathbf{S}_X^K$, $\mathbf{S}_Y^K$, $\mathbf{S}_C^K$, to V_X, V_Y, V_C, respectively, that there exists a code satisfying the following inequalities for any given $\delta > 0$.

$$Pr\{\mathbf{X}^K \neq G_X(V_X V_C)\} \leq \delta, \quad Pr\{\mathbf{Y}^K \neq G_Y(V_Y V_C)\} \leq \delta, \tag{1}$$

$$\frac{1}{K}(H(V_X) + H(V_C)) \leq H(X) + \delta, \frac{1}{K}(H(V_Y) + H(V_C)) \leq H(Y) + \delta \tag{2}$$

$$\frac{1}{K}(H(V_X) + H(V_Y) + H(V_C)) \leq H(XY) + \delta. \tag{3}$$

$$\frac{1}{K}H(\mathbf{X}^K|V_Y) \geq H(X) - \delta, \quad \frac{1}{K}H(\mathbf{Y}^K|V_X) \geq H(Y) - \delta. \tag{4}$$

$$\frac{1}{K}H(\mathbf{X}^K|V_X V_Y) \geq H(V_C) - \delta, \quad \frac{1}{K}H(\mathbf{Y}^K|V_X V_Y) \geq H(V_C) - \delta \tag{5}$$

where G_X and G_Y are decoders for X and Y, respectively. These conditions correspond to the following intuitive feeling.

1. (1): $\mathbf{X}^K$ and $\mathbf{Y}^K$ should be recovered from the corresponding private information and the common information.
2. (2)(3): Each private information and the common information should not include any redundancy.
3. (4): Each private information should be independent of the other information.
4. (5): The common information should convey the same amount of information about $\mathbf{X}^K$ and $\mathbf{Y}^K$.

However, for general correlated random variables X and Y, it is impossible to construct a code satisfying all (1)-(5).

Gács-Körner [1] and Wyner [2] defined two kinds of common information, say $C_{GK}(X;Y)$ and $C_W(X;Y)$, respectively, which satisfy that[3]

$$C_{GK}(X;Y) \leq I(X;Y) \leq C_W(X;Y) \leq \min(H(X), H(Y)). \tag{6}$$

$C_{GK}(X;Y)$ can be defined as the minimum rate of V_C that satisfies (1) and (2). On the other hand, $C_W(X;Y)$ can be defined as the maximum rate of V_C that satisfies (1) and (3). Hence, (3) is ignored in Gács-Körner's definition while (2) is ignored in Wyner's definition. Furthermore, the conditions (4) and (5) are not cared. In other words, the first intuitive feeling described above is emphasized in their definitions of common information.

In this paper, we define two new kinds of common information by putting emphasis on the third and forth intuitive feeling though the condition (1) is weakened to

$$Pr\{\mathbf{X}^K\mathbf{Y}^K \neq G_{XY}(V_X V_Y V_C)\} \leq \delta, \tag{7}$$

where G_{XY} is a decoder for X and Y.

The first common information $C_1(X;Y)$ is defined as follows.

$$C_1(X;Y) \triangleq \lim_{\delta \to 0} \inf_{(V_X, V_Y, V_C) \text{ satisfies (3), (4), (7)}} \frac{1}{K}H(V_C). \tag{8}$$

In other words, $C_1(X;Y)$ is the rate of the attainable minimum core V_C of $(\mathbf{X}^K, \mathbf{Y}^K)$ by removing each private information, which is independent of the other information, from $(\mathbf{X}^K, \mathbf{Y}^K)$ as much as possible.

In the second definition, we consider (V_X, V_Y) as noncommon information $V_{\overline{C}}$, and we define $C_2(X;Y)$ as follows.

$$C_2(X;Y) \triangleq \lim_{\delta \to 0} \sup_{(V_C, V_{\overline{C}}) \text{ satisfies (10)-(12)}} \frac{1}{K}H(V_C), \tag{9}$$

$$Pr\{\mathbf{X}^K\mathbf{Y}^K \neq G_{XY}(V_{\overline{C}} V_C)\} \leq \delta, \tag{10}$$

$$\frac{1}{K}(H(V_C) + H(V_{\overline{C}})) \leq H(XY) + \delta, \tag{11}$$

$$\frac{1}{K}H(\mathbf{X}^K|V_{\overline{C}}) \geq \frac{1}{K}H(V_C) - \delta, \quad \frac{1}{K}H(\mathbf{Y}^K|V_{\overline{C}}) \geq \frac{1}{K}H(V_C) - \delta. \tag{12}$$

In other words, $C_2(X;Y)$ is the rate of the attainable maximum core V_C such that if we loose V_C, then each uncertainty of $\mathbf{X}^K$ and $\mathbf{Y}^K$ becomes $H(V_C)$.

The following theorem holds for these $C_1(X;Y)$ and $C_2(X;Y)$.

Theorem

$$C_1(X;Y) = I(X;Y) \tag{13}$$

$$C_2(X;Y) = \min(H(X), H(Y)) \tag{14}$$

$$C_{GK}(X;Y) \leq C_1(X;Y) \leq C_W(X;Y) \leq C_2(X;Y) \tag{15}$$

Proof: Omitted.

The result of (13) coincides with our intuitive feeling that $I(X;Y)$ represents a kind of common information between X and Y.

On the other hand, the result of (14) does not coincide with our intuitive feeling though the definition (9) seems to be reasonable. This is caused from the fact that in addition to the common part, each private part can share the uncertainty each other by devising the encoding. As an example, consider the special case, $X = (S_X, S_C)$ and $Y = (S_Y, S_C)$, such that $S_X \in \{0, 1, \cdots, M_X - 1\}$, $S_Y \in \{0, 1, \cdots, M_Y - 1\}$, $H(S_X) = \log M_X$, $H(S_Y) = \log M_Y$, $M_X \leq M_Y$. Even for this case, $C_2(X;Y) = \min(H(X), H(Y))$ can be achieved by letting $V_{\overline{C}} = S_X \oplus S_Y$ and $V_C = (S_C, S_X)$ where $\oplus$ represents modulo M_Y summation.

References

[1] P. Gács and J. Körner, "Common Information is far less than mutual information" *Problems of Control and Information Theory*, vol.2, pp.149-162, 1973

[2] A. D. Wyner, "The Common Information of Two Dependent Random Variables", *IEEE Trans. on Inform. Theory*, vol.IT-21, no.2, pp.163-179, March 1975

[3] I. Csiszár and J. Körner, "Information Theory: Coding Theorems for Discrete Memoryless Systems", *Academic Press, Inc.*, 1981

ENTROPY AS A FUNCTION OF ALPHABET SIZE

Christoph G. Günther
Ascom Tech Ltd.
Segelhof
CH-5405 Baden, Switzerland
guenther@tech.ascom.ch

Walter R. Schneider
Asea Brown Boveri
Corporate Research
CH-5405 Baden, Switzerland
schneider@research.abb.ch

ABSTRACT

The entropy of a source with alphabet size n is between 0 and $\log n$. Frequently, it would be useful to have a more precise idea of what value should be expected. For that purpose, we compute the mean h_n and the variance σ_n^2 of the entropy with the average being taken over all probability distributions for a given alphabet size n. The largest difference between $\log n$, i.e., the value for equidistribution, and the mean h_n is obtained in the limit $n \to \infty$ and is equal to $1 - \gamma$ nats/symbol, with γ the Euler-Mascheroni constant. This has implications for the compression of data from memoryless sources. The variance σ_n^2 of many statistical systems scales such that σ_n^2/h_n goes to a constant in the limit $n \to \infty$. Surprisingly, in the present case, we have the much faster decay $\exp(h_n)\,\sigma_n^2 \to (\pi^2/3 - 3)$ as $n \to \infty$. The actual values of the variance are usually so small that the average value h_n can be substituted for the entropy in most applications.

1. INTRODUCTION

The Shannon entropy $H_n(p)$ appears as a central quantity in many communication problems. It is defined by

$$H_n(p) := -\sum_{i=1}^{n} p_i \log p_i .$$

Note that for convenience, we use natural logarithms.

The entropy $H_n(p)$ is easily evaluated for any given distribution. Often one faces, however, the problem of a typical value, i.e., a value that is typical for a given alphabet size n. Let us assume that we can compute the mean and variance of $H_n(p)$ with respect to $p = (p_1, \ldots, p_n)$, where p runs over all probability distributions. Then the mean value could be seen as being a typical value, whenever the variance is small.

The average which we consider is an unweighted average over all probability distributions, i.e., we assume that all probability distributions p from

$$S_n := \left\{ (p_1, \ldots, p_n) : \quad p_i \geq 0, \quad \forall i, \quad \sum_{j=1}^{n} p_j = 1 \right\}$$

are equally likely. The mean h_n and the variance σ_n^2 of the entropy $H_n(p)$ then become

$$h_n := \frac{1}{|S_n|} \int_0^1 dp_1 \ldots \int_0^1 dp_n \, \delta(p_1 + \ldots + p_n - 1) H_n(p),$$

$$\sigma_n^2 := \frac{1}{|S_n|} \int_0^1 dp_1 \ldots \int_0^1 dp_n \, \delta(p_1 + \ldots + p_n - 1)(H_n(p) - h_n)^2,$$

where δ is the Dirac distribution and where

$$|S_n| := \int_0^1 dp_1 \ldots \int_0^1 dp_n \, \delta(p_1 + \ldots + p_n - 1),$$

denotes the volume of S_n.

2. RESULTS

The computations are summarized as follows:

Theorem 1 *For any $n \geq 2$, the mean h_n and variance σ_n^2 of the entropy, when averaged over S_n are given by*

$$1) \qquad h_n = \sum_{k=2}^{n} \frac{1}{k} \tag{1}$$

$$2) \qquad \sigma_n^2 = \sum_{k=2}^{n} \frac{1}{k^2} + \frac{(n-1)}{(n+1)}\left(1 - \frac{\pi^2}{6}\right). \tag{2}$$

As an immediate consequence of the first part of the theorem, we have

Corollary 2 *Define $\Delta_n := \log n - h_n$ then for any $n \geq 2$:*

$$1) \qquad \log\left(\frac{n+1}{2}\right) < h_n < \log n \tag{3}$$

$$2) \qquad \Delta_{n+1} > \Delta_n \tag{4}$$

$$3) \qquad \lim_{n \to \infty} \Delta_n = 1 - \gamma, \tag{5}$$

where $\gamma = 0.5772156649015329\ldots$ denotes the Euler-Mascheroni constant, i.e., the constant that appears in Euler's infinite product representation of the Gamma function.

Corollary 2 tells that the discrepancy of the average entropy with respect to the value obtained for equidistribution increases monotonically to the value $1 - \gamma$ nats/symbol. The relative loss $(\log n - h_n)/h_n$ goes to 0. For a source with entropy h_n, this implies that coding can not reduce the average code word length of a memoryless source by more than 0.61 bits/symbol ($(1-\gamma)/\log 2 < 0.61$), whatever the size of n is. The relative gain of source coding goes to 0 with an increasing size of the alphabet.

Consider n independent random variables $\xi_1, \ldots, \xi_n$ that are identically distributed. If that distribution has a mean and a variance, say μ and σ^2, *then* $\sum_{i=1}^{n} \xi_i$ *has the mean* $\mu_n = n\mu$ *and the variance* $\sigma_n^2 = n\sigma^2$. Thus, the variance increases proportionally to n, i.e., proportionally to μ_n. This is what we are typically used to. In the present case, we have a rather different behaviour:

Corollary 3

1) σ_n is monotonically decreasing for $n \geq 3$.
2) $\lim_{n \to \infty} n\sigma_n^2 = \frac{\pi^2}{3} - 3$.

Corollary 3 means that the system literally freezes in an average behaviour. The variance goes exponentially fast to 0, in the sense that $\lim_{n \to \infty} \exp(h_n)\sigma_n^2 =$ constant. From a practical point of view this implies that the entropies of distributions with a large value of n need not be computed: they are close to the average value with a high probability. This is made more precise in the following table

n	h_n	σ_n
2	0.5	0.187
4	1.083	0.191
8	1.718	0.161
16	2.381	0.124
32	3.058	0.091
64	3.744	0.066
128	4.433	0.047

3. CONCLUSION

We have found a typical value h_n for the entropy of any source with an alphabet of n letters. This value deserves its name in the sense that $\sigma_n^2 \sim (\pi^3/3 - 3)\exp(-h_n)$ in the limit $n \to \infty$. The difference between the entropy for equidistribution $\log n$ and the mean h_n increases monotonically with $n \to \infty$ but is bounded by $1 - \gamma < 0.61 \log 2$ nats/symbol. *Thus, 0.61 bits/symbol is the maximal coding gain that can typically be expected for memoryless sources.* This shows that data compression strongly relies on memory.

GENERALIZING FANO'S INEQUALITY

Te Sun Han
Dept. Information Systems
Senshu University
Kawasaki 214, Japan

Sergio Verdú
Dept. Electrical Eng.
Princeton University
Princeton, NJ 08544

One of the most useful results in the Shannon theory is the lower bound on mutual information due to Fano

Theorem 1. Suppose that X and Y are random variables that satisfy:

a) X and Y take values on the same finite set with cardinality M

b) either X or Y is equiprobable.

Then,

$$I(X;Y) \geq P[X=Y] \log M - h(P[X=Y]) \tag{1}$$

where h is the binary entropy function.

The purpose of this paper is to give a more general version of the lower bound in Theorem 1 by dropping its assumptions.

The restriction that X and Y take values on the same set is made throughout for convenience in expressing the results. It is easy to see from the mutual information data processing theorem that it can be lifted by replacing $P[X=Y]$ by $P[X=\phi(Y)]$ where ϕ is an arbitrary function mapping the space of Y to the space of X. The assumption that at least one of the random variables is equiprobable is a nontrivial restriction, which we want to eliminate.

The power of Theorem 1 stems from its ability to lower bound the mutual information between two random variables in terms of a single parameter computable from their joint distribution: the probability that the random variables take the same value. Since it is possible to construct independent (nonequiprobable) random variables (X,Y) for any arbitrarily specified $P[X=Y]$, it is apparent that dropping assumption b) of Theorem 1 will require a lower bound that depends on the distribution of X and Y not only through $P[X=Y]$, but through some other, hopefully simple, quantity.

Consider the following result.

Theorem 2. Define the binary divergence function $d(x\|y)$ as the divergence between the two-mass distributions $(x, 1-x)$ and $(y, 1-y)$. If X and Y take values on the same set, then

$$I(X;Y) \geq d(P[X=Y] \,\|\, P[X=\bar{Y}]), \tag{2}$$

where $\bar{Y}$ is independent of X and has the same distribution as Y. Furthermore, equality holds in (2) if and only if

$$P_{XY}(x,y) = \begin{cases} \alpha P_X(x)P_Y(y) & x = y \\ \beta P_X(x)P_Y(y) & x \neq y \end{cases} \tag{3}$$

This work was supported in part by the Office of Naval Research under Grant N00014-90-J-1734.

Note that

$$P[X=\bar{Y}] = \sum_{\omega \in \Omega} P_X(\omega)P_Y(\omega) \tag{4}$$

i.e., the inner product between the marginals of X and Y which, in many cases is easy to obtain from the description of X and Y.

Condition (3) implies that the marginals are either nonoverlapping or both equiprobable.

We will now loosen (2) by applying the following lower bound on binary divergence

$$d(x\|y) \geq x \log \frac{1}{y} - h(x) \tag{10}$$

to Theorem 2, resulting in the following generalization of Theorem 1:

Theorem 3. If X and Y take values on the same set, then

$$I(X;Y) \geq P[X=Y] \log \frac{1}{P[X=\bar{Y}]} - h(P[X=Y]) \tag{11}$$

$$\geq P[X=Y] \log \frac{1}{\max_{\omega \in \Omega} P_X(\omega)} - h(P[X=Y]) \tag{12}$$

where by symmetry we can replace $\max_{\omega \in \Omega} P_X(\omega)$ by $\max_{\omega \in \Omega} P_Y(\omega)$.

It is tempting to strengthen the lower bound in Theorem 3 with

$$I(X;Y) \geq P[X=Y]\, H(X) - h(P[X=Y]). \tag{!?}$$

However, counterexamples to (!?) can be found. It is possible to modify the incorrect bound (!?) in terms of entropy and obtain the following result.

Theorem 4. Assume that X and Y take values on the same set and denote

$$p = \inf_{\omega \in \Omega} P[X=Y \,|\, X=\omega] = \inf_{\omega \in \Omega} P_{Y|X}(\omega|\omega) \tag{13}$$

Then,

$$I(X;Y) \geq p\, H(X) - h(P[X=Y]) \tag{14}$$

If, in addition, $p > 1 - \frac{1}{e}$, then

$$I(X;Y) \geq p\, H(X) - h(p). \tag{15}$$

Relations Between Entropy and Error Probability*

Meir Feder † Neri Merhav ‡

Abstract

The relation between the entropy of a discrete random variable and the minimum attainable probability of error made in guessing its value is examined. While Fano's inequality provides a tight *lower* bound on the error probability in terms of the entropy, we derive a converse result - a tight *upper* bound on the minimal error probability in terms of the entropy. As a consequence of this relation, a channel coding theorem for the equivocation is presented. At a rate $R < C$, where C is the channel capacity, it follows straightforwardly from the classical channel coding theorem and the bounds above that the equivocation can be made arbitrarily small (exponentially fast with the block length). This result is proved directly for DMC's, and from this proof it is further concluded that for $R \geq C$ the equivocation achieves its minimal value of $R-C$ at the rate of $n^{-1/2}$, where n is the block length.

In this work we explore the relationship between the entropy of a random variable and the minimal error probability in guessing its value. The well known Fano inequality [1] provides a tight *lower* bound on the error probability in terms of the entropy. We derive a converse result - a tight *upper* bound on the minimal error probability in terms of the entropy.

Specifically, denote by $\pi(X) = 1 - \max_x p(x)$ the minimal error probability associated with the random variable X and by $\pi(X|Y) = \int dP(y)[1 - p(\hat{x}|y)]$, where $\hat{x} = \hat{x}(y) = \arg\max_x p(x|y)$, the MAP error probability given an observation Y. Similarly, denote by $H(X)$ and $H(X|Y)$ the entropy and the conditional entropy (equivocation) respectively. Fano's inequality states that

$$H \leq \Phi(\pi) = h(\pi) + \pi \log(M-1), \tag{1}$$

where M is the size of the alphabet of X. We have shown a converse result

$$H \geq \phi^*(\pi) \tag{2}$$

where

$$\phi^*(\pi) = a_i\left(\pi - \frac{i-1}{i}\right) \quad \frac{i-1}{i} \leq \pi \leq \frac{i}{i+1}, \quad i = 1,\ldots,M-1, \tag{3}$$

and $a_i = i(i+1)\log\frac{i+1}{i}$. The region in the $\pi - H$ plane determined by inequalities (2) and (3) is depicted in Figure 1 for the case $M = 8$. The bounds above hold for $\pi(X)$ and $H(X)$, as well as for $\pi(X|Y)$ and $H(X|Y)$, and it can be shown that both bounds are sharp, i.e. each point on the bounds can be achieved with equality. We note that a weaker bound $H \geq 2\pi$, which coincides with (3) only at $0 \leq \pi \leq 1/2$ has been observed in, e.g., [2] and [3] pp.520-521.

To get the bound (3) we first calculated a function $\phi(\pi)$ which is the minimal entropy for each given value of error probability of a single random variable; the bound $\phi^*(\pi)$ is the largest convex function that is smaller or equal to $\phi(\pi)$.

It is interesting to note that the bounds above affirm the intuition that a random variable is totally random (i.e. $H = \log M$) iff it is totally unpredictable (i.e. its minimal error probability is $(M-1)/M$) and conversely, a random variable it totally redundant (i.e. its entropy is zero) iff it is fully predictable (its minimal probability of error is zero).

The relation above between entropy and error probability leads to a statement of the channel coding theorem in terms of the equivocation. As one immediately observe, the fact that zero equivocation is achieved iff zero error probability is achieved and the classical channel coding theorem imply that the equivocation of the codebook can be made arbitrarily small (exponentially fast with the block length) provided that $R < C$. It turns out that this observation can be proved directly, at least for DMC's. This proof provides an insight on the behavior of the equivocation at $R \geq C$.

Specifically, by applying a random coding upper bound *directly* to the equivocation, using techniques similar to Gallager's derivation of the coding theorem [3], it is shown that

$$H(\underline{X}|\underline{Y}) \leq \left(1 + \frac{1}{\rho}\right) 2^{-n[E_0(\rho,q) - \rho R]} \tag{4}$$

where $H(\underline{x}|\underline{Y})$ is the equivocation of the codebook,

$$E_0(\rho,q) = -\log \sum_y \left[\sum_x q(x)p(y|x)^{1/(1+\rho)}\right]^{1+\rho},$$

and where we identify $\max_{0\leq\rho\leq 1}[E_0(\rho,q) - \rho R]$ as the random coding exponent which is strictly positive as long as $R < C$.

The inequality (4) holds for any value of R. This bound on the equivocation is always useful, unlike the random coding bound on the error probability which becomes useless at it exceeds 1. When $R = C$ we find that the optimal ρ approaches zero, and by Taylor expansion of $E_0(\rho,q)$ about $\rho = 0$ we get

$$H(\underline{X}|\underline{Y}) \leq \alpha\sqrt{n}, \tag{5}$$

where $\alpha > 0$ is some constant.

Using (5) it is further easy to see that when $R > C$ there exists a codebook whose equivocation satisfies

$$\frac{1}{n}H(\underline{X}|\underline{Y}) \leq R - C + O\left(n^{-1/2}\right). \tag{6}$$

Since always $H(\underline{X}|\underline{Y}) \geq H(\underline{X}) - n \cdot \max_q I(\underline{X};\underline{Y}) = nR - nC$, we conclude that the equivocation, per input symbol, can be made exactly $R - C$, at a rate $O(n^{-1/2})$.

We finally note that other techniques for bounding the error probability can be used for bounding the equivocation. For example, it can be shown directly that the expurgated error exponent, which at low rates provides better exponent than the random coding exponent, is applied to the equivocation.

[1] R. Fano. Class notes for the course 6.574, transmission of information, Massachusetts Institute of Technology, 1952.

[2] M. E. Hellman and J. Raviv. "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Information Theory*, IT-16:368–372, July, 1970.

[3] R. G. Gallager. *Information Theory and Reliable Communications*. Wiley, New York, N.Y., 1968.

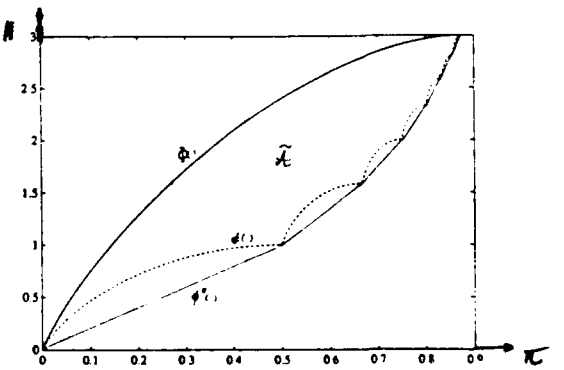


Figure 1: The Functions $\Phi(\alpha)$, $\phi(\alpha)$ and $\phi^*(\alpha)$ and the allowable region in the $\pi - H$ plane

*This research was supported in part by the Wolfson Research Awards administrated by the Israel Academy of Science and Humanities, at Tel-Aviv University.

†Meir Feder is with the Department of Electrical Engineering - Systems, Tel-Aviv University, Tel-Aviv, 69978, ISRAEL

‡Neri Merhav is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, 32000, ISRAEL

Generalized Cutoff Rates and Rényi's Information Measures

I. Csiszár (Budapest)

Abstract

Rényi's entropy and divergence of order α are given operational characterizations in terms of block coding and hypothesis testing, as so-called β-cutoff rates, with $\alpha = 1/1+\beta$ for entropy and $\alpha = 1/1-\beta$ for divergence. Out of several possible definitions of mutual information of order α (for channel W and input distribution P) we adopt

$$I_\alpha(P,W) = \min_Q \sum_x P(x) D_\alpha(W(\cdot|x)||Q).$$

This admits interpretation as a β-cutoff rate, with $\alpha = 1/1-\beta$ (at least for $\alpha \geq 1/2$), and so does $\max_P I_\alpha(P,W)$, the "Rényi capacity."

Geometrically, the β-cutoff rate for a discrete memoryless source or channel is the r-axis intercept of the tangent of slope β to the curve $e(r)$, where $e(r)$ is the exponent of the probability of error resp. of correct decoding for the best codes of rate r, according as r is an achievable rate or not. The ordinary cutoff rate of a DMC is the β-cutoff rate with $\beta = -1$. The β-cutoff rate for hypothesis testing has a similar geometric representation, $e(r)$ being the exponent of convergence of the probability of type 2 error to 0 or 1, for the best tests of sample size $n \to \infty$ with probability $\exp(-nr)$ of type 1 error.

Summary

Rényi [1] introduced a one-parameter family of information measures. His entropy of order α is

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \sum_x P^\alpha(x) \qquad (\alpha \neq 1) \tag{1}$$

and the divergence of order α is

$$D_\alpha(P||Q) = \frac{1}{\alpha-1} \log \sum_x P^\alpha(x) Q^{1-\alpha}(x) \qquad (\alpha \neq 1). \tag{2}$$

In the limit $\alpha \to 1$, the standard Shannon entropy and Kullback-Leibler divergence are recovered. Shannon's mutual information has several equivalent definitions whose "order α" extensions are no longer equivalent. Our results support the following definition of mutual information of order α, for a channel W with input distribution P:

$$I_\alpha(P,W) = \inf_Q \sum_x P(x) D_\alpha(W(\cdot|x)||Q). \tag{3}$$

Rényi's information measures enter useful error probability bounds, as observed by several authors. Still, few results are available that actually identify operationally defined quantities with information measures of order α. One such result, due to Campbell [2], characterizes entropy of order α in terms of exponential mean length of variable length codes.

In this paper, we give operational characterizations of Rényi's information measures (1), (2), (3) in terms of block codes and hypothesis tests, analogous to the familiar characterizations of the standard information measures. To this end, we introduce the concept of β-cutoff rates, generalizing the well-known concept of cutoff rate of a DMC that corresponds to $\beta = -1$.

I. Csiszár is with the Mathematical Institute of the Hungarian Academy of Sciences, H-1364 Budapest, POB 127, Hungary.

This research was supported by the Hungarian National Foundation for Scientific Research, Grant 1906.

Definition. The β-cutoff rate is

(i) for source coding: the smallest resp. largest s such that for every $r > 0$, the best codes of block-length n and rate $\leq r$ have probability of error

$$p_e < \exp\{n\beta(s-r) + o(n)\} \qquad \text{resp.}$$
$$p_e > 1 - \exp\{n\beta(s-r) + o(n)\}, \tag{4}$$

according as $\beta > 0$ or $\beta < 0$.

(ii) for hypothesis testing: the largest resp. smallest s such that for every $r > 0$, the best tests of sample size n and probability of type 1 error $\leq \exp(-nr)$ have type 2 error p_e satisfying (4), according as $\beta < 0$ or $\beta > 0$.

(iii) for channel coding: the largest resp. smallest s such that for every $r > 0$, the best codes of sample size n and rate r have average probability of error p_e satisfying (4), according as $\beta < 0$ or $\beta > 0$.

(iv) for channel coding with a fixed input distribution P: same as in (iii), but the codes are required to have codewords of the same type, approaching P as $n \to \infty$.

Theorem

(i) For a DMS with distribution P, the β-cutoff rate equals the Rényi entropy (1), with $\alpha = 1/1+\beta$, for all $\beta > -1, \beta \neq 0$.

(ii) For testing a simple hypothesis P against a simple alternative Q, the β-cutoff rate equals the Rényi divergence (2), with $\alpha = 1/1-\beta$, for all $\beta < 1, \beta \neq 0$.

(iii) For a DMC $\{W\}$, the β-cutoff rate with fixed input distribution P equals the Rényi mutual information (3), and the β-cutoff rate equals $\max_P I_\alpha(P,W)$, with $\alpha = 1/1-\beta$, for all $-1 \leq \beta \leq 1, \beta \neq 0$.

Remark. The "Rényi capacity" $\max_P I_\alpha(P,W)$ can be alternatively represented as

$$\max_P \frac{\alpha}{\alpha-1} \log \sum_y \left(\sum_x P(x) W^\alpha(y|x) \right)^{\frac{1}{\alpha}}$$

or as "information radius of order α"

$$\min_Q \max_x D_\alpha(W(\cdot|x)||Q).$$

This Theorem is a straightforward consequence of well known results on error exponents available, e.g., in Csiszár and Körner [3]. The reason the author still considers this Theorem remarkable is that it appears to be the first natural and unified operational characterization of Rényi's information measures.

References

[1] Rényi, A., On measures of entropy and information, in *Proc. 4th Berkeley Symp. Math. Statist. Probability, Vol. 1*, University of California Press, Berkeley, 1961, pp. 547-561.

[2] Campbell, L.L., A coding theorem and Rényi's entropy, *Information and Control*, Vol. 8, pp. 423-427, 1985.

[3] Csiszár, I. and Körner, J., *Information Theory: Coding Theorems for Discrete Memoryless Systems*, New York: Academic, 1981.

A Generalization of the Entropy Power Inequality with Applications to Linear Transformation of a White-Noise

Ram Zamir and Meir Feder

Dept. of Electrical Engineering - Systems, Tel-Aviv University
Tel-Aviv, 69978, ISRAEL

Abstract

We prove a generalisation of the Entropy-Power Inequality, and of the Fisher-Information Inequality, to multi-dimensional linear transformation of a vector with independent components, and use this generalisation in several applications.

Consider the (joint-) differential-entropy $h(A\underline{x})$, of a linear transformation $\underline{y} = A\underline{x}$, $\dim A = m \times n$, where $\underline{x} = x_1 \dots x_n$ is a continuous random vector and $h(\underline{y}) \triangleq E\{-\log f(\underline{y})\}$. In some cases, this entropy is easily calculated or bounded. If A is an invertible matrix, the linear transformation just scales and shuffles $\underline{x}$, thus the entropy is only shifted, i.e., $h(A\underline{x}) = h(\underline{x}) + \log|A|$, where $|\cdot|$ denotes absolute value of determinant. If A does not have a full row-rank, then $h(A\underline{x}) = -\infty$, since there is a deterministic relation between the components of $\underline{y}$. If $\underline{x} = \underline{x}^*$ is a Gaussian vector, the linear transformation A preserves the normality and so $h(A\underline{x}^*) = \frac{m}{2}\log(2\pi e|AR_xA^t|^{\frac{1}{m}})$, where R_x is the covariance matrix of $\underline{x}^*$.

In the above three cases $\underline{x}$ was an arbitrary random vector. In what follows we restrict $\underline{x}$ to have independent components. Suppose in addition that y is scalar, i.e., $y = a_1x_1 + \dots + a_nx_n$, then the entropy-power-inequality (EPI) can be used to lower bound its entropy. Specifically, by one of the equivalent forms of the EPI (see e.g. [1]),

$$h(\underline{a}^t\underline{x}) \geq h(\underline{a}^t\tilde{\underline{x}}) \tag{1}$$

where $\tilde{\underline{x}}$ is a Gaussian vector with independent components, such that $h(\tilde{x}_i) = h(x_i)$, $i = 1 \dots n$ and $\underline{a}^t = (a_1, \dots a_n)$. Note that an explicit calculation of the entropy in the RHS of (1) yields

$$h(\underline{a}^t\tilde{\underline{x}}) = \frac{1}{2}\log 2\pi e(\underline{a}^tP\underline{a}) = \frac{1}{2}\log 2\pi e(\sum_{i=1}^{n} a_i^2p_i) \tag{2}$$

where P is the covariance matrix of $\tilde{\underline{x}}$, i.e., it is a diagonal matrix whose diagonal values are the entropy powers $p_i = \frac{1}{2\pi e}2^{2h(x_i)}$. The inequality in (1) becomes equality iff $\underline{x}$ is Gaussian (or if $n = 1$).

In this paper, we generalize the lower bound (1) above, to the case where $\underline{y}$ may be a vector, and show:

Theorem 1 *For any matrix A and a vector $\underline{x}$ with independent components,*

$$h(A\underline{x}) \geq h(A\tilde{\underline{x}}) = \frac{m}{2}\log(2\pi e|APA^t|^{\frac{1}{m}}) \ . \tag{3}$$

Equality in (3) holds if $\underline{x}$ is Gaussian or if, after omitting the all-zero columns, A becomes invertible. When $\underline{x}$ has i.i.d. components and A is orthonormal (i.e., $|AA^t| = 1$), (3) becomes $\frac{1}{m}h(A\underline{x}) \geq h(x)$. One implication of this result can be interpreted as an increase of the entropy per degree of freedom after band-pass filtering of a white process. As the entropy becomes higher, the random vector becomes more Gaussian. A specific statement of this phenomena is given by the following application of Theorem 1:

Theorem 2 *For any matrix A and a vector $\underline{x}$ with independent components,*

$$\frac{1}{m}\mathcal{D}(A\underline{x}; A\underline{x}^*) \leq \max_{i=1\dots n} \mathcal{D}(x_i; x_i^*) \tag{4}$$

where m = Rank A, $\mathcal{D}(\underline{y}; \underline{y}^*) = \int f_{\underline{y}} \log \frac{f_{\underline{y}}}{f_{\underline{y}^*}}$ is the Kullback-Leibler-Distance (KLD) (or "information divergence") between $\underline{y}$ and $\underline{y}^*$, and $\underline{y}^*$ denotes a Gaussian vector with the same first and second moments as $\underline{y}$.

The mutual-information between any pair of orthogonal projections of a white Gaussian vector is zero (since they are independent). For the non-Gaussian case, we use Theorem 2 to prove:

Theorem 3 *Let $\underline{x} = x_1 \dots x_n$ be a vector with i.i.d. samples and let $A_l\underline{x}$ and $A_h\underline{x}$ be two orthogonal projections of $\underline{x}$ such that Rank $A_l = r$ and Rank $A_h = n - r$, then*

$$\frac{1}{r}I(A_l\underline{x}; A_h\underline{x}) \geq \mathcal{D}(x; x^*) - \frac{1}{r}\mathcal{D}(A_l\underline{x}; A_l\underline{x}^*) \tag{5}$$

where $\frac{1}{r}I(A_l\underline{x}; A_h\underline{x})$ is the mutual-information (per sample of $A_l\underline{x}$) between the two projections.

Note that if $\frac{1}{r}\mathcal{D}(A_l\underline{x}; A_l\underline{x}^*) \approx 0$ (for large enough n), i.e., $A_l\underline{x}$ approaches normality in a KLD sense, this mutual-information is lower bounded by the (positive) KLD of x. A simple example, for $r = 1$, is $A_l\underline{x} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i$ (the D.C. component of $\underline{x}$), where, by the strong form of the Central-Limit-Theorem of [2], $\mathcal{D}(A_l\underline{x}; A_l\underline{x}^*) \to 0$ as $n \to \infty$ (actually, for "nice" distributions the KLD decreases rapidly with n). Observe that the mutual information between the orthogonal projections of the non-Gaussian white noise, is bounded away from zero, in this example.

Motivated by the duality between EPI-type inequalities for various information theoretic measures (see [1]), an inequality analogous to (3) is derived for Fisher-Information matrix. Let $K(\cdot) = \int \frac{1}{f}\nabla f \cdot \nabla f^t$ denotes the Fisher-Information matrix, with respect to a translation parameter of a random vector with a density f, where ∇f is the gradient vector of f. Then,

Theorem 4

$$K(A\underline{x}) \leq K(A\hat{\underline{x}}) = \left(AK^{-1}(\underline{x})A^t\right)^{-1} \tag{6}$$

where $\hat{\underline{x}} = \hat{x}_1 \dots \hat{x}_n$ is a Gaussian vector with independent components, such that $K(\hat{x}_i) = K(x_i)$, $i = 1 \dots n$.

The matrix inequality (6) is in the sense that the difference matrix is positive semi-definite. Equality holds under the same conditions as in theorem 1.

References

[1] A. Dembo, T.M.Cover, and J.A.Thomas. Information theoretic inequalities. *IEEE Trans. Information Theory*, IT-37:1501-1518, Nov. 1991.

[2] A.R. Barron. Entropy and the central limit theorem. *The Annals of Probability*, 14, No. 1:336-342, 1986.

RATE-DISTORTION COMPUTATION AND STATISTICAL PHYSICS

Kenneth Rose
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

A new approach to rate-distortion computation and analysis is suggested in this work. We shall restrict our attention here to continuous amplitude memoryless sources. Much of the existing theory is formulated in terms of optimization over the output density, particularly the results leading to the Blahut algorithm. In the new approach we consider a mapping from the unit interval with the Lebesgue measure, to the output space. Instead of optimizing the output density directly, we optimize this mapping. The theoretical equivalence of the mapping approach (MA) to the traditional approach is intuitively obvious but can be formally shown by isomorphism theorems for topological measure spaces. Although equivalent in principle, the MA formulation is different, and by deriving the results from this angle, some new insights are gained, as well as a more efficient numerical approach to compute the rate-distortion function.

First, the mapping approach is presented and its equivalence to the usual approach is discussed. The optimality conditions are derived and are shown to be random-coding relatives of the Lloyd optimality conditions for optimal quantizer design.

Next, the MA formulation is used to prove that, for the squared error distortion, *the optimizing output density is purely discrete as long as the rate-distortion function has not merged with the Shannon lower bound.* In other words, except for the case that the bound is attained (e.g., Gaussian source for all positive rates), the output density consists of singularities. This could explain why explicit expressions for the rate-distortion function are so hard to obtain. In a paper that recently came to my attention [1], it is shown (using a different approach) that the optimizing output is discrete if the source density's support is not the entire space. This result is a special case of our result here, as for such sources the Shannon lower bound is strictly lower than the rate-distortion function at all nonzero distortions.

We then address the analysis of the evolution of the optimizing output densities as we decrease the distortion (that is, as we "crawl up" the rate-distortion curve). Here we start by showing that the functional that is minimized to find the optimizing density is the free energy of an appropriately defined statistical mechanics system. The slope parameter is simply related to the temperature in the physical analogy, and the optimizing output density is given by the isothermal equilibrium distribution at the given temperature. Thus, "crawling up" the rate-distortion curve is simply an annealing process in statistical mechanics. The analysis shows that the annealing process starts with a single output symbol at $R = 0$, and consists of a sequence of phase transitions which increase the number of symbols (or singularities) by splitting them. It is shown that the last phase transition occurs when the rate-distortion curve hits the Shannon lower bound, and where the singularities split, or rather, explode into a continuous distribution.

Finally, we discuss the applicability of the mapping approach to practical computation of rate-distortion functions. Discretization for numerical computation results in an algorithm whose performance differs from that of the Blahut algorithm (BA). BA optimizes over a grid of points in the output space to obtain an approximate solution (whose quality depends on the resolution of the grid). MA uses the mapping which adapts its effective grid to the source distribution and so is more efficient. Moreover, as long as the Shannon lower bound is not attained, the optimal density is discrete (and usually finite) so that few variables allow MA to find the exact solution, which BA approximates using the entire grid. Note also that once the Shannon lower bound is attained we can explicitly derive the solution, so numerical evaluation is no longer necessary. The MA based algorithm is closely related to our VQ design method [2]. Another relative is [3] where the derivation is constrained to a given alphabet size. The MA method allows the number of symbols to grow as necessary to obtain the unconstrained rate-distortion result.

Acknowledgements:

I am grateful to Robert M. Gray for prompting me to find deeper relations between the deterministic annealing approach to VQ and rate-distortion theory.

This work is supported by the Engineering Foundation with the cooperation of IEEE, grant RI-A-92-12.

REFERENCES

[1] S. L. Fix, "Rate distortion functions for squared error distortion measures," in *Proc. 16'th Annual Allerton Conf. on Commun., Contr. and Comput.*, Oct. 1978.

[2] K. Rose, E. Gurewitz, and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Transactions on Information Theory*, vol. 38, pp. 1249–1257, July 1992.

[3] W. A. Finamore and W. A. Pearlman, "Optimal encoding of discrete-time continuous-amplitude memoryless sources with finite output alphabets," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 144–155, Mar. 1980.

Zipf's Law and Information Complexity in an Evolutionary System

L.B.Levitin (Boston University, Boston, USA) and
B.Schapiro (NMI Reutlingen, Germany)

Zipf's law is a famous empirical law that is observed in the behavior of many complex systems of surprisingly different nature. Zipf [1] found a remarkable rank-frequency relationship in linguistics. If we consider a long text and assign ranks to all words that occur in the text in the order of decreasing frequencies, then the frequency f_r of a word satisfies the empirical law: $f_r = cr^{-\beta}$, where c and β are constants and $\beta \approx 1$. Zipf's law has been discovered independently in such diverse situations as distribution of biological species, distribution of income, distribution of city populations, etc.[2]

Most theoretical explanations of Zipf's law are based on the principle of the "least effort", "minimum cost" [3], "minimum energy" [4], or on other very specific assumptions which, in our opinion, call for further explanations.

This paper presents a model of the development of an evolutionary system in a form of a nonstationary branching Markov random process. We will formulate the model in the language of evolutionary dynamics, though it can be reformulated in terms of demography, linguistics, etc.

Consider an ecosystem consisting of populations $N_i(N)$ (i=1,2,...,A(N)) of species s_i, where A(N) is the number of different species at the N-th step of the process. The ecosystem is assumed to evolve according to the following rules:

1) At the (N+1)th step of the process an individual of species s_i is born with probability $\Pr\{N_i(N+1) = n_i + 1 \mid N_i(N) = n_i\} = \frac{1-c(N)}{N} n_i$ (1)

Here $N_i(N)$ is the random variable which is the population of species s_i at time N, i=1,2,...,A(N).

2) The probability that an individual of a new species $s_{A(N)+1}$ will be born at the (N+1)th step of the process (probability of a successful mutation) is

$$\Pr\{N_{A(N)+1}(N+1) = 1 \mid N_{A(N)+1}(N) = 0\} = c(N) \quad (2)$$

Set the initial conditions: $N_1(1) = 1$; $A(1) = 1$ (3)

Then for any N, $\sum_{i=1}^{A(N)} N_i(N) = N$. Formulae (1)-(3) define a branching Markov process.

We will analyze the behavior of the expected values $E(N_i)$ and the average frequencies $f_i(N) = E(N_i(N))N^{-1}$. Consider two special cases corresponding to two different assumptions about the mutation rate.

1. $c(N) = c = \text{const}, \quad c \ll 1$ (4)

Then, the expected number of species at step N is

$$E(A(N)) = 1 + (N-1)c \quad (5)$$

Calculation of the explicit expression of $E(N_i(N))$ is complicated by the fact that the step $N^{(i)}$ when the species s_i appears is a random variable. After quite an intricate derivation we obtain:

$$E(N_i(N)) = \left[N \left(\frac{c}{1-c} \right)^{i-1} \left(\sum_{j=0}^{i-2} \left(\frac{1-c}{c} \right)^j \frac{(-1)^{i-j}}{j+1} + (-1)^i \frac{c \ln c}{1-c} \right) \right]^{1-c} \quad (6)$$

Hence, for $c \ll 1$ and $i \gg 1$: $E(N_i(N)) \approx \left(\frac{cN}{i-1} \right)^{1-c}$, $f_i(N) \approx \frac{c^{1-c} N^{-c}}{(i-1)^{1-c}}$, (7)

which is Zipf's law (with the exponent slightly smaller than 1).

2. Now assume that the probability of mutation leading to the emergence of a new species decreases with time:

$$c(N) = bN^{-q}, \text{ where } q \ll 1. \quad (8)$$

Then the expected number of species grows slower than N:

$$E(A(N)) = \frac{b}{1-q} N^{1-q} \quad (9)$$

For large ranks i the frequencies are

$$f_i(N) \approx \left[\frac{b}{(1-q)i} \right]^{\frac{1}{1-q}} \exp\left[-\frac{b}{q} \left(\frac{b}{(1-q)i} \right)^{\frac{q}{1-q}} + \frac{b}{qN^q} \right] \quad (10)$$

This is also Zipf's law, since the exponential factor is almost constant for $b \ll 1$, $q \ll 1$. For example, if b=0.1, q=0.1, the factor changes from 0.45 to 1 when i changes from 1 to ∞. In contrast with case 1 (q=0), now the exponent in Zipf's law is larger than 1, and there exists a counterpart of the thermodynamic limit ($N \to \infty$) for the average frequencies:

$$f_i = \lim_{N \to \infty} f_i(N) = \left[\frac{b}{(1-q)i} \right]^{\frac{1}{1-q}} \exp\left[-\frac{b}{q} \left(\frac{b}{(1-q)i} \right)^{\frac{q}{1-q}} \right] \quad (11)$$

Let us address now the question of the complexity of the system described by our model. We expect intuitively that a "good measure" of complexity should reflect both "unpredictability" and "organization" (which implies memory) in the behavior of a complex system. We suggest as a measure of complexity at time N the mutual information between two successive states of the system S_N and S_{N-1}.

$$C_N = I(S_N; S_{N-1}) = H(S_N) - H(S_N | S_{N-1}) \quad (12)$$

This measure agrees with our intuition since it is nonnegative and vanishes for both extreme cases of chaotic (i.e. memoryless) systems and, on the other hand, strictly deterministic systems.

In our model the state S_N is a random vector with a random number A(N) of components: $S_N = (N_1(N), N_2(N), ..., N_{A(N)}(N))$ (13)

For large N we can approximately consider random variables $N_i(N)$ as independent. Then in case 1, approximately,

$$C_N \approx \frac{\pi^2}{6} cN - (1-c) \ln N \quad (14)$$

Thus, the limit complexity per one component of the system (one species) is

$$\tilde{C}_{\lim} = \lim_{N \to \infty} \frac{C_N}{E(A(N))} \approx \frac{\pi^2}{6} \text{ nats}, \quad (15)$$

or 2.37 bits per species.

Similar analysis in case 2 gives the same limit complexity per species (specific complexity) for $q \ll 1$. Apparently, this complexity is characteristic for all systems which obey Zipf's law with the exponent close to 1.

References

1. Zipf, G.K., The Psychobiology of Language. Houghton-Mifflin, Boston, 1935.

2. Studies on Zipf's Law. Ed. H. Guiter and M.V.Arapov. Studienverlag Brockmeyer, Bochum, 1982.

3. Mandelbrot, B., An Information Theory of the Statistical Structure of Language. In: Communication Theory. Ed. W.Jackson, London,1953, 486-502.

4. Shreider, Yu. A., Theoretical Derivation of Text Statistical Features (A Possible Proof of Zipf's Law). Problems of information Transmission, v.3, No.1, 1967, 57-63.

On Noiseless Diagnosis

Raymond W. Yeung
Department of Information Engineering
The Chinese University of Hong Kong, N.T., Hong Kong

Noiseless diagnosis has wide applications in source coding, decision table programming, medical diagnosis, database query processing, quality assurance in manufacturing, and pattern recognition. In this paper, we consider the following formulation of such a problem. Let F be the fault of a system which takes value in $\Omega = \{f_i, i = 0, \cdots, n-1\}$, and p_i be the probability of occurrence of f_i. Let $T = \{t_j\}$ be the set of tests available for diagnosing the system, and c_j and s_j be the cost and the number of possible responses of t_j, respectively. The set T is *sufficient* in the sense that it can distinguish all the possible faults of the system, and the tests in T are *noiseless* in the sense that for a given fault, when a particular test is applied, the response of the test is deterministic. We are interested in a testing tree which minimizes the expected cost to identify F, whose cost is denoted by C_{min}. Special cases of our model can be found in [1]-[3].

In our formulation, it is assumed that only one fault can occur. We note that this assumption is by no means restrictive, because if multiple faults can occur, we can always regard each possible combination of faults as a single fault, and reformulate the problem as a *single-fault* problem.

Define the *efficiency* of a test t_j by

$$e_j = \frac{\log s_j}{c_j}.$$

Since the number of possible responses of t_j is s_j, the maximum amount of entropy reduced when t_j is applied is $\log s_j$. This is achieved when all the responses are equally likely immediately before t_j is applied. Thus e_j gives the maximum amount of entropy reduced per unit cost when t_j is applied.

Assume without loss of generality that the tests in T are indexed such that

$$e_1 \geq e_2 \geq \cdots.$$

Now define a mapping $\rho : R^+ \rightarrow R^+$ as follows. First define $\rho(x)$ for the values

$$x_r = \sum_{j=1}^{r} e_j c_j,$$

$r = 0, 1, 2, \cdots$ by

$$\rho(x_r) = \sum_{j=1}^{r} c_j.$$

For a value between x_r and x_{r+1}, the value of $\rho(x)$ is defined as the interpolation of $\rho(x_r)$ and $\rho(x_{r+1})$. Thus

$$\rho(x) = \sum_{j=1}^{\gamma(x)} c_j + (x - x_{\gamma(x)})/e_{\gamma(x)+1}$$

where $\gamma(x)$ is the largest integer r such that

$$x \geq \sum_{j=1}^{r} e_j c_j.$$

Theorem 1 (Lower Bound) *For any testing tree, $C \geq \rho(H)$.*

The following lemma, which is of fundamental interest, is instrumental in the proof of Theorem 1. Basically, it is a generalization of Shannon's entropy bound to non-D-ary trees.

Lemma 1 *For a testing tree, define the descendancy matrix $[a_{ik}]$, where*

$$a_{ik} = \begin{cases} 1 & \text{if } f_i \text{ is a descendant of non-terminal node } k \\ 0 & \text{otherwise} \end{cases}$$

Let r_k be the number of branches of non-terminal node k. Then

$$\sum_i p_i \sum_k a_{ik} \log r_k \geq H.$$

Toward obtaining upper bounds on C_{min}, we introduce the notion of *irreducibility* of a sufficient test set.

Definition 1 *A test set T is irreducible if T is sufficient and no proper subset of T is sufficient.*

Lemma 2 *If an irreducible test set contains a test with d possible responses, the size of the test set is at most $n - d + 1$.*

Theorem 2 *Let the tests in a sufficient test set T be indexed such that $s_1 \leq s_2 \leq \cdots$. Then the size of an irreducible subset of T is at most j^*, where j^* is the largest integer j satisfying*

$$n - s_j + 1 \geq j. \tag{1}$$

Theorem 3 (Universal Upper Bound) *Let the tests in a sufficient test set T be indexed such that $s_1 \leq s_2 \leq \cdots$. Then C_{min} is upper bounded by the total cost of the most expensive j^* tests in T.*

The universal upper bound on C_{min} does not depend on $\{p_i\}$. This bound is particularly useful when $\{p_i\}$ is unknown. We also obtain a refined upper bound on C_{min} which depends on $\{p_i\}$.

References

[1] D. A. Huffman, "A method for the construction of minimum redundancy codes," *Proc. IRE*, 40, 1090-1101, 1962.

[2] E. N. Gilbert and E. F. Moore,"Variable-length binary encodings," *Bell Syst. Tech. J.*, vol. 38, no. 4, pp. 933-968, 1959.

[3] K. R. Pattipati and M. G. Alexandridis, "Applications of heuristic search and information theory to sequential fault diagnosis," *IEEE Trans. Syst. Man Cybern.*, vol. 20, no. 4, pp. 872-887, Jul/Aug 1990.

Upper Bound for Uniquely Decodable Codes in a Binary Input N-User Adder Channel

by

Shraga Bross and Ian F. Blake
Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario

Abstract

The binary input N-user adder channel models a communication media accessed simultaneously by N users. In this model each user transmits binary sequences and the channel's output on each bit slot equals the sum of the corresponding N inputs. A *uniquely decodable* code for this channel is a set of N codes – a code for each of the N users – such that the receiver can determine all possible combinations of transmitted codewords from their sum. Van-Tilborg presented a method for determining an upper bound on the size of a uniquely decodable code for the two-user binary adder channel. He showed that for sufficiently large block length this combinatorial bound converges to the corresponding capacity region boundary.

In the present work we use a similar method to derive an upper bound on the size of a uniquely decodable code for the binary input N-user adder channel. The new combinatorial bound is iterative – i.e., the bound for the $(N-1)$-user case can be obtained by projecting the N-user bound on $(N-1)$ combinatorial variables and in particular it subsumes the two-user result. For sufficiently large block length the N-user bound converges to the capacity region boundary of the binary input N-user adder channel.

Summary

The binary input N-user adder channel is a discrete memoryless channel accessed by N users. It is assumed that each user transmits binary sequences, bit and block synchronism is maintained, and the channel's output on each bit slot equals the sum of the corresponding N inputs. A *uniquely decodable* (UD) code for this channel is a collection of N block-length n codes – $(C_1, C_2, \cdots, C_N)$ – such that the sums $c_1 + c_2 + \cdots + c_N$ for any $(c_1, c_2, \cdots, c_N) \in C_1 \times C_2 \times \cdots \times C_N$ are different. This enables the receiver to uniquely determine all possible combinations of transmitted codewords from their sum. The coding problem is to find a UD code which maximizes the product $|C_1| \cdot |C_2| \cdots |C_N|$ – i.e., a UD code having the maximum rate-sum.

Van Tilborg considered the binary ($N = 2$) adder channel [1,2]. He showed that the size of any uniquely decodable block code pair (C_1, C_2) of length n is upper bounded by

$$|C_1| \cdot |C_2| \leq \sum_{k=0}^{n} \binom{n}{k} \min \left\{ 2^k , 2^{(n-k)} \right\} \qquad (1)$$

Furthermore, for sufficiently large n the rate-sum of the combinatorial upper bound on (1) converges to the capacity region boundary of the binary adder channel.

In the present work we use a technique which resembles Van Tilborg's method to derive an upper bound on the size of any UD block code for the N-user adder channel. We prove the following result

Theorem : Let $(C_1, C_2, \cdots, C_N)$ be a uniquely decodable block code of length n for the N-user adder channel. Then

$$|C_1| \cdot |C_2| \cdots \cdot |C_N| \leq \sum_{\substack{k_1, k_2, \cdots, k_{N-1} \\ 0 \leq k_1 + k_2 + \cdots + k_{N-1} \leq n}} \binom{n}{k_1} \binom{n-k_1}{k_2} \cdots$$

$$\binom{n - \sum_{i=1}^{N-2} k_i}{k_{N-1}} \cdot \min \left\{ \max \left(2^{k_1}, 2^{k_2}, \cdots, 2^{k_{N-1}} \right) , 2^{(n - \sum_{i=1}^{N-1} k_i)} \right\} \qquad (2)$$

The upper bound (2) is iterative – i.e., the $(N-1)$-user bound can be obtained by projecting the r.h.s. of (2) on a subspace of $(N-1)$ combinatorial variables (e.g. by setting $k_{N-1} = 0$).

For $N = 3$ the bound admits the form

$$|C_1| \cdot |C_2| \cdot |C_3| \leq \sum_{\substack{k,l \\ 0 \leq k+l \leq n}} \binom{n}{k} \binom{n-k}{l} \cdot$$

$$\min \left\{ \max \left(2^k, 2^l \right) , 2^{n-(k+l)} \right\} \qquad (3)$$

which yields Van Tilborg's result (1) upon projection on the $N = 2$ plane.

The asymptotic behavior of the rate-sum corresponding to the r.h.s. of (2) is investigated. We show that for sufficiently large n the rate-sum is lower bounded by

$$\sum_{i=1}^{N} R_i \geq 1 + \frac{1}{2} \log_2 N \, . \qquad (4)$$

The lower bound (4) is identical to the capacity region boundary for $N = 2$ and is shown to be fairly close to the capacity region boundary for $N \geq 3$.

References

1. H. Van Tilborg, "An Upper Bound for Codes in the Two-Access Binary Erasure Channel," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 1, pp. 112-116, Jan. 1978.

2. H. Van Tilborg, "Upper Bounds on $|C_2|$ for a Uniquely Decodable Code Pair (C_1, C_2) for a Two-Access Binary Adder Channel," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 386-389, May 1983.

Coding for the Synchronized Multiple-Access Binary Adder Channel with Idle Sources

Y. W. Wu
Center for Signal Warfare
Vint Hill Farms Station
Warrenton, VA 22186

S. C. Chang
Department of ECE
George Mason University
Fairfax, VA 22030

Coding techniques for the synchronized multiple-access binary adder channel with idle sources are studied. Based on Lindstrom's combinatory detection algorithm, a class of uniquely decodable multiple-user codes is constructed. The rate sums of these codes are asymptotically equal to the maximum achievable values. Each constituent code has a zero vector, and two nonzero vectors which are 1's complement of each other. The source information bits "0" and "1" are encoded by using these two nonzero complementary vectors, and the idle state of source is represented by the zero vector. This approach is quite similar to the direct-sequence spread-spectrum multiple-access method. This coding mechanism will provide an exploratory methodology to fill the gaps among random access collision resolution, multiple-user information theory and spread spectrum.

ON CYCLIC CODES FOR THE T-USER Q-ARY ADDER CHANNEL

Valdemar C. da Rocha Jr.
Communications Research Group - CODEC
Department of Electronics and Systems
Federal University of Pernambuco
50741 Recife PE BRASIL

Abstract: This paper addresses the construction of q-ary cyclic codes for the synchronous noiseless T-user q-ary adder channel (T-QAC). This construction is adaptive in the sense that the decoder will correctly identify any t active users, $0 \le t \le T$, and correctly recover their respective messages, i.e., any subset of t active users (unknown in advance to the decoder) will be uniquely decoded. A very low complexity decoding procedure is given and it is shown that the maximum achievable sum rate is 1.

INTRODUCTION

In a recent paper [1] Mathys introduced a class of codes for the synchronous noiseless T active out of N multiple-access channel which is a discrete-time real adder channel without feedback with N real or binary inputs. These codes are uniquely decodable and have a sum rate that approaches 1 if the decoder is informed of which T or less users are active. That sum rate is reduced to a value of at most 1/2 if the decoder has to identify the subset of active users which in this case is limited to at most T/2.

We remark that two desirable properties of codes designed for a code division multiple-access (CDMA) communication system are the possibility of identification by the decoder of the active users without sacrificing code rate and the availability of a low complexity decoding procedure.

In what follows we prove a theorem which allows the use of q-ary cyclic codes in a synchronous noiseless T-user real adder channel in such a manner that they are uniquely decodable. The maximum sum rate achieved is 1. This relatively low maximum sum rate is compensated for by the fact that the resulting decoder satisfies the two desirable properties mentioned above.

We consider the factorization of x^n-1 over GF(q), assuming that n and q are relatively prime, which we denote as (n,q)=1. The condition (n,q)=1 implies that x^n-1 has no repeated irreducible factors over GF(q). Let $g_1(x),g_2(x),\ldots,g_T(x)$ denote a set of T polynomials which are factors of x^n-1 and are pairwise relatively prime over GF(q). We note that $\max \sum_i \deg[g_i(x)]=n$. Since $g_i(x)$ and $(x^n-1)/g_i(x)=h_i(x)$, $1 \le i \le T$, are relatively prime polynomials and g(x) has degree at least 1, it follows by the greatest common divisor theorem for polynomials that there exists $\beta_i(x)$ such that

$$\beta_i(x)h_i(x) \equiv 1 \mod g_i(x), \quad 1 \le i \le T. \tag{1}$$

Let us assume a noiseless synchronous T-QAC. Let $m_i(x)$ denote the message polynomial for user i. Let $h_i(x)$ be the generator polynomial of the cyclic code allocated to user i. The codewords of user i are generated in the usual manner by computing $m_i(x)h_i(x)$ but, before being transmitted, each codeword is multiplied by $\beta_i(x)$ and reduced modulo x^n-1. Obviously the operations of encoding and multiplying by $\beta_i(x)$ can be done simultaneously.

THEOREM: Let $C_1,C_2,\ldots,C_T$ be blocklength n q-ary cyclic codes with message polynomials $m_1(x),m_2(x),\ldots,m_T(x)$ and generator polynomials $h_1(x),h_2(x),\ldots,h_T(x)$, respectively. Then the t-tuple $(C_1,C_2,\ldots,C_t)$, $1 \le t \le T$, is uniquely decodable on the synchronous noiseless t-user q-ary adder channel and has a maximum sum rate of 1 achieved when t=T.

PROOF: We note that $\max \deg[m_i(x)] = \deg[g_i(x)]-1$, $1 \le i \le T$. By the Chinese remainder theorem for polynomials [2,pp.287-288] it follows that the polynomial sum $r(x) = m_1(x)h_1(x)\beta_1(x) + m_2(x)h_2(x)\beta_2(x)+\ldots+m_t(x)h_t(x)\beta_t(x)$ over GF(q), and $1 \le t \le T$, is uniquely determined by $m_1(x),m_2(x),\ldots,m_T(x)$ when $\deg[m_i(x)] < \deg[g_i(x)]$, $1 \le i \le t$ and $\deg[r(x)] < n$. Therefore it follows that r(x), considered as a real sum of polynomials, is also uniquely determined by $m_1(x),m_2(x),\ldots,m_T(x)$. Since each code C_i has $\deg[g_i(x)]$ information symbols and $\max \sum_i \deg[g_i(x)]=n$, it follows that the maximum sum rate rate is 1. □

The situation when $m_i(x)=0$ may be confused by the decoder with the situation when user i is not active. Such ambiguities can be avoided, for example, by forbidding the messages $m_i(x)=0$.

To decode the information sent by user i, i.e., to extract $m_i(x)$ from r(x), we simply apply the Chinese remainder theorem *in reverse order*, i.e., we compute over GF(q) the remainder of the division of r(x) by $g_i(x)$, $1 \le i \le t$.

Since $g_i(x)$ is a factor of $m_j(x)h_j(x)\beta_j(x)$, $1 \le j \le t$, $j \ne i$, it follows that $r(x) \equiv m_i(x)h_i(x)\beta_i(x) \mod g_i(x)$. However, from (1) and the assumption that $\deg[m_i(x)] < \deg[g_i(x)]$ it follows that $r(x) = m_i(x) \mod g_i(x)$.

A CLASS OF EQUAL RATE BINARY CYCLIC CODES

Let $n=2^m-1$ be a Mersenne prime. It is well known that, except for x-1, all the remaining (n-1)/m irreducible factors of x^n-1 have degree m. We can therefore construct a class of equal rate binary cyclic codes for the T-user binary adder channel (T-BAC) as follows. Let T be a divisor of (n-1)/m, i.e., Ts = (n-1)/m. Let $g_i(x)$, $1 \le i \le T$, be chosen as the product of s distinct degree m irreducible factors of of x^n-1 and such that $(g_i(x),g_j(x))=1$, $i \ne j$. The generator polynomial for user i, denoted by $h_i(x)$, is as previously defined. The binary codes constructed when T=2, i.e., codes for the 2-BAC, in general do not satisfy the sufficient condition for unique decodability given in [3].

ACKNOWLEDGEMENT

Partial support by the Brazilian National Council for Scientific and Technological Development (CNPq) under the grant No. 304214/77 is gratefully acknowledged.

REFERENCES

[1] P.Mathys,"A class of codes for T active users out of N multiple-access communication system", IEEE IT Trans. vol. 36, No.6, pp.1206-1219, November 1990.

[2] R.E.Blahut, Theory and Practice of Error Control Codes, Addison-Wesley, 1983.

[3] V.C. da Rocha Jr. and J.L.Massey,"A new approach to the design of codes for the binary adder channel", presented at the Third IMA Conference on Cryptography and Coding, Cirencester, England, December 15-18, 1991.

Coding for the Gaussian Multiple-Access Channel: An Algebraic Approach

Bixio Rimoldi*
Washington University
Department of Electrical Engineering
Electronics Systems and Signals Research Laboratory
St. Louis, MO 63130, USA

Abstract Given any group G, the G-adder channel is the channel with T inputs taking values in G and output equal to the sum (over G) of the inputs. An F-adder channel is a G-adder channel where G is the additive group of a finite field F. Similarly, the R-adder channel is the one corresponding to the usual field R of real numbers. The Gaussian multiple-access channel is the cascade of the R-adder channel with the (single-user) additive white Gaussian noise channel. Multiple-access multiple-rate codes for F-adder channels are defined and two constructions for such codes are given. In order to use such codes on the Gaussian multiple-access channel, the latter is decomposed into a number, say l, of F-adder channels. This is done via a construction involving a lattice with sufficient coding gain to reduce the error-probability to a negligible value and sublattices of it by means of which we form a suitable chain of finite quotient groups. The multiple-access codes described are well suited for use with random-access protocols with multiple reception.

SUMMARY

In this paper we present a new approach to construct multiple-access codes for F-adder channels, for the R-adder channel, and for the Gaussian multiple-access channel (see the Abstract for the definition of these channels).

The first part of our paper focuses on F-adder channels. We begin by introducing a set of definitions that we deem convenient for an algebraic approach to coding for F-adder channels. Then two constructions leading to multiple-access multi-rate codes with sum-rate constraint are given.

Let $R^i = K^i/N$ be the largest rate needed at node i, $i = 1, 2, \cdots, T$, where N is the blocklength common to all codes. A *multi-rate code* $\mathcal{D}^i$ for node i is a set of $K^i + 1$ linear block codes, one for each rate r in the set $\mathcal{R}^i := \{k/N : k \in \{0, 1, \cdots, K^i\}\}$. The need for multi-rate codes derives from the source model which is assumed to be bursty (see Gallager's comments below).

A *multiple-access multi-rate code* is a set of T (one for each channel input node) multi-rate codes. The multiple-access multi-rate codes obtained via our constructions have *sum-rate constraint* in the sense that decoding upon observation of the received sum-word is successful, provided that the sum-rate satisfies $\sum_{i=1}^{T} r^i \leq R$, where r^i is the rate of the linear block code used at node i and R is a design parameter not exceeding 1.

In the second part of our talk we focus attention on the Gaussian multiple-access channel. By means of T "modulators" and a "demodulator," we decompose the Gaussian multiple-access channel into a number, say l, of independent F-adder channels that are used as described in the first part of our presentation. The decomposition procedure can be summarized as follows. We start with an appropriate lattice S_0 as input signal set to to transform the Gaussian multiple-access channel into a virtually error-free R-adder channel (See e.g. [1]). From S_0 and any sublattice S_1 one obtains a quotient group S_0/S_1. We assume that the choice of S_0 and S_1 are such that S_0/S_1 is isomorphic to the additive group of a finite field F. This allows us to decompose the R-adder channel into an F-adder channel and an independent residual R-adder channel with inputs in S_1. The procedure can be repeated with the residual channel, provided that the F-adder channel is used to transmit codewords of a multiple-access code over F. If no power constraint is given, this procedure can be repeated indefinitely. However, If a power constraint is given, one has to stop after a finite number of steps l, namely when the only element of S_l that satisfies the power constraint is the zero element.

Our approach addresses the criticism raised by Gallager [5, page 124] when he observes that: "[There are] three bodies of research on multiaccess channels, each proceeding in virtual isolation from the others and each using totally different models." Gallager refers to the fact that the research on multiple-access channels has concentrated either on the bursty arrival of messages (collision resolution research) or on the noise and interference aspects of the multiple-access channel (information theoretic approach) but not on both. The information theoretic approach does not take into account the source model since one generally assumes sources producing information at some average rate. Unfortunately, in order to see this average rate one has to smooth out the source by averaging over a long time. This introduces unacceptable delays. Our approach addresses both aspects. In particular, the problem of bursty arrival of messages is addressed by having multi-rate codes. If it the source model is such that one cannot guarantee that the sum-rate constraint is fulfilled at all times, like when the arrival statistic is Poisson, then one can use our multiple-access multi-rate codes with a random-access protocol with multiple-reception as described in [2] - [4].

References

[1] J, H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Springer-Verlag, New York, 1988.

[2] M. Kavehard, "An accessing technique for information packet networks," in *Proc. IEEE Electron. Aerospace syst. Conv.*, Nov. 1981, pp. 42-45.

[3] B. S. Tsybakov, V. A. Mikhailov, and N. B. Likhanov, "Bounds for packet transmission rate in a random multiple access system," *Prob. Paredachi Informatsii*, vol. 19, no. 1, pp. 61-81, 1983.

[4] N. Mahravari, "Random-access communication with multiple reception," *IEEE Trans. Inform. Theory*, vol. 36, pp. 614-622, May 1990.

[5] R. Gallager, "A perspective on multiaccess channels," invited paper in *IEEE Trans. Inform. Theory*, vol. 31, pp. 124-142, March 1985.

*Supported in part by National Science Foundation grant NCR-9109944.

OPTIMAL MULTIUSER CODES FOR THE REAL ADDER CHANNEL

A. Brinton Cooper, III
U.S. Army Research Laboratory
Advanced Computing and Information Sciences
Aberdeen Proving Ground, Maryland 21005

Brian Hughes
Department of Electrical and
Computer Engineering
The Johns Hopkins University
Baltimore, Maryland 21218

Abstract

A new family of uniquely decodable binary codes is presented for the T-user real adder channel. The codes consist of T individual codebooks, each containing only two codewords, one of which is the all-zero sequence. These codes achieve a sum rate that is equal, asymptotically in T, to the sum capacity. An iterative decoding algorithm is also presented. Applications are discussed to codes for T active users out of M potential users, and to superimposed codes.

Summary

The *T-user real adder channel* is a multiple access channel with output $y = x_1 + x_2 + \ldots + x_T$, which is the real sum of a set of binary input symbols, one from each of T users. A *T-user code* for this channel is a set $(C_1, C_2, \ldots, C_T)$ of binary block codes having a common length N. The *sum rate* is $R_{sum}(T) = R_1 + R_2 + \ldots + R_T$, where R_i is the rate of the code C_i. A T-user code is *uniquely decodable* (UD) if all of the sums, formed by taking one codeword from each user, are distinct.

Chang and Weldon [1] showed that the *sum capacity* of this channel satisfies

$$C_{sum}(T) \approx (1/2)\log_2 T \text{ as } T \to +\infty\,. \qquad (1)$$

They also presented a family of UD T-user codes for which $R_{sum}(T) \approx C_{sum}(T)$ as $T \to +\infty$.

Our investigation of T-user codes for the real adder channel is motivated by an interest in codes for T active users out of M [2], and superimposed codes [3], for which the present codes can serve as building blocks. In this application, the codes of [1] are inadequate because the vast majority of users have only non-zero codewords, so these users must be active at all times.

We therefore consider T-user codes $(C_1, \ldots, C_T)$ of the following form: each code C_i consists of only two codewords, *one of which is the all-zero sequence.* A T-user code of this form can be described by a $T \times N$ binary *generator matrix* B, the i^{th} row of which is the nonzero codeword of C_i.

Supported in part by ARO Grant DAAL03-89-K-0130.

We specify a family of T-user codes by recursively constructing the corresponding generator matrix. The first code is the trivial single-user code with $B_0 = [1]$ and $T_0 = N_0 = 1$. The rule for constructing B_j from B_{j-1} is

$$B_j = \begin{bmatrix} B_{j-1} & B_{j-1} & 0_{T_{j-1}} \\ B_{j-1} & \bar{B}_{j-1} & 1_{T_{j-1}} \\ I_{j-1} & 0_{j-1} & 0_{N_{j-1}} \\ 0'_{N_{j-1}} & 1'_{N_{j-1}} & 1 \end{bmatrix} \quad j = 1, 2, 3, \ldots \qquad (2)$$

where $\bar{B}_{j-1}$ is the one's complement of B_{j-1}, I_{j-1} is the identity matrix of dimension N_{j-1}, 0_{j-1} is a square, all-zero matrix with dimension N_{j-1}, 0_N is the all-zero $N-$tuple, and 1_N is the all-one $N-$tuple.

Theorem: For any positive integer j, the matrix B_j in (2) defines a uniquely decodable binary T_j-user code of blocklength N_j, where

$$T_j = (j+1)2^j \quad \text{and} \quad N_j = 2^{j+1} - 1\,. \qquad (3)$$

◇

Note that as $T_j \to +\infty$

$$R_{sum}(T_j)/C_{sum}(T_j) \to 1. \qquad (4)$$

We also present an iterative decoding algorithm and a brief description of new T-out-of-M user codes and superimposed codes which can be constructed from the T-user codes.

References

[1] S. C. Chang and E.J Weldon, Jr, "Coding for T-user multiple-access channels," *IEEE Transactions on Information Theory*, IT-25 (6), pp. 684-691, November 1979.

[2] P. Mathys, "A class of codes for a T active users out of M multiple access communication system," *IEEE Transactions on Information Theory*, IT-36 (6), pp. 1206-1219, November 1990.

[3] T. Ericson and L. Györfi, "Superimposed codes in R^n," *IEEE Transactions on Information Theory*, IT-34 (4), pp. 877-880, July 1988.

Two-Decodable Coding of the Two-User Binary Adder Channel

Jian-Jun SHI and *Yoichiro WATANABE*

Department of Electronics, Doshisha University,
Kyoto, 602 JAPAN.

Abstract: A two-decodable coding scheme for the two-user binary adder channel is proposed, where the first code of the code pair is constrained to a class of linear codes.

1. Introduction

This paper deliberates on how to construct a two-decodable code pair (C,S) for the two-user binary adder channel. We use the fact that when the first code C is given *a priori*, a maximum independent set of the δ-order graph $G_C^{(\delta)}$ associated with C achieves the highest rate of the second code S, which is proposed by Kasami and Lin in 1983[1].

For a restricted model of C, it is possible to evaluate a lower bound of the independence number of $G_C^{(2)}$ and to propose a practical construction scheme of the two-decodable code pair. The two-order graph $G_C^{(2)}$ associated with C is decomposed into layers, each of which consists of mutually isomorphic subgraphs. It is easy to calculate their independence numbers[3]. The sum of the independence numbers of non-adjacent subgraphs is the lower bound which we will propose.

2. Lower bound

The code C to be considered here is an (n,k) linear code with a generator matrix $\Gamma=[I_k P_1 \cdots P_k]$, where I_k is a $k\times k$ identity matrix, and P_j $(j=1,\cdots,k)$ is a $k\times x_j$ matrix with all entries of 1 in the j-th row and entries of 0 elsewhere, and $x_1+x_2+\cdots+x_k=n-k$, $x_j\geq 0$[2].

For the 2-order graph $G_C^{(2)}$, its vertex set $V=\{0,1\}^n$ can be divided into the partition $V = V_0 \cup V_1 \cup \cdots \cup V_{2^{n-k}-1}$, where V_i, $i=0,1,\cdots,2^{n-k}-1$, is a coset of C. Let $L_i=0\ldots0a_{i1}^{(1)}\ldots a_{ix_1}^{(1)}\cdots a_{i1}^{(k)}\ldots a_{ix_k}^{(k)}$ be a coset leader of V_i. Define $Q_i=\{l \mid a_{i1}^{(l)}\ldots a_{ix_l}^{(l)}\neq 0,\ l=1,\cdots,k\}$. Let F_i be a subgraph of $G_C^{(2)}$ induced by V_i. The subgraph F_i belongs to the $|Q_i|$'th layer, and its independence number is equal to $2^{|Q_i|}$.

Put a set $m_i=\{m_{i1},\cdots,m_{ik}\}$, where the m_{il} is the number of "1" in the block $a_{i1}^{(l)}\ldots a_{ix_l}^{(l)}$. The number of F_j's such that $m_i=m_j$, $j=0,1,\cdots,2^{n-k}-1$, is caculated as $\prod_{l=1}^{k}\binom{x_l}{m_{il}}$. In order to select mutually non-adjacent F_i's, let M be a subset of $\{m_i,\ i=0,1,\cdots,2^{n-k}-1\}$ such that $m_i,m_j\in M$ satisfy one of the following two cases:

(a) If $m_{il}+m_{jl}\neq x_l$, $l=1,\cdots,k$, then $\sum_{l=1}^{k}(m_{il}-m_{jl})\geq 2$.

(b) If there exists an index p such that $m_{ip}+m_{jp}=x_p$, then $\sum_{l=1,l\neq p}^{k}(m_{il}-m_{jl})\geq 1$.

The number of non-adjacent F_i's depends on M. The above arguments are summarized in the following theorem.

Theorem : Let C be an (n,k) linear code with the generator matrix Γ, then a lower bound of the independence number of $G_C^{(2)}$ is given by

$$\alpha(G_C^{(2)})\geq\max_{M}\sum_{i=1}^{|M|}2^{|Q_i|}\prod_{l=1}^{k}\phi(m_{il})\binom{x_l}{m_{il}},$$

where $\phi(m_{il})=\begin{cases}0.5 & m_{il}/x_i=0.5\\ 1 & \text{otherwise}\end{cases}$. □

3. Conclusion

The construction scheme of the two-decodable code pair is proposed as follows: Let C be an (n,k) linear code with the generator matrix Γ, and choose out a set M that makes an independent set larger. This independent set is the code S. Thus the two-decodable code pair (C,S) is obtained. It is confirmed that there exist generator matrices Γ's such that the lower bounds are equal to $\alpha(G_C^{(2)})$.

The authors are grateful to Dr. H. Harada for his instruction in this paper.

References

[1] T. Kasami, S. Lin, V. K. Wei and S. Yamamura, "Graph theoretic approaches to the code construction for the two-user multiple-access binary adder channel," *IEEE Trans. Inform. Theory*, **IT-29**, pp.114-130 (1983).

[2] F. Guo and Y. Watanabe, "Graphs Associated with a Linear code," *IEICE Trans.* **E 74**, pp.49-53 (1991).

[3] J-J. Shi and Y. Watanabe, "Two-Decodable Code Pair for the Two-User Binary Adder Channel," Presented at *1991 ISCOM Int. Symp. Com.* Tainan, Taiwan.

LINEAR CODES FOR AN AWGN MULTIPLE ACCESS CHANNEL WITH PARTIAL ACCESS.

Gregory Poltyrev and Jakov Snyders

Dept. of Electrical Engineering - Systems, Faculty of Engineering
Tel-Aviv University
Tel-Aviv, 69978, ISRAEL

Abstract

A method of transmitting information through an AWGN multiple access binary adder channel (BAC) will be addressed. We shall consider the following procedure of access to the channel: there are N users but only m, $m < N$, users are active (are transmitting their messages) during a fixed period of communication; the transmission is completely synchronized; the subset of the active users is known to the receiver. Such situation will be named transmission through BAC with a partial access (BACPA). If $N >> m$ then time sharing is a very ineffective method for the BACPA. Indeed, since the subset of active users is unknown to each of the users, the time must be shared between all N users. Consequently, the overall transmission rate R_{ov} is given by $R_{ov} = \frac{m}{N}R \ll 1$, where R is the coding rate of each user. We shall show that effective transmission through a BACPA can be realized by means of linear codes. More specifically, we shall show that for any noiseless BACPA it is possible to construct N linear codes such that $R_{ov} = 1$ and the decoding error probability equals 0. For the case of AWGN BAC, we shall show that transmission by means of linear codes can have even better characteristics than time sharing.

Summary

The following situation of transmitting information through a multiple access channel is addressed. A binary real adder channel (BAC) perturbed by additive white Gaussian noise (AWGN) is considered. The BAC can be described as follows: the inputs corresponding to the users are binary, i.e., the input alphabet of each user is $X = \{-1,1\}$; the output of the channel at instance j is equal to $y^j = \sum_{i=1}^{m} x_i^j + z^j$, where m is the number of active users, x_i^j is the input signal of the ith user at instance j and z^j; $j = 1,2,\ldots$, are iid Gaussian random variables. We shall consider the following procedure of access to the channel: there are N users but only m, $m < N$, users are active (are transmitting their messages) during a fixed period of communication; the transmission is completely synchronized; the subset of the active users is known to the receiver. Such situation will be named transmission through BAC with a partial access (BACPA).

We shall call the coding for BAC for the case $N = m$, i.e., when all existing users are simultaneously active, coding for transmission through BAC with a complete access (BACCA). The simplest, and usually used, method of transmission through BACCA is time sharing. In that case, under the condition that all users use codes with the same coding rate R, the information transmission rate of each user is equal to $\frac{R}{m}$ and the overall rate R_{ov} of transmission through the BACCA is given by $R_{ov} = R < C_0 \leq 1$, where C_0 is the capacity of the AWGN one-way channel with binary input. It is known that there are coding methods for the BAC for which $R_{ov} > C_0$ [1]. A remarkable fact is that the capacity of the AWGN BAC is achieved by uniform distribution on the inputs of the users ($p(x=-1)=p(x=1)=0.5$). Consequently, the capacity of the BAC can be attained on the ensemble of binary linear codes. The possibility of employment of linear codes simplifies considerably the coding, and frequently also the decoding. It should be noted here that linear codes can not realize the zero error of decoding probability under transmission through the BAC without noise with overall rate $R_{ov} > 1$ [2]. This implies that the value of the decoding error probability for linear codes can be worse than for nonlinear ones. But the simplicity of realizing the coding and decoding can in many cases be an acceptable price for such deterioration.

If $N >> m$ then time sharing is a very ineffective method for the BACPA. Indeed, since the subset of active users is unknown to each of the users, the time must be shared between all N users. Therefore, the overall transmission rate R_{ov} is given by $R_{ov} = \frac{m}{N}R \ll 1$, where R is the coding rate of each user. We shall show that effective transmission through a BACPA can be realized by means of linear codes. More specifically, we shall show that for any noiseless BACPA it is possible to construct N linear codes such that $R_{ov} = 1$ and the decoding error probability is equal to 0.

Bounds on the number of users, N, for a given m will be presented. In the case of AWGN BAC, any m active users share the same binary linear (n,k) code C. This means that the code of any user is either a subcode of C or some coset of the subcode. We construct, by means of the random coset method [3], an upper bound on the error decoding probability. This bound enables us to show that, for the case of AWGN BAC, transmission by means of linear codes outperforms, with respect to the decoding error probability, time sharing schemes. Parameters that play a significant role in determination of the probability of decoding error will be considered.

References

[1] R. Alswede, "Multi-way communication channels", *Proc. 2nd Int. Symp. Inform. Theory*, Tsahkadsor, Armenian S.S.R., (1971), *pp.* 23-52, Publishing House of the Hungarian Academy of Science, 1973.

[2] T. Kasami, S. Lin, V.K. Wei, S. Yamamura, "Graph theoretic approaches to the code construction for the two-user multiple-access binary adder channel", *IEEE Trans. Information Theory*, vol.IT-29, no.1, *pp.* 114-130, Jun. 1983.

[3] G. Poltyrev, "About improving the upper bound on error decoding probability for codes with complicated structure", *Problemy Peredachi Informatii*, vol.23, No.4, pp. 5-18, 1987.

CODING FOR THE F-ADDER CHANNEL: TWO APPLICATIONS OF REED SOLOMON CODES

Rüdiger Urbanke Bixio Rimoldi*
Washington University
Department of Electrical Engineering
Electronics Systems and Signals Research Laboratory
St. Louis, MO 63130, USA

Abstract

Given any finite field F, the F-adder channel is the channel whose inputs are elements of F and the output is the sum (over F) of the inputs. It is shown how Reed Solomon (RS) codes can be used to obtain multiple-access multiple-rate codes of convolutional type for the F-adder channel. It is also shown that when the F-adder channel is noisy, the codewords of a multiple-access multi-rate code for the F-adder channel can be protected in a simple and flexible manner by means of RS codes.

I Introduction

In a companion paper [1] multiple-access multiple-rate codes for the F-adder channel have been defined and two constructions of such codes have been given. All codes in [1] are of block type (as opposed to convolutional type). This paper presents two applications of Reed Solomon (RS) codes to coding for the F-adder channel. The first application results in multiple-access multi-rate codes of convolutional type. This is done in section II. In section III we assume that the F-adder channel is noisy and show how to combine multiple-access coding and error protection in a flexible way.

II Convolutional Codes for the F-Adder Channel

In [1] we explicitly assumed that the channel is the F-adder channel where F is a finite field. One can easily verify that definitions and constructions still apply if we replace F by $R = F[D]$, the ring of polynomials over F[1]. In this way, generator matrices over R defining block type multiple-access codes over R can be seen as generator matrices of convolutional type multiple-access codes[2] over F. Since adding elements of R is the same as transforming these elements into sequences over F, adding them componentwise, and transforming back the resulting sequence to a ring element, one can use the convolutional codes obtained in this way as multiple-access codes for the F-adder channel (as opposed to the $F[D]$-adder channel).

Example 1 Consider the following 7×3 generator matrix.

$$\boldsymbol{h}^T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & D & D^2 \\ 1 & D^2 & D+D^2 \\ 1 & 1+D & 1+D^2 \\ 1 & D+D^2 & D \\ 1 & 1+D+D^2 & 1+D \\ 1 & 1+D^2 & 1+D+D^2 \end{pmatrix}. \quad (1)$$

The elements of $\boldsymbol{h}^T$ are over the polynomial ring $R = F[D]$ with $F = GF(2)$. Assume that 7 users are sharing a binary-adder channel. Assign to user i the rate 1/3 convolutional encoder having as generator matrix the i-th row[3] of $\boldsymbol{h}^T$. The receiver will be able to decode the messages upon observation of the channel output, provided that no more than 3 users are active and provided that the receiver knows which users are active. This is true since any 3 rows of $\boldsymbol{h}^T$ are linearly independent. Notice that decoding is unique, provided that the sum rate is not larger than unity. Multiple-access codes having this property were denoted *optimal* in [1].

The key to obtaining optimal multiple-access codes of convolutional type for the F-adder channel lies in the ability to find an $n \times (n-k)$ matrix over $R = F[D]$ with the property that all collections of $(n-k)$ rows are linearly independent. The values of n and $(n-k)$ are design parameters that depend on the number of users and on how many of them are allowed to be active concurrently.

Let $F = GF(p^m)$ be any finite field. Let $E = GF\left(p^{lm}\right)$ be any finite extension field of F. Let n divide $p^{lm}-1$. Let $\boldsymbol{h}^T$ be the transposed parity check matrix of a (n,k) RS code over E. Then it is readily checked that $\boldsymbol{h}^T$ generates an optimum convolutional type multiple-access code, if we view its elements as polynomials over F. We note that the specific binary example presented above has been derived by this procedure with $F = GF(2)$, $l = 3$, $n = 2^3 - 1$ and $k = 4$. A similar construction method for such matrices has been proposed in [2] (there the motivation was to find channel correcting convolutional codes).

III Error Control Coding via RS Codes

Let $\boldsymbol{h}^T$ be the transposed parity check matrix of a (n,k) RS code over $GF(p^m)$. Assume that we want to construct a multiple-access code to operate over a noisy F-adder channel and hence need an error correcting scheme to secure the multiple-access codewords. Let $(n-k)$, the length of the multiple-access codewords, divide $p^m - 1$, and assume that we want to be able to correct up to t errors per block.

Partition the n rows of $\boldsymbol{h}^T$ into generator matrices and let $\boldsymbol{B}^i$ be the generator matrix of user i. Let $\tilde{\boldsymbol{c}}^i$ be the encoded message of user i. Before transmission, user i sets the last $2t$ components of $\tilde{\boldsymbol{c}}^i$ to zero and takes the (inverse) Fourier transform. Clearly the resulting *outer* code will be a RS code with error correcting capability t (the last $2t$ frequency components of $\boldsymbol{c}$ are zero). This scheme works because the transposed parity check matrix $\tilde{\boldsymbol{h}}^T$ of a $(n, k+2t)$ RS code can be derived from the transposed parity check matrix $\boldsymbol{h}^T$ of a (n,k) RS code by deleting the last $2t$ columns. Hence by setting the last $2t$ components of $\tilde{\boldsymbol{c}}^i$ to zero and taking the Fourier transform we actually base the multiple-access codes on a $(n, k+2t)$ RS code and embed this in an outer $(n-k, n-k-2t)$ RS code for the error correction. Needless to say, t may take on any value $0 \le t < \frac{n-k}{2}$. This error correction scheme does not decrease the number of available rows of $\boldsymbol{h}^T$, but reduces the maximum possible sum rate by a factor $\frac{n-k-2t}{n-k}$. We see that this error correction schemes allows a flexible choice between low error probability and high sum rate.

References

[1] B. Rimoldi, "Coding for the Gaussian Multiple Access Channel: An Algebraic Approach," in *International Symposium on Information Theory*, (Austin, TX), IEEE, Jan. 1993.

[2] P. Piret, *Convolutional Codes - An Algebraic Approach*. Cambridge, Massachusetts: The MIT Press, 1988.

[1] More generally R could be a commutative ring with a unit element with respect to multiplication and no zero divisors.

[2] We assume that the information sequences have finite length, so that they can be represented by a polynomial.

[3] More generally, the number of rows assigned to users may vary in order to account for users having dissimilar rate requirements.

*Supported in part by National Science Foundation grant NCR-9109944.

Joint Signal Detection (D) and Estimation (E) Under Prior Uncertainty: New Results

by
David Middleton
127 E. 91 Street, New York, NY 10128, U.S.A.†

Abstract:

When estimation of signal waveform or signal parameters takes place under prior uncertainty as to whether or not the signal is actually present [$p(H_1) < 1$, with $\therefore q(H_0) > 0$] estimation based on the assumption that $p(H_1) = 1$ can result in estimates that can be seriously in error. Moreover, such estimators and estimates are themselves biased, with unknown bias if the "noise only" ($q(H_0) > 1$) state is not properly accounted for.

The present paper extends the original analyses of Middleton and Esposito [1a,1b], and more recent work of the present author [2], to include canonical estimation in generalized noise for least mean square error (LMSE) estimators and for (unconditional) maximum likelihood estimators (UMLE's), which last were not available before in the case of the UMLE. [In addition, the verbal presentation includes new *threshold* results, obtained for *correlated noise samples* (the author's so-called quasi-equivalent (QE) noise models), where only the first order pdf $w_1(z)$ and covariance k_Z of the noise process are available [2].]

Summary:

In many practical signal processing situations where estimation in noisy environments of signal waveform (S) or parameters (θ) is required (e.g., classification and localization of targets, measurements, remote sensing, etc.), it is often not known *a priori* whether or not the desired signal is present. Detection and estimation are then jointly required, so that both a correct estimator, i.e., one that is unbiased and optimal, can be constructed, and an associated (optimal) detector employed, which in turn can "validate" (i.e., accept or reject) the estimate. The *a priori* probabilities are denoted by $q(H_0)$, $p(H_1)$, with $q+p = 1$, $0 \le (q,p) \le 1$, and the desired (optimum) estimators, γ^*, are denoted by $\gamma^*_{p<1}$ here, with $\overline{\gamma^*_{p<1}}^{H} = p\bar{\theta}$, where $H = H_0 + H_1$, for these estimators to be unbiased. As usual, optimality is defined in terms of minimum average "risk" or cost. [New canonical results for jointly optimum threshold D and E in generalized noise, for both coherent and incoherent reception, *when the noise samples are correlated*, are discussed in the presentation.]

Furthermore, we consider weak coupling only, between detector and estimator, with the added simplification that *in detection* the cost of declaring the signal absent when the signal is present, and vice versa, is independent of the signal. Then detection and estimation can be carried out independently, in parallel, with the convention that the (optimum) estimator $\gamma^*_{p<1}$ is rejected if the probability of correct detection [$P_D = p(H_1)(1-\beta)$] is smaller than some preselected value. Failure to account for the fact that $p(H_1) < 1$ can have serious consequences in applications of the estimator:

1. Optimum LMSE Estimators, $p(H_1) < 1$

Here we shall summarize the principal results recently obtained (see also [1], [2], and refs.) for the quadratic cost function (QCF), which leads to LMSE estimators. We have, generally, the optimum estimator

$$\boxed{\gamma^*_{p<1}\Big|_{QCF} = \left(\frac{\Lambda_J}{1+\Lambda_J}\right)\gamma^*_{p=1}\Big|_{QCF}}\ , \text{cf. (3.7, [1].} \tag{1.1}$$

where Λ_J is the generalized likelihood ratio:

$$\Lambda_J \equiv \mu\left(F_J(x|S(\theta))\right)_\theta / F_J(x|0); \mu \equiv \frac{p(H_1)}{q(H_1)}; \begin{cases} x = V/\sqrt{\psi} = \{x_j\}; J = MN \\ j = (m,n) \end{cases}$$

with space (m,M)–time (n,N), $m = 1,\ldots,M$; $n = 1,\ldots,N$; $j = m,n$, sampling; x = normalized data vector of $MN = J$ elements, and $F_J(x|S)$ = j-fold pdf of x, the generalized noise given the signal (vector) $S = [S_j]$, etc. The associated Bayes risk ($= C_o \cdot$ LSME) is expressed formally as

$$R(\sigma,\gamma^*)|_{p<1;QCF} = C_o\overline{|\gamma^* - \theta|^2} = C_o\left\{\overline{q\left(\gamma^*_{p<1}\Big|_{QCF}\right)^2}^{H_o} + \overline{p\left|\theta - \gamma^*_{p<1}\Big|_{QCF}\right|^2}^{H_1}\right\}, \tag{1.2}$$

Specifically, the general result $\gamma^*_{p=1}\Big|_{QCF}$ for estimating the $\theta_{\hat{m}}, \hat{m} = 1,\ldots,\hat{M}$, out of $(\theta_{\hat{m}},\theta') = \theta$ parameters is given by:

$$\boxed{\gamma^*_{p=1}\Big|_{QCP} = \int_{\theta_{\hat{m}}} \hat{\theta}\sigma(\hat{\theta}_m)e^{\ell^{(21)}(x|\hat{\theta})_{QCF}}d\hat{\theta},} \tag{1.3}$$

where $\ell^{(21)} = \log\left(F_J(x|S(\theta_{\hat{m}},\theta'))\right)_{\theta'} - \log\left(F_J(x|S)\right)_\theta$, with $\theta = (\hat{\theta},\theta')$, $\hat{\theta} = \left[\theta_{\hat{m}}\right]$. cf. (3.8), [1].

2. Optimum UML Estimators ($p(H_1) < 1$)

Other new results recently obtained by the author concern the UMLE, which are derived from an appropriate "simple" cost function. We use a "strict" form, which now yields a Bayes risk‡

$$R^*_E(\sigma,\delta^*)|_{SCF} = C_o\left\{A_M - \int_\Gamma\left[qF_J(x|0)\delta\left(\gamma^*_{p<1}(x) - 0\right)dx + p\int_\Omega F_J(x|S(\theta))\delta\left(\gamma^*_{p<1}(x) - \theta\right)\sigma(\theta)d\theta\right]\right\} \tag{2.1}$$

Maximizing the integrand of $\int_\Gamma(\ )dx$, minimizes R_E ($= R^*_E$); e.g., the extremal condition determining $\gamma^*_{p<1}(x)$ is $\frac{\partial}{\partial\gamma}\left\{\delta(\gamma - 0) + p\,F(x|S(\gamma)\sigma(\gamma))\right\}_{\gamma\to\gamma^*_{p<1}} = 0$.

Using the fact that $\int_{\Gamma_o} F_J(x|0)dV = \int_\Gamma F_J(x|0)\delta\left(\gamma^*_{p=1}(x) - 0\right)dx = q_{\hat{M}}(0|H_0)_{\gamma^*_{p=1}}$

where $q_{\hat{M}}\left(y|H_0\right)_{\gamma_{p=1}}$ is the $\hat{M}$-fold pdf of $\gamma^*_{p=1} = y = \left[y_1,\ldots,y_{\hat{M}}\right]$ and Γ_O is the domain of x for which $\gamma^*(x)_{p=1} = 0$, while Γ = domain of all x, *and the requirement that $\gamma^*_{p<1}$ must be unbiased*, gives, after some manipulation, the desired (new) result for the optimum estimator here:

$$\boxed{\gamma^*_{p<1}(x)\Big|_{SCF} = \begin{cases} \gamma^*_{p=1}(x)_{SCF},\ x \in \Gamma_1 : \gamma^*_{p=1}(x) \neq 0 \\ -\overline{\gamma^*_{p=1}(x)_{SCF}}^{H_0} \Big/ \left[1 + \Lambda_J(x)\right]q_{\hat{M}}\left(0|H_0\right)_{\gamma^*_{p<1}},\ x \in \Gamma_0 : \gamma^*_{p=1}(x) = 0 \end{cases}} \tag{2.2}$$

Here $\gamma^*_{p=1}\Big|_{SCF}$ is found in the usual way [1a],

$$\frac{\partial}{\partial\hat{\theta}}\left\{\log\sigma(\hat{\theta}) + \ell^{(10)}(x|\hat{\theta})_{SCF}\right\}_{\hat{\theta}=\gamma^*} = 0, \tag{2.3}$$

where $\ell^{(10)} \equiv \log\left(F_J(x|S(\hat{\theta},\theta'))\right)_{\theta'} - \log F_J(x|0)$.

The associated Bayes risk is obtained here by inserting (2.2) back into (2.1) and employing integration procedures like the above. For $p = 1$, only the first term of (2.2) applies.

3. Concluding Remarks

As noted at the beginning, failure to account for the fact that $p(H_1) < 1$ can not only lead to erroneous (and biased) estimates, but also these can be sufficiently inaccurate as to have serious consequences. For example, a difference in, say, the mean estimate of threshold signal amplitude (power) of ~10% vis-à-vis the correct ($p < 1$) value, corresponding to $p \equiv 0.9$ (vs. $p = 1$) produces an error of 10% in the average minimum detectable signal. This can be 2.5 to 3.0 dB for –25 dB or –30 dB for the latter: serious amounts, for instance, in "Matched Field Processing", where one tries to keep signal degradation below 1 dB for effective matching of the propagation model to the received data.

References

[1a]. D. Middleton and R. Esposito, "Simultaneous Optimum Detection and Estimation of Signals in Noise," IEEE Trans. Information Theory, Vol., IT-14, No. 3, May 1968, pp. 434–444.

[1b]. _____, "New Results in the Theory of Simultaneous Optimum Detection and Estimation of Signals in Noise," *Problemy Peredachi Informatsii*, Vol. 6, No. 2, April–June, 1970, pp. 3–20; Engl. trails., pp. 93–106, Consultants Bureau, NY., (Plenum), 1973.

[2]. D. Middleton, "Threshold Detection and Estimation in Correlated Interference," paper 2A2, pp. 7–12, *Proceedings*, 9th Intl. Zürich Symposium on EMC: "EMC '91." Switzerland, March 12–14, 1991.

†Based on work supported under Grant N00014-91-J-4131, Code 1114 SE (Dr. R. N. Madan), Office of Naval Research.

‡ Equations (3.15), (3.16), [1a] contain unnecessary integrals over V, since F_n, $w(\gamma)$, $\delta(\gamma - V) \ge 0$.

OPTIMUM INCOHERENT DETECTION OF FADING SIGNALS IN NON-GAUSSIAN NOISE

E. Conte, M. Di Bisceglie, M. Lops
Dipartimento di Ingegneria Elettronica, Università di Napoli, Via Claudio 21, 80125 Napoli, Italia

Abstract

The problem of detecting one out-of M fading signals in a Spherically symmetric noise process is addressed. We show that the classical detector is canonically optimum regardless the fading and the noise models. An example is worked out for the case of Middleton Class-A distributed noise and Nakagami fading.

Noise and fading models

A compound-Gaussian process [1] can be thought of as the product of a modulating non-negative, wide-sense stationary process, $s(t)$ say, and a Gaussian, possibly complex, one, $g(t)$ say, independent of $s(t)$, namely $c(t) = s(t)\,g(t)$. Obviously, not all processes are amenable to such a representation; precisely, the admissibility condition the common distribution of the quadrature components of the noise should fulfill is

$$f_{cI}(x) = f_{cQ}(x) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 s^2}} e^{-\frac{x^2}{2\sigma^2 s^2}} f(s)\,ds \qquad (1)$$

where σ^2 is the common variance of the quadrature components of the Gaussian process and $f(s)$ is the first-order pdf of the random process $s(t)$. Among the marginal pdf's complying with (1) we cite the Middleton Class-A distribution, the Generalized Gaussian, the Generalized Cauchy, the Generalized Laplace [2]. In keeping with theoretical considerations, supported by experimental evidence, we assume that the bandwidth of $s(t)$ is much smaller than that of $g(t)$; so, on sufficiently short time intervals, the modulating process is practically a random constant and the overall noise process degenerates into a Spherically symmetric random one. When such a model is in force, the spectral properties of the process reproduce, except for a scale factor, those of the Gaussian noise.

As to the channel, we assume the flat-flat fading model: the useful received signals are hence related to the transmitted waveforms through the complex factor $\alpha = Ae^{j\phi}$ with A –the random gain of the channel– arbitrarily distributed and ϕ –the received phase– uniformly distributed in $[0, 2\pi)$.

Synthesis of the optimum detector

The M-ary detection problem can be stated as

$$H_i \;:\; \mathbf{r} = Ae^{j\phi}\mathbf{p}_i + \mathbf{c} \quad i = 1, 2, \ldots, M \qquad (2)$$

where $\mathbf{r}$, $\mathbf{p}_i$, $\mathbf{c}$ are complex, N-dimensional vectors representing the corresponding waveform signals as N diverges. For equally likely hypotheses and signals with equal energy $\mathcal{E}_p$, minimizing the error probability requires choosing the hypothesis that maximizes the function

$$\overline{\frac{1}{(2\pi s^2\mathcal{N}_0)^N} e^{-\frac{\|\mathbf{r}\|^2 + A^2\mathcal{E}_p}{2s^2\mathcal{N}_0}} I_0\left(\frac{A|\mathbf{r}\cdot\mathbf{p}_i|}{s^2\mathcal{N}_0}\right)} \qquad (3)$$

where the bar denotes expectation with respect to A and s, $I_0(\cdot)$ is the modified Bessel function of first kind and order zero and $2\mathcal{N}_0$ is the noise power spectral density. It can be shown that this is equivalent to maximizing $|\mathbf{r}\cdot\mathbf{p}_i|$. Thus, the optimum receiver is the classical minimum-distance detector, regardless the first-order distributions of the noise and of the fading. As to the correlated case, it can be easily managed by introducing a linear filter which whitens the received observations [2].

Receiver performance in Nakagami fading

The performance of the above receiver in non-Gaussian noise can be evaluated by simply averaging s out of $P(e|s)$, the error probability under Gaussian noise.

Consider the Nakagami m-distribution, namely

$$f(A) = \frac{2m^m A^{2m-1}}{\Gamma(m) A_{rms}^{2m}} e^{-m(A/A_{rms})^2} \qquad (4)$$

where A_{rms} is the channel root mean square gain, and m is a shape parameter ruling the fading depth. For the case of M orthogonal signals embedded in noise with Middleton Class-A pdf, we obtain

$$P(e) = \sum_{i=1}^{\infty} \sum_{k=1}^{M-1} \binom{M-1}{k} \frac{(-1)^{k+1}\epsilon_i}{k+1} \left[\frac{s_i^2 m(k+1)}{\gamma_R k + s_i^2 m(k+1)}\right]^m$$

where $\epsilon_i = e^{-\nu}\nu^i/i!$ ν a shape parameter, $s_i^2 = (i/\nu + \lambda)/(1+\lambda)$ λ the ratio of the power of the Gaussian componenent to that of the impulsive one and γ_R denotes the Signal-to-Noise Ratio (SNR). Results indicate that when deep fading is present, (e.g. m=1) the shape parameter of the noise is almost uninfluential, while, for increasing m, spikier noise results into worse performance: however, the detection loss, as measured with respect to the Gaussian case, approaches a constant value (depending on m) as SNR diverges.

References

E. Conte, G. Galati, M. Longo, "Exogenous modelling of non-gaussian clutter", *J. Inst. Electron & Radio Eng.*, 1987, No.57, pp.191-197.

E.Conte, M. Di Bisceglie, M. Longo, M. Lops, "Canonical Detection in Spherically Invariant Noise", *IEEE Trans. on Communications*, (under revision).

Detection of Time-frequency Concentrated Transient Signals

Thomas P. Krauss, *Student Member, IEEE*, Thomas W. Parks, *Fellow, IEEE*, and Ram G. Shenoy, *Member, IEEE*

SUMMARY

We consider the problem of detection of time-frequency concentrated transient signals in white Guassian noise. There are many instances in which the detection of time-frequency concentrated transients can be useful: any case in which the class of transients to be detected is known to have a certain time-frequency signature, but the exact time samples are not known. Examples of such classes are speech and animal sounds, sonar and radar return pulses, seismic signals, and underwater acoustic transients.

Recent study of the Weyl correspondence as it relates to time-frequency representations [1,2] has been fruitful in that a way to associate a function in the time-frequency plane (a.k.a. "Phase Space" or the "Wigner plane") with a linear operator has been suggested and studied. This linear operator has several useful properties: as shown in [1], it is a self-adjoint operator provided the function in the Wigner plane, or the "symbol" as it is called, to which it corresponds is real. Also a real symbol can be reconstructed by taking a weighted sum of the Wigner distributions of the eigenfunctions of its corresponding linear operator (with the eigenvalues as the weights).

One of the properties of the Weyl correspondence most important in the application of detection of time-frequency concentrated signals is the fact that the (double) integral of the symbol multiplied by the Wigner distribution of a signal, i.e. their inner product, is equal to the inner product of the image of the signal under the symbol's corresponding operator and the signal itself. That is, if $P(t,f)$ is the symbol (a function of time and frequency), and $W_{x,x}(t,f)$ is the (auto) Wigner distribution of the signal x, then the following equality holds:

$$(P, W_{x,x}) = (L_P x, x)$$

where L_P is the linear operator corresponding to the symbol P. The utility of this property is this: we define a signal as concentrated in a region if the integral of its Wigner distribution over the region is large [3]. If the symbol is a "mask", or the indicator function of some region R in the time-frequency plane, then the Rayleigh quotient $(L_P x,x)/(x,x)$ can be considered the degree of concentration of x in the region R. Hence the problem of detecting transient signals that have a large degree of concentration in a particular time-frequency region becomes one of detecting signals for which the Rayleigh quotient $(L_P x,x)/(x,x)$ is larger than some positive threshold. Such a class of signals is called a cone-class and our detection problem is to detect signals in such a set.

Our solution to this problem is an application of the generalized likelihood ratio test (GLRT) assuming the following two hypotheses:

$$H_0: r = w$$

$$H_1: r = w + s$$

where w is a zero mean Gaussian noise process of known variance (with covariance matrix $\sigma^2 I$), and s is an unknown signal in the cone-class $C_{L_P(\mu)}$ defined by

$$C_{L_P(\mu)} = \{x: \frac{(L_P x, x)}{(x, x)} \geq \mu\}$$

The generalized (log) likelihood ratio is

$$\Lambda(r) = (r,r) - (r-\hat{s}, r-\hat{s})$$

where $\hat{s}$ is the closest approximation to r in $C_{L_P(\mu)}$. Detector performance is analyzed for a class of random concentrated signals, and for a class of acoustic well-logging signals. As the concentration level μ approaches the largest eigenvalue λ_1 of the operator L_P the problem approaches subspace detection [4, pp.145-147] in the eigenspace of λ_1.

The theory is applied here to the discrete-time, discrete-frequency case, where signals are parameterized completely by a finite-number of samples. The Wigner distribution and the linear operator corresponding to an arbitrary symbol can be computed without approximation in this case; properties of each are verified.

REFERENCES

[1] R. G. Shenoy and T. W. Parks. The weyl correspondence and time-frequency analysis. Submitted to IEEE Trans. Acoust. Speech Signal Process.

[2] F. Hlawatsch and W. Kozek. Time-frequency representation of linear time-varying systems using the weyl symbol. In *Proc. IEE Sixth Int. Conf. on Digital Signal Processing in Communications*, 1991.

[3] R. G. Shenoy and T. W. Parks. Time-frequency concentrated basis functions. In *Proc. ICASSP 90*, pages 2459–2462, April 1990.

[4] Louis L. Scharf. *Statistical Signal Processing; Detection, Estimation, and Time Series Analysis.* Addison-Wesley, 1991.

This work was supported by ONR contract N00014-91-1161.

T. P. Krauss was with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853. He is now with the MathWorks, Inc., Natick, MA 01760

T. W. Parks is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

R. G. Shenoy is with Schlumberger-Doll Research, Ridgefield, CT 06877.

QUICKEST DETECTION OF AN ABRUPT CHANGE IN A RANDOM SEQUENCE WITH FINITE CHANGE-TIME

Yong Liu and Steven D. Blostein
Department of Electrical Engineering, Queen's University
Kingston, ON, Canada K7L 3N6

Quickest change detection has a wide variety of applications, including search radar, digital signal processing, image processing, monitoring communication channels, and fault detection [1, 2]. In this paper a modified Shiryayev criterion [9] is used to study the problem of quickest detection of an abrupt change in a random sequence with independent and identical distributions before and after the change. The modified Shiryayev criterion minimizes expected delay

$$D_m \triangleq E\{N - m + 1 | N \geq m\} \tag{1}$$

subject to a given false alarm probability

$$\alpha \triangleq P\{N < m\} \leq \gamma \tag{2}$$

and a given false alarm average run length (ARL)

$$N_{fm} \triangleq E(N | N < m) \geq t \tag{3}$$

where N is random stopping variable, m is non-random unknown change-time, γ and t are two constants. It is noted that if Eq.(3) is ignored and m is random, then the above criterion is identical to Shiryayev's ([9], Eqs.(4.130)(4.131), p.198). The difference between the modified Shiryayev criterion and the criteria used in [6, 7] is that the former considers the false alarm probability while the latter do not. The modified Shiryayev criterion has important applications in a number of situations where a change occurs in a finite time. It has been shown [9] that a procedure of Girshick-Rubin [3] and Shiryayev [8, 9] (GRS procedure) is optimal when the unknown change-time is geometrically distributed, where the Shiryayev criterion is used. In this paper the unknown change-time is assumed to be any non-random integer instead of a geometrically distributed random variable assumed by Shiryayev [9].

A new procedure is proposed in this paper to approach the problem defined above, which minimizes an asymptotic risk (a linear combination of D_m, α and N_{fm}) and is referred to as a minimum asymptotic risk (MAR) procedure. The decision statistic (MAR statistic) is

$$\hat{r}(x_1^n) = \frac{T_n \pi_0 c_0 - (1 - \pi_0)}{T_n \pi_0 + 1 - \pi_0} n \tag{4}$$

where $x_1^n \triangleq (x_1, \ldots x_n)$ is observations, π_0 and c_0 are design parameters, and T_n is the well-known CUSUM statistic. The CUSUM statistic can be written as $T_n = \max\{T_{n-1}, 1\}\frac{f_1(x_n)}{f_0(x_n)}$, where f_0 and f_1 are densities before and after the change respectively. The decision rule of the MAR procedure is that one stops and decides a change has occurred as soon as

$$\hat{r}(x_1^n) > \hat{r}(x_1^{n-1}). \tag{5}$$

It has been shown [4] that the MAR procedure is asymptotically optimal. The optimality is formally expressed as the following theorem:

Theorem 1 *When (i) the one-sample likelihood ratio $\frac{f_1(x_1)}{f_0(x_1)} \neq 1$ for all x, or (ii) the false alarm probability α goes to zero, we have*

$$D_{bm} \leq D_m \tag{6}$$

subject to

$$N_{bfm} \leq N_{fm} \tag{7}$$

and

$$\alpha_b = \alpha, \tag{8}$$

where D_{bm} and D_m denote the expected delay for the MAR procedure and any other procedure, respectively. N_{bfm} and N_{fm} denote the corresponding false alarm ARLs, and α_b and α the false alarm probabilities.

For nonasymptotic situations, simulation results reported in [4, 5] reveal that the MAR procedure compares favorably with the CUSUM, GRS and moving window fixed sample size procedures, where the modified Shiryayev criterion was used.

We have also observed that the MAR procedure is very insensitive to the choice of design parameter c_0. It can be shown [4] that $c_0 = \pi_0$ satisfies Theorem 1.

References

[1] Basseville, M. (1988) 'Detecting changes in signals and systems: a survey,' *Automatica*, Vol.24, No.3, pp.309-326.

[2] Basseville, M. and A. Benveniste(ed.) (1986), *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-verlag, New York.

[3] Girshick, M.A. and H. Rubin (1952), 'A Bayes approach to a quality control model,' *Annals of Mathematical Statistics*, vol.23, 114-125.

[4] Liu, Y. and S. D. Blostein (1992a), 'Quickest detection of an abrupt change in a random sequence with finite change-time,' *IEEE Transactions on Information Theory*, (submitted)

[5] Liu, Y. and S. D. Blostein (1992b), 'A modified Bayesian procedure for quickest signal change detection,' *Proceedings of the Biennial Symposium on Communications*, Kingston, Ontario, Canada, May 1992, pp.406-409.

[6] Lorden, G. (1971), 'Procedures for reacting to a change in distribution,' *Annals of Mathematical Statistics*, 42(6): 1897-1908.

[7] Roberts, S.W. (1966), 'A comparison of some control chart procedure,' *Technometrics*, Vol.8, No.3, August.

[8] Shiryayev, A.N. (1963), 'On optimum methods in quickest detection problems,' *Theory of Probability and its Applications*, Vol. 8, No.1.

[9] Shiryayev, A.N. (1978), *Optimal Stopping Rules*, Springer-Verlag, New York.

DECENTRALIZED ENCODING FOR LINEAR ESTIMATION OF A REMOTE SOURCE

M. Di Bisceglie, M. Longo
Dipartimento di Ingegneria Elettronica, Università di Napoli, Via Claudio 21, 80125 Napoli, Italia

Abstract

N separated sensors transmit noisy observations of a source to a central processor for final estimation; local encoding of each observation is performed prior to transmission. The issue is to devise encoding-decoding schemes for this decentralized context. The present scheme is composed of N scalar quantizers, each with a corresponding decoder, and a linear combination of individual estimates. This structure ensures that the decoder size increases only linearly with N. Through an example involving 2 sensors with a Gaussian model of source and observations, this scheme is compared to a previously considered unconstrained decoding scheme and to a distortion–rate bound.

Problem statement

With reference to Figure 1, X is a random variable representing the source and $Y_1, \ldots Y_N$ are noisy, possibly correlated, observations taken at each of N separated sensors. Due to communications constraints, Y_i is compressed by a scalar quantizer Q_i of rate R_i producing index Z_i, $i = 1, \ldots, N$. Quantizers cannot share their data,

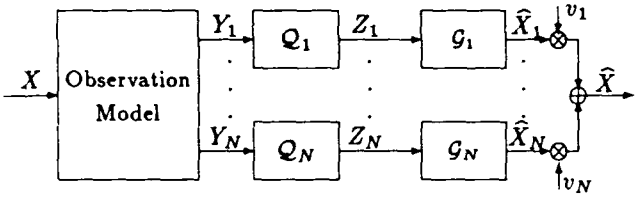


Figure 1: Schematic of the encoding-decoding system.

hence they do not cooperate. The indices are transmitted to a central processor. Reproduction of the remote source is accomplished by

1. A bank of N decoders g_i, each producing an estimate $\widehat{X}_i$ of the remote source based on Z_i.
2. Linear combination of individual estimates to yield the final estimate $\widehat{X} = \underline{v}^T \underline{\widehat{X}}$ where $\underline{\widehat{X}} = (\widehat{X_1}, \ldots \widehat{X_N})^T$.

This structure ensures that the size of the decoding table is in the order of $2^{R_1} + \cdots 2^{R_N}$, hence increases linearly with N; in contrast, if no such structure were imposed (as in [1]), the size would be $2^{R_1} \times \cdots 2^{R_N}$, hence would increase exponentially with N. Therefore, the problem is:

given: the source density $p(x)$, the observation model $p(y|x)$ and a pointwise distortion measure $d(x, \hat{x})$,

find: N pairs (q_i, g_i) of encoding-decoding rules, and a weighting vector $\underline{v}$,

such that: The average distortion $D = Ed(X, \widehat{X})$ is as small as possible.

Decoupled solution

A simple suboptimum approach is to isolate the design of the (q_i, g_i) pairs, $i = 1, \ldots, N$, from the selection of $\underline{v}$. Design of the pair (q_i, g_i) can be accomplished as in unifilar encoding–decoding of a remote source, namely alternatingly improving q_i into q_i^+ and g_i into g_i^+ according to

$$q_i^+(y_i) = \underset{z_i}{\operatorname{argmin}} \int_X d(x, \hat{x} = g(z_i)) p(x, y_i) dx \quad (1)$$

$$g_i^+(z_i) = \underset{\hat{x}_i}{\operatorname{argmin}} \int_X d(x, \hat{x}_i) p(x | z_i = q_i(y_i)) dx \quad (2)$$

Upon reception of the collection Z_i, $i = 1, \ldots, N$, a bank of N decoders, one for each encoder, produces centrally the vector of the individual optimal estimates $\widehat{X}_i$, $i = 1, 2, \ldots, N$, according to (2). The optimum weighting vector can be found as that vector which minimizes the distortion $D = E_{X\underline{Z}} d(X, \underline{v}^T \underline{\widehat{X}}(\underline{Z}))$.

Example

To assess the proposed scheme, also in comparison with unconstrained scheme [1], we adopt the quadratic distortion measure $d(x, \hat{x}) = (x - \hat{x})^2$ and we consider the case of a Gaussian source, $X \sim \mathcal{N}(0, 1)$, and 2 sensors affected by additive, zero-mean Gaussian noise with common variance σ^2 and correlation coefficient ρ. The optimum weighting vector is then $\underline{v} = E[X \underline{\widehat{X}}^T] \Sigma^{-1}$, where Σ is the covariance matrix of $\underline{\widehat{X}}$.

Results are presented in Table I which shows the performance of the present scheme (labelled as I), the performance of the unconstrained scheme [1](labelled as II) and a distortion bound for decentralized schemes [1]. Notice that the loss of the present scheme becomes significant only for highly adverse correlation and for high variance ratio $\gamma = 1/\sigma^2$.

Table I
Distortion results ($R_1 = R_2 = 3$).

	γ = 0 dB			γ = 20 dB		
ρ	I	II	OPTA	I	II	OPTA
-0.99	0.039	0.039	0.005	0.029	0.010	0.002
0	0.349	0.348	0.051	0.036	0.015	0.006
0.99	0.506	0.502	0.091	0.046	0.018	0.010

References

[1] M. Di Bisceglie, M. Longo, "Quantization for decentralized estimation from correlated data," pres. at *IEEE ISIT 90*, San Diego CA., Jan. 1990 pp.34-35.

Model Based Motion Field Estimation

Christoph Stiller and Frank Müller

Institute for Communication Engineering, Aachen University of Technology (RWTH)
5100 Aachen, Germany, Phone: +49-241-807677, Fax: +49-241-807669

Introduction

The estimation of motion and boundaries of objects in an image sequence is an important issue for efficient video compression. It allows exploitation of the strong statistical bindings of image intensities along the motion trajectories [1, 4].

The idea of object-oriented motion estimation is to subdivide the scene into regions of continuous motion. Thus discontinuities in the motion field may only occur at region boundaries. Ideally, each region uniquely corresponds to one surface of a moving object in the 3-D real world.

The motion field is regarded as a pair of random fields $V = (U, L)$, where U denotes a field of one motion vector per pixel, and L denotes a generic segmentation field. The segmentation field groups the motion vectors into several continuous regions.

Image Model

This contribution follows a model based Bayesian approach. The model considers MSE of motion compensated prediction, motion discontinuities and uncovered regions. The resulting estimation criterion is derived straight forward as MAP-criterion with help of the model assumptions.

Given the samples a (= previous frame) and b (= next frame) of the random fields A and B, the objective is to find the motion field sample $v^* = (u^*, l^*)$ of V of maximum a posteriori probability

$$P(V=v^*|A=a,B=b) = \max_v \{P(V=v|A=a,B=b)\}$$
$$= \frac{P(B=b|V=v^*,A=a)\cdot P(V=v^*|A=a)}{P(B=b|A=a)}, \qquad (1)$$

where the reformulation of the objective function follows from Bayes rule.

The first factor of the numerator is described by a so called *observation model*. In this contribution it is assumed to depend on the displaced frame difference (dfd) only. The dfd is modeled segmentwise stationary obeying a white, zero-mean *generalized gaussian distribution* in each segment.

In each segment not corresponding to decovered background ML-estimates for σ and ν are substituted back into the objective function. In decovered regions a likelihood that is slightly lower than the one of the least likely region is imposed. This assures that motion is estimated, whereever a reasonable correspondence between regions of successive frames can be established.

The second factor of the numerator in (1) mainly accounts for the spatio-temporal statistical bindings within a motion field and is described by a *motion model* representing prior expectations on V. The principle of minimum description length [3] is adopted, which assigns each sample v a probability according to its content of decision

$$P(V=v|A=a) = P(V=v) = 2^{-C(v)},$$

where $C(v)$ denotes the code length of a (lossless) *contour/texture code*. The contour code describes the segmentation matrix l while the texture code segmentwise describes the motion vectors u assuming strong statistical bindings of neighboring motion vectors belonging to the same object (label). It can be shown that V is a Gibbs/Markov random field (cf. [5]).

Combining the two models, according to (1), the MAP-criterion can be derived as

$$-\log(P(V=v|A=a,B=b)) =$$
$$\sum_{k=1}^{K} \frac{N_k}{\hat{\nu}_k}\left[1+\log\left(\frac{\hat{\nu}_k \sum_{i\in G_k}|x_i|^{\hat{\nu}_k}}{N_k}\right)\right] + N_k\cdot\log\left(\frac{\Gamma(1/\hat{\nu}_k)}{\hat{\nu}_k}\right)$$
$$+\log(2)\cdot C(v) + \text{const.}, \qquad (2)$$

where N_k denotes the number of pixels in the k-th region G_k and x_i denotes a pixel of the dfd. (2) is locally minimized employing iterated conditional modes [2] providing fast convergence.

References

[1] J.K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images - a review. *Proceedings of the IEEE*, 76(8):917 - 935, August 1988.

[2] J. Besag. On the statistical analysis of dirty pictures. *Journal Royal Statistical Society, Ser.B*, 48(3):259 - 302, 1986.

[3] Y.G. Leclerc. Constructing simle stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73 - 102, 1989.

[4] H.G. Musmann, P. Pirsch, and H.-J. Grallert. Advances in picture coding. *IEEE Proceedings*, 73(4):523 - 548, April 1985.

[5] C. Stiller and D. Lappe. Gain/cost controlled displacement-estimation for image sequence coding. In *Proc. ICASSP91, Toronto, Canada*, volume 4, pages 2729 - 2732, May 1991.

MULTI-GRID METHODS FOR MEAN FIELD THEORY IN EM PROCEDURES FOR MARKOV RANDOM FIELDS

JUN ZHANG AND BINGLAI CHEN
DEPARTMENT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
UNIVERSITY OF WISCONSIN-MILWAUKEE
MILWAUKEE, WI 53201

Abstract — In this short paper, we describe how multi-grid methods, originally developed for numerical solution of differential equations, can be used in the mean field calculations for Markov random fields in EM procedures to reduce computation and improve convergence.

1. Introduction:

Many problems in image processing and computer vision can be formulated as *incomplete data* problems [1], where part of the data is not observable (hidden) and needs to be estimated along with the data model (in particular, model parameters). The Markov random field (MRF) has been demonstrated as a very general and effective model for the hidden variables since it captures the underlying physical processes and constraints (e.g., [2]).

The EM (expectation-maximization) algorithm [1] is an effective maximum-likelihood (ML) procedure for parameter and hidden variable estimation in incomplete data problems. However, when the hidden variables are modeled as MRF's, it runs into difficulty due to the exponential complexity in the calculation of the conditional mean of the MRF required by the E-step. To overcome this difficulty, we have developed an iterative procedure [3] based on the mean field theory (MFT) of statistical mechanics. This approach provides a mathematically sound (in some sense optimal) approximation that can be calculated (without the exponential complexity).

While the efficacy of the MFT approach has been demonstrated in our previous work in image segmentation and image restoration, it is prone to a problem common to many iterative procedures — as the data size increases, the amount of computation increases drastically and the convergence slows down. In fact, this problem has plagued numerical methods for differential equations for quite some time until recently when a solution, known as the multi-grid (MG) method, has been developed. In this paper, we summarize our work in applying MG methods to the mean field calculations in EM procedures, while more details can be found in [4].

2. MFT in EM Procedures:

Let $\mathbf{S}$ be a 2-D lattice with a neighborhood system and $\mathbf{u} = \{u_i, i \in \mathbf{S}\}$ be an MRF. Then it is well known that $\mathbf{u}$ has a Gibbs distribution:

$$p(\mathbf{u}) = Z^{-1} \exp\left[-\beta U(\mathbf{u})\right] = Z^{-1} \exp\left[-\beta \sum_c V_c(\mathbf{u})\right] \qquad (1a)$$

where $U(\mathbf{u})$ is the energy function, $V_c(\mathbf{u})$'s are the clique potentials, and Z is the partition function.

It is not difficult to see that the direct calculation of the mean of $\mathbf{u}$, $< \mathbf{u} >$ is exponentially complex since one needs to sum over all possible configurations of $\mathbf{u}$. The MFT suggests an approximation [3]: the influence of u_j, $j \neq i$, in the calculation of $< u_i >$ can be approximated by that of $< u_j >$ i.e.,

$$< u_i > \simeq Z_i^{mf'^{-1}} \sum_{u_i} u_i \exp[-\beta U_i^{mf'}(u_i)], \qquad (2a)$$

where

$$U_i^{mf'} = V_c(u_i) + \sum_{j \in \mathbf{N}_i} V_c(u_i, < u_j >). \qquad (2b)$$

and $Z_i^{mf'}$ is a normalization factor. In an EM procedure where the hidden variables are modeled as MRF, the mean field theory of (2a)-(2b) can be used in the E-step to evaluate the Q function:

$$Q(\Phi|\Phi^{(p)}) = < \log p(\mathbf{y}|\mathbf{u}, \Phi) + \log p(\mathbf{u}|\Phi)|\mathbf{y}, \Phi^{(p)} >, \qquad (3)$$

where $\mathbf{y}$ is the observed data, Φ is the model parameter vector, p represents the pth iteration.

3. Application of MG Methods:

For the sake of simplicity, we illustrate the MG ideas through a two-grid method. Let $\mathbf{S}$ be the *fine grid*, now denoted by $\mathbf{S}^h$; one can then generate a *coarse grid* $\mathbf{S}^H$ by merging neighboring sites in $\mathbf{S}^h$ (e.g., merging every four neighboring sites into one site). The two-grid method achieves computation reduction and convergence acceleration by alternating mean field calculations between the fine and coarse grids rather than just on the fine grid (e.g., fine first, then coarse, then fine). Two important problems in this method are: how to transfer between $\mathbf{u}^h$ and $\mathbf{u}^H$ and how to define the coarse-grid energy function $U^H(\mathbf{U}^H)$. The solution to the first problem is easy – through an interpolator and a restriction operator, i.e., $\mathbf{I}_H^h : \mathbf{u}^H \rightarrow \mathbf{u}^h$ and $\mathbf{I}_h^H : \mathbf{u}^h \rightarrow \mathbf{u}^H$, respectively. The second problem is, on the other hand, more difficult. We have experimented with two strategies: the "fractal" method (energy function the same in different grids, similar to [5]) and Galerkin's method ($U^H(\mathbf{u}^H) = U^h(\mathbf{I}_H^h \mathbf{u}^H)$). Typical results by Galerkin's method are shown in Fig. 1 for image segmentation. To achieve the same segmentation quality (MSE of classification), the MG scheme uses only 2/3 of the time used by a fine-grid only scheme (the saving is much large for a larger image, e.g., 256 × 256).

4. Summary:

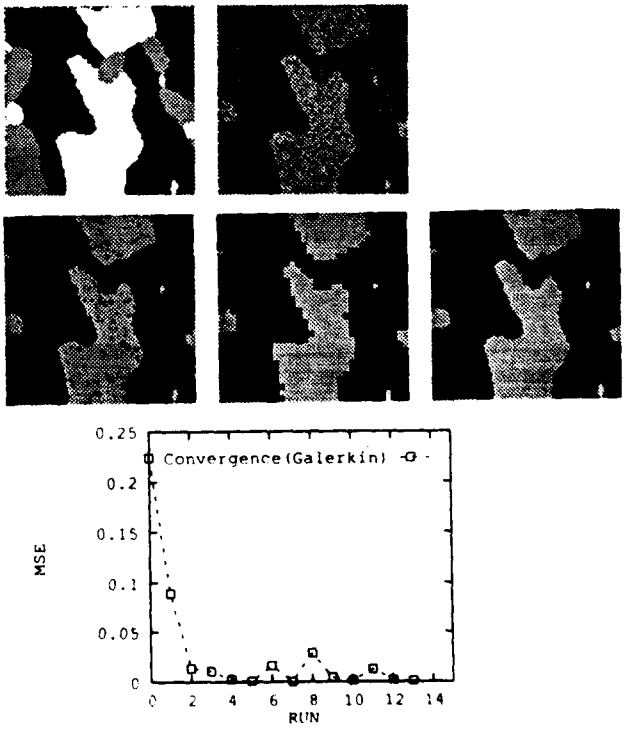
In this paper, we have described the application of MG methods in mean field calculations in EM procedures where the hidden variables are modeled as MRF's. This approach achieves computation reduction and acceleration of convergence through alternating the calculation of the mean field among different grids (fine-coarse-fine). Both fractal and Galerkin coarsening provide good results.

References

[1] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Roy. Soc. Statist., Series B., No. 1, pp. 1-38, 1977.
[2] J. Marroquin, S. Mitter, and T. Poggio, "Computer vision," J. Amer. Stat. Association, Vol. 82, pp. 76-89, March, 1987.
[3] J. Zhang, "The mean field theory in EM procedures for Markov random fields," IEEE Trans. SP., Vol. 40, pp. 2570-2583, Oct. 1992.
[4] J. Zhang and B. Chen, in preparation.
[5] C. Bounman and B. Liu, "Multiple resolution segmentation of textured images," IEEE Trans. PAMI, Vol. 13, pp. 99-113, Feb., 1991.

Figure 1: An Example in Image Segmentation.

The first row shows the true "region map" (what the segmentation should be like) and the observed image. The second row shows the results of three iterations on the fine-grid, a coarse-grid, and the fine-grid, respectively. The convergence is illustrated by the MSE (between consecutive segmentations in the iterations). Galerkin's method is used for coarsening.



Maximum Likelihood Parameter Estimation of the Harmonic, Evanescent, and Purely Indeterministic Components of Homogeneous Random Fields *

Joseph M. Francos, Anand Narasimhan† and John W. Woods
Electrical Computer and Systems Engineering Department
Rensselaer Polytechnic Institute, Troy, NY 12180-3590

The paper presents a solution to the general problem of fitting a parametric model to observations from a single realization of a 2-D homogeneous random field with mixed spectral distribution. So far, the problem has been largely unsolved. Existing methods either assume the field has an absolutely continuous spectral distribution and try to fit white noise driven linear models to the observed field, or treat the special case of estimating the parameters of a sinusoidal signal in white noise. The existence of evanescent random fields has not received attention in the estimation literature, although the evanescent components have major impact on the structure and properties of the random field, as they result in directional attributes in the observed realizations. We present a maximum-likelihood solution to this estimation problem.

On the basis of a 2-D Wold-like decomposition [1], the random field is decomposed into a sum of two mutually orthogonal components: a *purely-indeterministic* field and a *deterministic* one. The 2-D deterministic random field is further orthogonally decomposed into a *half-plane deterministic* field and a *generalized evanescent* field. The generalized evanescent field is a linear combination of a countable number of mutually orthogonal *evanescent* fields. A typical example of an evanescent field is a 2-D separable random field which is the product of a 1-D purely-indeterministic random process along one axis and a harmonic 1-D process in the orthogonal dimension. The above decomposition implies a similar decomposition of the spectral measure of the regular random field into a sum of mutually singular spectral measures, each associated with a different component of the spatial decomposition. The spectral distribution function of the purely indeterministic field is the absolutely continuous component of the regular field spectral distribution. The spectral measure of the deterministic field is concentrated on a set of Lebesgue measure zero in the 2-D frequency plane. For practical applications, these results suggest that the deterministic field "spectral density function" has the form of a sum of 1-D and 2-D Dirac delta functions. The *harmonic* random field, which generates the 2-D delta functions of the "spectral density" is one of the components of the half-plane deterministic random field. The 1-D delta functions which are supported on lines of rational slope result from the generalized evanescent random field.

Hence, in this paper we concentrate on a solution to the problem of estimating the parameters of the harmonic and generalized evanescent components of the field in the presence of an unknown colored noise generated by the purely-indeterministic component, jointly with estimating the purely-indeterministic component parameters. We assume that the purely-indeterministic component can be modeled by a 2-D AR model.

The suggested algorithm is a two-stage procedure. In the first stage we obtain a suboptimal initial estimate for the parameters of the spectral support of the evanescent and harmonic components by solving the set of 2-D overdetermined normal equations for the parameters of a high-order linear predictor of the observed data. In the second stage we refine these initial estimates by iterative maximization of the conditional likelihood of the observed data, which is expressed as a function of only the parameters of the spectral supports of the evanescent and harmonic components. This representation is possible due to a parameter transformation developed in this work. The solution for the unknown spectral supports of the harmonic and evanescent components reduces the solution for the other unknown parameters of the field, to a linear least squares solution. Experimental evidence is presented to demonstrate the high accuracy of the estimates for each of the random field components: harmonic, evanescent of any orientation, and purely indeterministic.

*This work was partially supported by NSF grant MIP-9120377.

†Currently with the I.B.M. T. J. Watson Research Center, NY 10598.

References

1. J. M. Francos, A. Z. Meiri and B. Porat, "A Wold-Like Decomposition of 2-D Discrete Homogeneous Random Fields", submitted for publication.

QUANTIZED RECEIVER R_o BOUNDS AND THE ASYMPTOTIC RELATIVE EFFICIENCY OF QUANTIZED DETECTORS

Marcos O. Cimadevilla
E-Systems
Greenville Division
Greenville, TX 75403-6056

Jerry D. Gibson
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843-3128

SUMMARY

Wozencraft and Kennedy [1] and Massey [2] have argued that the R_o criterion is the logical choice for the design of modulation systems for coded digital communications. This paper develops a relationship between the change in the R_o parameter for quantized receivers and the asymptotic relative efficiency (ARE) of quantized detectors, which sheds light on both receiver design and performance analysis.

Considering only binary input communication systems with Q output symbols and assuming that the transmission of data is corrupted by zero mean, additive white Gaussian noise, the quantized receiver has

$$R_o' = 1 - \log_2\left[1 + \sum_{h=1}^{Q} \sqrt{q_{1h}q_{2h}}\right] \tag{1}$$

where $q_{1h} = \frac{1}{\sqrt{2\pi}} \int_{x_{h-1}}^{x_h} e^{-(x-\sqrt{E_N})^2/2} dx$ and $q_{2h} = \frac{1}{\sqrt{2\pi}} \int_{x_{h-1}}^{x_h} e^{-(x+\sqrt{E_N})^2/2} dx$. Here, the x_h specify the end points of the quantization interval Δ_h, and E_N is defined as the signal energy per dimension. The expression for the cutoff rate of a DMC when employing an optimum receiver with no quantization ($Q = \infty$) is given by [3]

$$R_o = 1 - \log_2(1 + e^{-E_N/\mathcal{N}_o}), \tag{2}$$

where $\mathcal{N}_o/2 = \sigma^2$ is the noise power spectral density of the additive white Gaussian noise (AWGN) channel.

In order to determine the amount of degradation due to receiver quantization, we must first determine the $\{x_h\}$ that maximize (1). One finds that the necessary condition for the $\{x_h\}$ to be optimal is that they satisfy the condition [2] $x_h = (\ln[\lambda(b_h)\lambda(b_{h+1})]^{\frac{1}{2}})/2\sqrt{E_N}$, where $x_o = -\infty$ and $x_Q = \infty$ and $\lambda(b_h) \triangleq P(b_h|\sqrt{E_N})/P(b_h|-\sqrt{E_N}) = q_{1h}/q_{2h}$. Here the b_h corresponds to the output level associated with the Δ_h interval. An iterative numerical technique can be used to solve for the set of $\{x_h\}$.

Having obtained the threshold levels $\{x_h\}$ that maximize R_o', we may now determine the amount of degradation introduced by certain quantization schemes by comparing the difference between R_o and R_o'.

Next we consider the optimum quantization of data where the quantized data are to be used to form a test of hypothesis for signal detection. In particular, the optimum quantizer is defined as the one that maximizes detection efficacy. The problem that we consider here is the detection of a constant positive signal s in additive noise with a symmetric density function $f(x)$. For testing the hypothesis $H : \bar{r}$ consists of noise only, versus the alternative $K : \bar{r}$ consists of signal and noise, where $\bar{r}$ is the received vector, the generalized test statistic, based on quantized data, may be described by [4] $S = \sum_{i=1}^{N} r_i' s_i$ where the $\{r_i'\}_{i=1}^{N}$ are the quantized data and the $\{s_i\}_{i=1}^{N}$ is the known signal sequence representing the *dc* signal. The efficacy $\mathcal{E}_s$ for this test statistic for the constant-signal case (i.e. $s_i = s =$ constant for all i) may be written as [4]

$$\mathcal{E}_S = \frac{2\{\sum_{j=1}^{Q/2} l_j[f(a_j) - f(a_{j-1})]\}^2}{\sum_{j=1}^{Q/2} l_j^2[F(a_{j-1}) - F(a_j)]}. \tag{3}$$

Here the a_j specify the end points of the Q input ranges and the l_j correspond to the output levels of each input range. The vectors $\bar{a}$ and $\bar{l}$ that maximize (3) when $f(x)$ is the normalized Gaussian density function are given by $l_j = \int_{a_j}^{a_{j-1}} x f(x) dx / \int_{a_j}^{a_{j-1}} f(x) dx$ and $a_j = (l_{j+1} + l_j)/2$ [4].

In order to evaluate the performance of the quantized detectors, we compare them to the linear detector since this is the optimum detector when detecting a *dc* signal in Gaussian noise [5]. The test statistic of the linear detector is given by $S_L = \sum_{i=1}^{N} r_i$ and its corresponding efficacy is $\mathcal{E}_{S_L} = 1/\sigma^2$ where σ^2 is the variance of the noise density.

We obtain the ARE's of the quantized detectors relative to the linear detector by taking the ratio of their efficacies,

$$E_{S,S_L} = \frac{\mathcal{E}_S}{\mathcal{E}_{S_L}} = \frac{2\sigma^2\{\sum_{j=1}^{Q/2} l_j[f(a_j) - f(a_{j-1})]\}^2}{\sum_{j=1}^{Q/2} l_j^2[F(a_{j-1}) - F(a_j)]}. \tag{4}$$

Taking the background noise density to be the normalized Gaussian density function, we obtain the desired results.

It is shown that in the vanishing signal/large sample size case, the two efficiency measures give the same results for all values of Q.

REFERENCES

[1] J.M. Wozencraft and R.S. Kennedy, "Modulation and demodulation for probabilistic coding", *IEEE Transactions on Information Theory*, vol. IT-12, pp. 291-297, July 1966.

[2] J.L. Massey, "Coding and modulation in digital communications", in International Zurich Seminar on Digital Communications, (Zurich, Switzerland), March 1974.

[3] J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering*. New York: John Wiley & Sons, 1965.

[4] S.A. Kassam, "Optimum quantization for signal detection", *IEEE Trans. Commun.*, vol. COM-25, pp. 479-484, May 1977.

[5] J.D. Gibson and J.L. Melsa, *Introduction to Nonparametric Detection with Applications*. New York: Academic Press, 1975.

A NEW PROCEDURE FOR DECODING CYCLIC AND BCH CODES UP TO ACTUAL MINIMUM DISTANCE

G. L. Feng and K. K. Tzeng

The Center for Advanced Computer Studies, University of Southwestern Louisiana, Lafayette, LA 70504 and the Department of Electrical Engineering and Computer Science, Lehigh University, Bethelehem, PA. 18015

Abstract

In this paper, a new procedure for decoding cyclic and BCH codes up to their actual minimum distance is presented. Previous algebraic decoding procedures for cyclic and BCH codes such as the Peterson decoding procedure and our procedure using nonrecurrent syndrome dependence relations can be regarded as special cases of this new decoding procedure. With the aid of a computer program, it has been verified that, using this new decoding procedure, all binary cyclic and BCH codes of length 63 or less can be decoded up to their actual minimum distance. The procedure incorporates an extension of our Fundamental Iterative Algorithm and the complexity of this decoding procedure is $O(n^3)$.

Summary

For some years, algebraic decoding of cyclic and BCH codes has been restricted by the minimum distance bounds of the codes. Previous algebraic decoding algorithms (Berlekamp-Massey, Euclidean, and our generalizations) have aimed at solving Newton's identities which can be viewed as a set or sets of linear recurrences. We have recently introduced a procedure that frees the decoding of cyclic and BCH codes from the confinement of the bounds and can decode many cyclic and BCH codes up to their actual minimum distance [1]. In our recent procedure, the decoding is accomplished through the determination of nonrecurrent dependence relations among the syndromes. However, the application of this procedure depends on a condition that has to be satisfied for a code to be so decoded. Thus, that decoding procedure is still short of the desired final goal on achieving decoding of all cyclic and BCH codes up to their actual minimum distance. In this paper, we present a new decoding procedure that does not depend on the satisfaction of this condition. We show that, for a code with actual minimum distance d to correct up to $t = \lfloor (d-1)/2 \rfloor$ errors, all that is required is that a (2t+1)×(2t+1) syndrome matrix can be so formed that the syndromes above the minor diagonal are all known and those at the minor diagonal are some unknowns and their conjugates. With reference to the table of codes listed in van Lint and Wilson's paper [2] and with the aid of a computer program, the existence of at least one such matrix for each code has been verified for all binary codes of length 63 or less. Thus, to say the least, the procedure is capable of decoding all binary cyclic and BCH codes of length ≤ 63 up to their actual minimum distance. We have also demonstrated the existence of such syndrome matrices for some codes of length greater than 63. The procedure is a very general one and includes previously mentioned algebraic decoding procedures as special cases. It can be applied to the decoding of codes of any length for which such syndrome patterns exist.

This work was supported by the National Science Foundation under Grant NCR-9016095.

More specifically, the syndrome matrix S referred to in this paper is of the following form:

$$\begin{bmatrix} S_b & S_{b+i_1} & S_{b+i_2} & \cdots & S_{b+i_{2t-2}} & S_{b+i_{2t-1}} & S_{b+i_{2t}} \\ S_{b+j_1} & S_{b+i_1+j_1} & S_{b+i_2+j_1} & \cdots & S_{b+i_{2t-2}+j_1} & S_{b+i_{2t-1}+j_1} & \\ S_{b+j_2} & S_{b+i_1+j_2} & S_{b+i_2+j_2} & \cdots & S_{b+i_{2t-2}+j_2} & & \\ \vdots & \vdots & \vdots & & \vdots & & \\ S_{b+j_{2t-1}} & S_{b+i_1+j_{2t-1}} & & & & & \\ S_{b+j_{2t}} & & & & & & S_{b+i_{2t}+j_{2t}} \end{bmatrix}$$

where the triangular portion of S above the minor diagonal consists of known syndromes and the syndromes at the minor diagonal of S are some unknowns and their conjugates.

Under the assumption that ν errors actually occurred where $\nu \le t$, then there exist at most ν columns of S which are linearly independent. The other columns are then dependent on these columns. A major step for this decoding procedure is then to determine the unknown syndromes through the linear dependence relations among the columns of S. In this paper, we show that this can be accomplished through an extention of the Fundamental Iterative Algorithm we first introduced in [3].

Once $S_0, S_1, S_2, \cdots, S_{n-1}$ are computed, the error vector can be determined through an inverse Fourier transform of the syndrome vector $(S_0, S_1, S_2, \cdots, S_{n-1})$.

We note that the decoding of the (41,21,9) quadratic residue code [4] can be much more easily handled by this new procedure.

References

[1] G.L. Feng and K.K. Tzeng, "Decoding cyclic and BCH codes up to actual minimum distance using nonrecurrent syndrome dependence relations," *IEEE Trans., Inform. Theory*, vol. IT-37, pp. 1716-1723, Nov. 1991.

[2] J. van Lint and R.M. Wilson, "On the minimum distance of cyclic codes," *IEEE Trans., Inform. Theory*, vol. IT-32, pp. 23-40, Jan. 1986.

[3] G.L. Feng and K.K. Tzeng, "A Generalization of the Berlekamp-Massey Algorithm for Multisequence Shift-Register Synthesis with Applications to Decoding Cyclic Codes," *IEEE Trans., Inform. Theory*, vol. IT-37, pp. 1274-1287, Sept. 1991.

[4] I.S. Reed, T.K. Truong, X. Chen, and X. Yin, "The Algebraic Decoding of the (41,21,9) Quadratic Residue code," *IEEE Trans., Inform. Theory*, vol. IT-38, pp. 974-986, May 1992.

A New Remainder Based Decoding Algorithm for Reed-Solomon Codes

by

Tomik Yaghoobian and Ian F. Blake
Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario

Abstract

Conventional decoding techniques for decoding cyclic codes require the computation of power sum syndromes which can often account for a significant portion of the decoder computations. Since the syndromes can be computed from the remainder polynomial, the polynomial obtained by dividing the received polynomial by the code generator polynomial, it follows that this polynomial contains all the information required to decode. Thus one might hope for a decoding technique that uses the remainder polynomial directly. Berlekamp and Welch have given such an algorithm which requires the sequential testing of the parity check locations and updating of four polynomials. Whiting in his doctoral thesis has given a modification of this procedure that makes more efficient the evaluation and updating of these polynomials.

The present work derives a new algorithm using only the remainder polynomial. A new key equation is derived which may be solved by the usual Euclidean algorithm. The advantages of this approach are discussed and compared to the original algorithm and a performance of the algorithm in terms of computational and circuit complexity is considered.

Summary

Conventional decoding techniques for cyclic codes require the computation of the power sum syndromes. These are given by the evaluation of the remainder polynomial, the polynomial obtained by dividing the received word by the code generator polynomial, at the roots of the generator polynomial. Such computations can absorb a significant part of the decoding effort. As the remainder polynomial contains all the information required to decode, it might be hoped to derive a technique that uses the remainder polynomial directly. Berlekamp and Welch [1],[2] devise such an algorithm which requires a sequential test of the parity check locations, the evaluation of discrepancies at these locations and the updating of various estimates. The procedure involves four polynomials and there appears to be no obvious way to split the algorithm into the more conventional determination of the error locator and evaluator polynomials via either the Berlekamp-Massey or Euclidean algorithm approach, followed by a Chien search. The Berlekamp-Welch algorithm was further considered in the thesis by Whiting [3], which also contains an excellent description of the algorithm itself. (As far as the authors are aware, the algorithm itself has not appeared in the literature.) He devised an efficient linear scaling technique for the updating of the polynomials.

The present work derives a new remainder based algorithm. It develops a key equation that uses the remainder polynomial, expressed in terms of the Lagrange interpolation polynomials, which allows solution for the error locator and evaluator polynomials by the conventional Euclidean algorithm technique. In particular it shows that if the remainder polynomial of the Reed-Solomon code with parameters $(n, k, d = n - k + 1 = 2t + 1)$, then

$$F(x) = \sum_{k=0}^{d-2} r_k h_k(x) , \quad h_k(x) = c_k \prod_{i=0, i \neq k} (x - \alpha^i)/(\alpha^k - \alpha^i)$$

where c_k depends on the code parameters only and $h_k(x)$ is a Lagrange polynomial, then if there exist polynomials

$$N(x),\ W(x),\ \deg N(x) < \deg W(x) \leq t$$

which satisfy the key equation

$$W(x)F(x) \equiv N(x) \bmod p(x) = (x-1)(x-\alpha)\cdots(x-\alpha^{2t-1})$$

then $W(x)$ is the error locator and $N(x)$ is closely related to the usual error evaluator polynomial. Thus the Reed-Solomon code can be decoded directly from the remainder polynomial, bypassing the need for the syndrome polynomial.

The relationship of this algorithm to the usual form of the Berlekamp-Welch algorithm and the implications of this form of decoding in terms of implementation are discussed and examples are given.

References

[1] L.R. Welch and E.R. Berlekamp, A new Reed-Solomon decoding algorithm, International Symposium on Information Theory, Ste. Jovite, Quebec, 1983.

[2] E.R. Berlekamp and L.R. Welch, Error correction for algebraic block codes, *US Patent No. 4,633,470*, September, 1983.

[3] D.L. Whiting, Bit-serial Reed-Solomon decoders in VLSI, *Ph.D. Thesis, California Institute of Technology, California*. 1985.

Inverterless Cauchy Cells for A Systolic Reed-Solomon Encoder

M. A. Hasan and V. K. Bhargava
Department of Electrical & Computer Engineering
University of Victoria
Victoria, BC, Canada

Summary

Consider an $(n,\ k)$ Reed-Solomon (RS) code of length $n = q - 1$ and redundancy $r = n - k$ over the finite field GF(q). The usual implementation of the RS encoder consists of an r stage feedback shift register [1]. In some very high speed applications, the presence of the accompanying global feedback path restricts the speed of the encoder.

Recently, Seroussi has proposed an architecture for the RS encoder [2]. Unlike the usual implementation of the RS encoder, Seroussi's architecture does not require any global feedback path. Furthermore, the architecture is of systolic type and has modular a structure– it consists of one pre-processing cell and r *Cauchy* cells [2]. This modularity feature of the encoder makes it suitable for hardware implementation.

The circuit complexity of Seroussi's RS encoder depends essentially on the Cauchy cells. Each Cauchy cell computes one parity symbol for the RS code and contains one parallel type divider for the finite field GF(q). Unfortunately, the realization of a divider is much more complicated than that of a multiplier [3]. Let M denote the circuit complexity of a parallel type multiplier of GF(q), where $q = p^m$, p is prime and m is a nonzero positive integer. Then the circuit complexity of a *modular* parallel divider is, in general, $O(mM)$ and that of Seroussi's RS encoder is $O(rmM)$.

In this paper, we extent Seroussi's work. It is shown here that the Cauchy cell can be implemented without any divider. The proposed Cauchy cell also has a shorter logic path and yields an RS encoder which has a circuit complexity $O(rM)$.

Let $\alpha,\ \alpha^2,\ \cdots, \alpha^{r-1}$ be the roots of the RS code where α is a primitive element of GF(q). In systematic form, the generator matrix of the RS code can be written as

$$\mathbf{G} = [\mathbf{I}|\mathbf{A}],$$

where $\mathbf{I}$ is the $k \times k$ identity matrix and $\mathbf{A}$ is a $k \times r$ matrix where the matrix elements belong to GF(q). $\mathbf{A}$ is called a *Cauchy* matrix and its elements are given as follows [2]:

$$A_{i,j} = \frac{u_i v_j}{x_i + y_j}, \quad 0 \le i \le k-1,\ 0 \le j \le r-1, \qquad (1)$$

where

$$x_i = -\alpha^{n-1-i}, \quad 0 \le i \le k-1,$$

$$y_j = \alpha^{n-1-k-j}, \quad 0 \le j \le r-1,$$

$$u_i = \frac{1}{\prod\limits_{\substack{0 \le l \le k-1 \\ l \ne i}} (\alpha^{n-1-i} - \alpha^{n-1-l})}, \quad 0 \le i \le k-1,$$

$$v_j = \prod_{0 \le l \le k-1} (\alpha^{n-1-k-j} - \alpha^{n-1-l}), \quad 0 \le j \le r-1.$$

Consider the codeword $(d_0,\ d_1,\ \cdots,\ d_{k-1},\ p_0,\ p_1,\ \cdots,\ p_{r-1})$, where d_i $(i = 0,\ 1,\ \cdots,\ k-1)$ and p_i $(i = 0,\ 1,\ \cdots,\ r-1)$ are the data and parity symbols, respectively. For $0 \le i \le k-1$, $0 \le j \le r-1$, let

$$B_{i+1,j} = (x_i + y_j) B_{i,j}, \qquad (2)$$

$$R_{i+1,j} = (x_i + y_j) R_{i,j} + d_i u_i r_{i,j}, \qquad (3)$$

with $B_{0,j} = 1$ and $R_{0,j} = 0$. Then it can be shown that

$$p_j = R_{k,j}, \quad j = 0,\ 1,\ \cdots,\ r-1. \qquad (4)$$

As the computation of $R_{i+1,j}$ $(0 \le i \le k-1,\ 0 \le j \le r-1)$ requires only multiplication and addition operations, the Cauchy cells, each of which computes one parity symbol, can be designed without any divider/inverter.

The computation of $d_i u_i$ in (3) is done in the pre-processing cell of the RS encoder, and one Cauchy cell can recursively compute one parity symbol with two multiplications and two additions in each time step. However, the multiplications can be performed in parallel resulting in a logic path consisting of one multiplier and two adders. This logic path is shorter than that of [2] which consists of one divider and two adders.

References

[1] E. R. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[2] G. Seroussi, "A systolic Reed-Solomon encoder," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1217-1220, July 1991.

[3] M. A. Hasan and V. K. Bhargava, "Bit-serial systolic divider and multiplier for GF(2^m)," *IEEE Trans. Comput.*, pp. 972-980, Aug. 1992.

A NEW LOOK AT THE KEY EQUATION

Patrick Fitzpatrick,
IFI Institute of Advanced Microelectronics,
NMRC, Cork, Ireland.
email: fitzpat@bureau.ucc.ie

Abstract

We describe a new algorithm based on Gröbner bases of modules for solving for the pair (a,b) the multivariable polynomial congruence $as \equiv b \bmod I$ where I is an ideal in $k[x_1,\ldots,x_n]$ and s is given. The restriction to one variable gives a new approach to decoding BCH and (classical) Goppa codes.

1 Introduction

Let $\mathbf{A} = k[x_1,\ldots,x_n]$ where k is a field. Interpreting a polynomial s of total degree m as a truncation of a formal power series we may ask for relatively prime polynomials a,b where $a(0) \neq 0$ such that the expansion of b/a as far as terms of degree m is equal to s. This problem may be regarded as a special case of solving (for a and b) the congruence

$$as \equiv b \bmod I, \tag{1}$$

where I is an ideal in $\mathbf{A}$ and s is a given polynomial. In general we require that a and b be relatively prime but drop the condition that $a(0) \neq 0$.

In the 1-variable problem, I is the ideal generated by a single polynomial and it is well known that in this case the solution may be determined using the extended Euclidean algorithm or the Berlekamp-Massey algorithm; neither of these is valid for $n > 1$. Sakata [3] has given an extension to n variables of the Berlekamp-Massey algorithm and in [2] we gave a corresponding generalization of the method based on the extended Euclidean algorithm. The natural context for such a generalization is that of *Gröbner* bases of polynomial modules. We proved in [2] that Gröbner bases of modules can be used to solve congruence (1) for any ideal I.

The 1-variable technique can be applied to the decoding problem for BCH and (1-variable) Goppa codes where it provides a new theoretical derivation of an algorithm for solving the key equation which *in practice* is equivalent to that based on the extended Euclidean algorithm. This new theoretical underpinning is in a sense more "natural" than either of the two classical methods. In the following sections we outline a direct derivation of this algorithm. For further details *cf.* [1].

2 The solution module

The 1-variable form of congruence (1) is

$$as \equiv b \bmod y. \tag{2}$$

where s (the syndrome polynomial) has degree at most $2t-1$ and y is a polynomial of degree $2t$ (x^{2t} or the Goppa polynomial). The set of *all* pairs (a,b), without restriction, satisfying (2) forms an $\mathbf{A}$-module G. Any such module has a finite basis and it is easy to prove that $B = \{(0,y),(1,s)\}$ is a basis of G. Our aim is to determine from B another basis B' which contains (a scalar multiple of) the specific required solution (σ,ω) — in the usual notation — where $\sigma(0) = 1$.

We impose a total order $<$ on the set of *terms* $\{(x^r,0),(0,x^r)\}$ in $\mathbf{A}^2$ by interleaving them as follows:

$$(1,0) < (0,1) < (x,0) < (0,x) < (x^2,0) < (0,x^2) < \cdots$$

This ensures that $(x^p,0) > (0,x^q)$ if and only if $p > q$. Each pair (a,b) can be expressed as a finite sum $\Sigma\gamma_j \mathbf{t}_j$ where $\gamma_j \in k$ and $\mathbf{t}_j$ is a term and the *leading term* of a pair (a,b) is that term in the decomposition which is greatest relative to $<$. If the leading term of (a,b) has the form $(x^r,0)$ we say its leading term is *on the left* while if it has the form $(0,x^r)$ then its leading term is *on the right*. Note that the leading term of (a,b) is on the left if and only if $\delta a > \delta b$ (where δ denotes degree).

A pair (a,b) can be *reduced* by a pair (a',b') if the leading term of (a',b') is on the left and $\delta a \geq \delta a'$, or if the leading term of (a',b') is on the right and $\delta b \geq \delta b'$. The *reduction step* will be defined with reference to the left hand side—an analogous definition applies on the right. Suppose $a = a_l x^l + \cdots + a_0, a' = a'_m x^m + \cdots + a'_0$ where $a_l \neq 0, a'_m \neq 0, l \geq m$ and the leading term of (a',b') is on the left. Then we say (a,b) is *reduced by* (a',b') *to* $(a'',b'') = (a,b) - (a_l/a'_m)x^{l-m}(a',b')$. It is clear that $\delta a'' < \delta a$.

We call a basis D of a module M a *reduced basis* if none of its elements can be reduced by any other.

Theorem 2.1 *(i) Let M be a module. Then a reduced basis of M consists either of a single element or of two elements $\{(a_1,b_1),(a_2,b_2)\}$ where (a_1,b_1) has leading term on the left and (a_2,b_2) has leading term on the right. Moreover, in the latter case $\delta a_1 > \delta a_2$ and $\delta b_1 < \delta b_2$.*
(ii) Let D be a reduced basis of M and let $(a,b) \in M$. Then the leading term of (a,b) is a multiple of the leading term of an element of D.
(iii) A reduced basis of M is a Gröbner basis.

Since we are interested in the module G arising in the decoding application we may assume that G contains an element (σ,ω) where σ and ω are relatively prime and $\delta\sigma \leq t, \delta\omega < \delta\sigma$. In particular the leading term of (σ,ω) is on the left. (We use the condition $\sigma(0) = 1$ in Step 2 of the algorithm below.)

Theorem 2.2 *(σ,ω) is an element of least leading term in G. Every reduced basis of G contains a scalar multiple of (σ,ω).*

As a consequence of this theorem we have the following algorithm for solving the key equation.

Algorithm
1. Reduce the basis $B = \{(0,y),(1,s)\}$ to a reduced basis $B' = \{(u_1,v_1),(u_2,v_2)\}$ where (u_1,v_1) has leading term on the left.
2. Set $(\sigma,\omega) = u_1(0)^{-1}(u_1,v_1)$.

References

[1] P. Fitzpatrick, A new derivation of an algorithm for solving the key equation, (submitted for publication).

[2] P. Fitzpatrick, J. Flynn, A Gröbner basis technique for Padé approximation, *J. Symbolic Computation*, 13 (1992), 133-138.

[3] S. Sakata, Extension of the Berlekamp-Massey algorithm to n dimensions, *Information and Computation*, 84 (1990) 207-239.

ON THE MINIMUM CODE LENGTH OF S-STEP (T, U) PERMUTATION DECODABLE CYCLIC CODES

Anader Benyamin-Seeyar, Tho Le-Ngoc, and Ming Jia
Department of Electrical and Computer Engineering, Concordia University,
1455 de Maisonneuve West, Montreal, Quebec, Canada H3G 1M8

Summary

One variation of error-trapping decoding which is known as "permutation decoding" was introduced by Prange[1]. A serial decoder based on this treatment was given by MacWilliams, who made use of code preserving (T, U) permutation sets to obtain k error-free positions from which the rest of the codeword could be reconstructed [2]. Recently, the exact lower bounds on the code length n for two and three steps (T, U) permutation decodable (PD) $(n, k, 2t+1)$ cyclic codes have been found [3-5]. In this paper, we extend those results on the code length for any s-step (T, U) permutation decodable cyclic codes with odd valued error-correcting capability t. In addition, an optimum permutation step which makes the most efficient improvement in the code rate of PD cyclic code is also given. Since the derivation of these results involves only the error position, the results are applicable to cyclic codes over GF(2^m).

Main Results

The exact lower bounds on the code length n for two and three step (T, U) permutation decodable $(n, k, 2t + 1)$ cyclic codes are given in [3, 4]. Here we extend the results on the code length for s-step (T, U) permutation decodable cyclic codes with odd-valued t only. The results are presented in the form of theorems. Note that the subscrips "e" and "o" are used in the manuscript to indicate the even and odd valued variables respectively.

Theorem 1: The $(n, k_e, 2t_o + 1)$ codes with $n = t_o(k_e - 2^{s-1} + 1) + 2^{s-1}$, $k_e \geq t_o(2^{s-2} - 1) + 1$ for $t_o = 5$ and $k_e \geq t_o(2^{s-2} - 1) - 2^{s-2} + 3$ for $t_o \geq 7$, are s-step PD.

Theorem 2: The $(n, k_o, 2t_o + 1)$ codes with $n = t_o(k_o - 2^{s-1} + 2) + 2^{s-1}$, $k_o \geq t_o(2^{s-2} - 1)$ for $t_o = 5$ and $k_o \geq t_o(2^{s-2} - 1) - 2^{s-2} + 2$ for $t_o \geq 7$, are s-step PD.

Theorem 3: The $(n, k_o, 2t_o + 1)$ codes with $n = t_o(k_o - 2^{s-1} + 2) + 2^{s-1}$, $k_o = t_o(2^{s-2} - 1) - 2^{s-2}$ for $t_o \geq 7$, are not s-step PD.

Corollary: The $(n, k_e, 2t_o + 1)$ codes with $n = t_o(k_e - 2^{s-1} + 1) + 2^{s-1}$, $k_o = t_o(2^{s-2} - 1) - 2^{s-2}$ for $t_o \geq 7$, are not s-step PD.

Now we present the bounds on the code rate R_c for binary $(n, k, 2t + 1)$ permutation decodable cyclic codes.

Theorem 4: The code rate of s-step PD codes with $k > k^*$ and $t_o = 5$ is less than $1/(t_o - 2 + 2/t_o)$, where $k_o^* = t_o(2^{s-2} - 1) - 2$ and $k_e^* = k_o^* - 1$.

Theorem 5: The code rate R_c of s-step PD codes with $k > k^*$ and $t_o \geq 7$ is less then $1/(t_o - 2)$, where $k_o^* = t_o(2^{s-2} - 1) - 2^{s-2}$ and $k_e^* = k_o + 1$.

The above results present the bounds on the code length and the code rate of s-step PD codes. Clearly, when permutation steps s increases, R_c increases. Next, we show how R_c increases with respect to s.

Suppose that the code rate of a s-step PD is R_{cs}, then $\Delta R_{cs} = R_{c(s+1)} - R_{cs}$ is not the same for different s. This can shown as follows:

$$\frac{\partial R_c}{\partial s} = \frac{k_o 2^{s-2}(t_o - 1)\ln 2}{n^2} > 0 \tag{5}$$

$$\frac{\partial^2 R_c}{\partial s^2} = \frac{k_o(k_o + 2)t_o(t_o - 1)2^{s-2}\ln^2 2}{n^3} > 0 \tag{6}$$

So when $k > k^*$, ΔR_{cs} increases as the permutation steps increases. Consider the case of $t_o = 7$ as an example (the case of $t_o = 3$ and 5 are similar), for s-step permutation,

$$k_s^* = t_o(2^{s-2} - 1) - 2^{s-2} \tag{7}$$

$$\Delta k_s^* = k_{s+1}^* - k_s^* = k_s^* + t_o \tag{8}$$

Therefore, when each more permutation is applied, k^* increases $k^* + t_o$.

On the other hand, when $k < k^*$, the code length n does not keep decreaseing at the rate of $\Delta n = (\Delta k)t$, so the increasing in the code rate R_c is very slow. That is to say, ΔR_{cs} increases as permutation steps s increases, until to this extent that k^* large enough and $k \leq k^*$ being satisfied. The step corresponding to the largest ΔR_{cs} is the optimum permutation step which makes the most efficient improvement in the code rate of PD codes. When the error correcting capability of a code is goven as t_o, k^* corresponding to every permutation steps s is known; therefore, the optimum permutation step can be estimated. From the implementation point of view, this optimum step determines the number of U- perumtation steps used for achieving higher code rate R_c.

From the results above, the relation between R_c, k and s as shown in Figure 1 can be obtained, where $R = 1/(t_o + 2 + 2/t_o)$ and $R' = 1/(t_o - 2)$ for $t_o = 5$ and $t_o \geq 7$ respectively. For the region of $k < k^*$, the code rate of PD codes is around R.

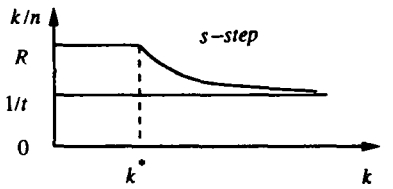


Figure 1: The relation between R_c, k and s

As conclusions, the following comments can be made:

1. When $k \to \infty$, the code rate of s-step PD codes decreases and approaches the code rate of the codes which can be decoded by error-trapping decoding.

2. When $k \to k_+^*$, R_c approaches to $1/(t_o - 2 + 2/t_o)$ and $1/(t_o - 2)$ for $t_o = 5$ and $t_o \geq 7$ respectively.

3. For each additional step, k^* increases more than double. In the region of $k < k^*$, the code length n does not keep decreasing at the rate of $\Delta n = (\Delta k)t_o$, so the increasing in the code rate R_c is very slow and R_c is around the value R.

4. $\Delta R_{cs} = R_{c(s+1)} - R_{cs}$ increases as the permutation steps s increases until k^* is large enough and $k < k^*$ being satisfied. Therefore, there existes an optimal step which makes the most improvement in the code rate of PD codes. Given a t-error correcting $(n, k, 2t + 1)$ cyclic code, then the optimum step s can be determined. In this way the steps needed to decode a certain code can be estimated.

References

[1] E. Prange, "The use of Information Sets in Decoding Cyclic Codes," *IEEE Trans. Inform. Theory*, vol. 8, pp. 85-89, Sept. 1962.

[2] F. J. MacWilliams, "Permutation Decoding of Systematic Codes," *Bell Syst. Tech. J.*, vol. 43, part I, pp. 485-505, Jan. 1964.

[3] A. Benyamin-Seeyar, S. S. Shiva, and V. K. Bhargava "Capability of the Error-trapping Technique in Decoding Cyclic Codes," *IEEE Trans. Inform. Theory*, vol. 32, No. 2, pp. 166-180, Mar. 1986.

[4] M. Jia, A. Benyamin-Seeyar, and T. Le-Ngoc, "Exact Lower Bounds on the Codelength of Three-Step Permutation-Decodable Cyclic Codes," *IEEE Trans. Inform. Theory*, vol. 38, No. 6, pp. 1812-1817, Nov. 1992.

EFFICIENT CODING/DECODING STRATEGIES FOR CHANNELS WITH MEMORY *

Cuong Hon Lai and Samir Kallel
Department of Electrical Engineering
The University of British Columbia
Vancouver, B. C., Canada, V6T 1Z4

Abstract

Many digital communication channels are affected by errors that tend to occur in bursts. This paper proposes two new schemes for burst error correction. Both schemes employ a combination of two codes. In the first scheme, one of the codes is used for random error correction and for burst detection while the other one is used only for burst recovery. In the second scheme, one of the codes is used for burst detection and for channel state estimation. With the second scheme, both codes are used for error correction. Unlike existing burst-error-correcting schemes, it is shown that the proposed schemes are adaptive to channel conditions and less sensitive to errors in the guard space. For the same delay, the proposed schemes offer better performance than the interleaving schemes. When the channel is heavily corrupted by bursts, the improvement is even more pronounced.

Summary

Many digital communication channels are affected by errors that tend to occur in clusters or bursts [1]. Several schemes for burst error correction on these channels have been reported [2–5]. One approach is to use special codes designed exclusively for burst error correction [4]. These so-called burst-error-correcting codes perform relatively well over channels with short bursts, but perform poorly when the channels are corrupted with long bursts. Another conventional approach is to interleave channel symbols prior to transmission. With interleaving, burst errors are spread over many symbols, and can thus be viewed as random errors. However, for channels with long bursts, interleaving schemes need extremely long delay to be effective, which might not be tolerated in some applications. Another approach is Gallager's burst-finding scheme [2]. In this scheme, a rate 1/2 systematic convolutional code is used with a modified majority logic decoding. Gallager's scheme sacrifices random-error-correcting capability in exchange for better burst correction. A modified version of this scheme was recently suggested by Schlegel and Herro [5]. This scheme is essentially the same as Gallager's scheme except that majority logic decoding is replaced by a modified Viterbi decoding algorithm. Both Gallager's burst-finding scheme and Schlegel and Herro's scheme are extremely sensitive to errors in the guard space.

Two efficient coding and decoding strategies are proposed in this paper. Both schemes employ a combination of two punctured convolutional codes and a burst detection procedure. Burst detection is accomplished by observing the increment in the cumulative path metrics from Viterbi decoding. Scheme 1 uses two punctured convolutional codes with different memories. In this scheme, a code with a relatively short memory is used with Viterbi decoding for random error correction and for burst detection. The other code which has a much longer memory is used with backward sequential decoding to recover burst errors. Normally, the decoder operates in the random mode and it uses the received sequence corresponding to the code with the shorter memory. An abrupt increase in the cumulative path metrics indicates that the channel is most likely in a burst. The decoder then switches from the random mode to the burst mode, and starts burst error recovery. In the burst mode, starting from a chosen state, the decoder employs a backward sequential decoding algorithm to recover the corrupted data. When the channel becomes quiet, the path metrics are relatively constant, and the decoder returns to the random mode.

Scheme 2 uses two punctured codes that are derived from the same original convolutional code with complementary perforation patterns. One code sequence is transmitted after a delay from the transmission of the other code sequence. The first code sequence is used with a Viterbi decoder to detect bursts using the same burst detection procedure as in Scheme 1. The burst detection procedure also serves for estimating the channel state. Both received sequences are then used by a second Viterbi decoder which uses the channel state information provided by the first decoder.

The proposed schemes are adaptive to channel conditions. The parameters of the decoders can be chosen to optimize the performance of the schemes. For the same delay, these schemes outperform the conventional interleaving schemes when the channel is heavily corrupted by bursts. While Gallager's burst-finding scheme and Schlegel and Herro's scheme are sensitive to errors in the guard space, the proposed schemes can tolerate high error rates in the guard space.

References

[1] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. of IEEE*, vol. 66, pp. 724–744, July 1978.

[2] R. G. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[3] G. D. Forney, "Burst-correcting codes for the classic bursty channel," *IEEE Trans. Comm.*, vol. 19, pp. 772–781, Oct. 1971.

[4] S. Lin and D. J. Costello, *Error Control Coding*. NJ: Prentice-Hall, 1983.

[5] C. B. Schlegel and M. A. Herro, "A burst-error-correcting Viterbi algorithm," *IEEE Trans. Comm.*, vol. 38, pp. 285–291, Mar. 1990.

* This research was supported by the National Sciences and Engineering Research Council of Canada.

Comparison of Erasure-and-Error Decoding Schemes

Takeshi Hashimoto

Dept. Elect. Eng., Univ. Electro-Communications
Chofugaoka 1-5-1, Chofu, Tokyo 182, JAPAN
e-mail: hasimoto@liszt.ee.uec.ac.jp

Erasure-and-error decoding is a general form of decoding for reliable communications and, at the same time, the basis of important channel coding schemes such as coded (or hybrid) ARQ and concatenated coding. There are several schemes discussed in the context of information theory. Those are Forney's scheme, the threshold decoding discussed in Gallager's textbook, the likelihood-ratio decision, the use of error-detecting codes, and their modifications. Most of the schemes may be described, in terms of a reliability measure $Q(\mathbf{y}, m)$, in such a manner that the decoded message is accepted only if $Q(m) > T$ for a specified T and an erasure is declared otherwise. For example, $Q(\mathbf{y}, m) = \log \frac{P(\mathbf{y}|\mathbf{x}_m)}{\sum_{m' \neq m} P(\mathbf{y}|\mathbf{x}_{m'})}$ in Forney's scheme, $Q(\mathbf{y}, m) = \log \frac{P(\mathbf{y}|\mathbf{x}_m)}{q(\mathbf{y})}$ in the threshold decoding, and $Q(\mathbf{y}, m) = \log \frac{P(\mathbf{y}|\mathbf{x}_m)}{\max_{m' \neq m} P(\mathbf{y}|\mathbf{x}_{m'})}$ in the likelihood-ratio decision. An interesting variation of the likelihood-ratio decision is Kudryashov's scheme where $Q(\mathbf{y}, m) = \log \frac{P(\mathbf{y}|\mathbf{x}_m)}{\max_{m' \neq m} P^{1-\sigma}(\mathbf{y}|\mathbf{x}_{m'}) f^{\sigma}(\mathbf{y})}$ is used. Between these schemes, Forney's scheme is known to be the best one but requires much computation. Thus, suboptimal, but simpler schemes are preferred in real applications. However, when we are to select one between these schemes, we realize that there do not exist enough discussions on the relationship between the respective performances.

In this paper, we consider the upper bounds of P_{era}, erasure probability, and P_{uer}, undetected error probability, of the respective schemes and compare them in a systematic manner. Known bounds for the respective schemes are insufficient for our purpose. For example, most of them are presented in terms of different exponent functions, the performance is under-estimated in the case of the likelihood-ratio decision, there are some confusions concerning to the analysis of threshold decoding as seen in Gallager's textbook, and the performance bound of the scheme based on error-detecting codes has not been considered in the Shannon-theoretic context. A reason for some of them is that the performance of an erasure-and-error decoding scheme is frequently considered in terms of the asymptotic error exponent as the blocklength goes to infinity and, at the same time, the exponent of the erasure probability goes to zero. Even though this gives the best theoretically attainable error exponent, it usually has nothing to do with the observable performance because of the decoder complexity.

We consider the performance bound of a given scheme in the following form (or its minor variation) such that

$$P_{uer} \leq \exp\{-N E_{scheme}(R, R_o)\}$$

for R_o satisfying

$$P_{era} \leq \exp\{-N E_{sp}(R_o)\},$$

where R is the coding rate and N is the block length. We call E_{scheme} the error exponent of the scheme. This form of the bound is really required in many applications since the tradeoff between P_{era} and P_{uer} is an important problem.

In the discussion, we carefully distinguish several threshold decoding schemes. Although the same $Q(\mathbf{y}, m)$ is used, there are the weak threshold (Th) decision, the strong threshold (STh) decision where the uniqueness of the decision is required, and the maximum likelihood threshold (MlTh) decision where the maximum likelihood m is tested. These show somewhat different characteristics and performances.

As a result of analysis, we show that $E_{Th}(R, R_o) \leq E_{STh}(R, R_o)$, that $E_{STh}(R, R_o)$, $E_{MlTh}(R, R_o)$, and $E_{Ed}(R, R_o)$ for the scheme based on error-detecting codes, have almost the same bound, and that $E_{Forney}(R, R_o)$ is strictly superior to the rest. An interesting point is, for a binary symmetric channel, that $E_{Lr}(R, R_o)$ for the likelihood-ratio decision and $E_{Kudryashov}(R, R_o)$ are very close to $E_{Forney}(R, R_o)$. A reason of the last result may be that $Q(\mathbf{y}, m)$ is basically the ratio of likelihood functions in these schemes.

Fault-Tolerant Distributed Decoding of Cyclic Block Codes

Ahsun H. Murad and Thomas E. Fuja†
Department of Electrical Engineering / Systems Research Center
University of Maryland, College Park, MD 20742

Abstract

Suppose one is given M (possibly corrupted) codewords from M (possibly different) codes, each over $\mathbf{F}_q$; suppose further that the codewords have a single symbol in common. The *common-symbol decoding problem* is that of estimating the symbol in the common position. In [1], a solution to this problem was presented for a very restricted case. This talk presents a general solution that contains the familiar one-step majority-logic decoding as a special case. This algorithm leads naturally to a decoder structure suitable for fault-tolerant decoding of cyclic block codes; the resulting architecture undergoes graceful degradation with increasing component failures. Bounds on decoder performance under various kinds of partial decoder failures are presented.

Summary

One-step majority-logic decoding [2], is one of the simplest algorithms for decoding cyclic block codes. However, it is an effective decoding scheme for very few codes. In [1], the authors presented a generalization based on the following *common-symbol decoding approach*.

Given $C_1, C_2, \ldots, C_M$ a set of M linear block codes over $\mathbf{F}_q$, let $\mathbf{c}_1 \in C_1, \mathbf{c}_2 \in C_2, \ldots, \mathbf{c}_M \in C_M$ be a set of codewords with the first symbol in common—i.e., $c_{1,1} = c_{2,1} = \cdots = c_{M,1}$. The symbols making up the codewords are transmitted over a channel (the common-symbol only once), and errors occur. Let $\mathbf{r}_i = \mathbf{c}_i + \mathbf{e}_i \quad \forall i = 1, 2, \ldots, M$ be the received, corrupted codewords. (Of course, $r_{1,1} = r_{2,1} = \cdots = r_{M,1}$.)

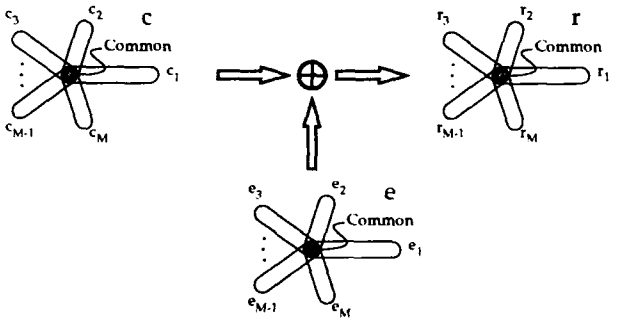


Fig 1. M codewords sharing a symbol being transmitted over a channel.

It was shown in [3] that the common-symbol can be estimated *correctly* provided no more than $\lfloor(\delta-1)/2\rfloor$ errors occur in all the M codewords. Furthermore, if δ is even, and $\delta/2$ errors occur, then such an event can be detected. Here,

$$\delta \triangleq \sum_{i=1}^{M} \delta_i \quad -(M-1), \qquad \text{where} \qquad \delta_i \triangleq \min_{\substack{\mathbf{x},\mathbf{y}\in C_i \\ x_1 \neq y_1}} \mathrm{d_H}(\mathbf{x},\mathbf{y}).$$

That is, δ_i is the minimum Hamming distance between codewords in C_i, that differ in the first (common) position. The *common-symbol decoding problem* is that of performing this decoding. (In [1], a solution was provided for the simple case of $q = 2, M = 2$, and $\delta_1 = \delta_2$.)

Let $\mathcal{D}_i$ be a bounded-distance decoder for the code C_i defined as:

$$\mathcal{D}_i(\mathbf{r}_i) \triangleq \begin{cases} \underset{\mathbf{x}\in\mathbf{S}_i(\mathbf{r}_i)}{\operatorname{argmin}}\ \mathrm{d_H}(\mathbf{x},\mathbf{r}_i) & \text{if } \mathbf{S}_i(\mathbf{r}_i) \neq \varnothing; \\ ? & \text{otherwise;} \end{cases} \qquad \text{where}$$

$$\mathbf{S}_i(\mathbf{z}) \triangleq \{\mathbf{y} \in C_i : \mathrm{d_H}(\mathbf{y},\mathbf{z}) \leq \left\lfloor \frac{\delta_i - 1}{2} \right\rfloor, \text{ or } \mathrm{d_H}(\mathbf{y},\mathbf{z}) = \frac{\delta_i}{2}, y_1 \neq z_1\}.$$

That is, (i) if there exist codewords within $\lfloor(\delta_i-1)/2\rfloor$ of the received n_i-tuple $\mathbf{r}_i$, then $\mathcal{D}_i$ maps $\mathbf{r}_i$ onto the closest of them; (ii) if not, and δ_i is even, then it erases the first symbol and tries again; (iii) If both of the above fail, then the error is uncorrectable (indicated by mapping to ?). Next, define the following:

(i) $\mathbf{U} \triangleq \{i : \mathcal{D}_i(\mathbf{r}_i) = ?\}$;

(ii) $\hat{\mathbf{e}}_i \triangleq \mathbf{r}_i - \mathcal{D}(\mathbf{r}_i)$; $\quad \eta_i \triangleq \hat{e}_{i,1}$; $\quad \tau_i \triangleq \mathrm{wt}(\hat{\mathbf{e}}_i)$. $\quad \forall i \notin \mathbf{U}$;

(iii) $\mathbf{E}_\alpha \triangleq \{i : i \notin \mathbf{U}, \mathcal{D}_i(\mathbf{r}_i) = \alpha\}$. $\quad \forall \alpha \in \mathbf{F}_q$;

(iv) $g_\alpha \triangleq \begin{cases} 0 & \alpha = 0, \\ |\{i : \delta_i \text{ is even}, \tau_i = \delta_i/2, \eta_i \neq \alpha\}| & \text{otherwise.} \end{cases}$

(v) $N(\alpha) \triangleq \sum_{i\in\mathbf{E}_\alpha} \tau_i \quad + \sum_{i\in\mathbf{U}^c-\mathbf{E}_\alpha} (\delta_i - \tau_i) \quad + \sum_{i\in\mathbf{U}} \left\lceil \frac{\delta_i+1}{2} \right\rceil$
$+ g_\alpha - (M-1)\,\mathrm{wt}(\alpha), \quad \forall \alpha \in \mathbf{F}_q.$

Common-Symbol Decoding Algorithm: *Assume no more than $\lfloor\delta/2\rfloor$ errors have occurred. Then, if there exists a unique $\alpha = \alpha^* \in \mathbf{F}_q$, that minimizes $N(\alpha)$, the error in the common position is given correctly as α^*. Moreover, if there does not exist such a unique α^*, then δ is even, and exactly $\delta/2$ errors have occurred.*

For the case of one-step majority-logic decoding, the J orthogonal parity-checks correspond to simple parity-check codes with $\delta_i = 2, \quad \forall i = 1, 2, \ldots, J$, and the algorithm reduces to the familiar one-step majority-logic decoding.

The common-symbol decoding problem may be applied to decoding cyclic block codes because of the following observation: For many large, powerful codes, it is possible to break up any codeword from the large code into M codewords (from smaller, weaker codes) that share a single symbol in common—just as one-step majority-logic decoding may be viewed as breaking up a codeword into a number of codewords, each from a simple parity-check code, with a single symbol in common.

This suggests the following distributed approach to decoding cyclic codes: (i) Break the received (possibly corrupted) codeword from the powerful code into M (possibly corrupted) codewords from smaller codes; (ii) decode these (smaller) codewords in parallel; (iii) pool the results of the individual decoders (η_i's, τ_i's, and $\mathbf{U}$) to decode the symbol in the common position; (iv) repeat with cyclic shifts to decode all symbols of the code.

Fault-Tolerant Decoding: Suppose, of the M decoders, the ith decoder were to fail (i.e. produces unreliable values for η_i and τ_i). Then if this fact is known, simply ignoring the result of the ith decoder allows us to correct the common symbol provided no more than $\lfloor(\delta-\delta_i)/2\rfloor$ errors have occurred. However, if this fact is not known, then only $\lfloor(\delta-1)/2\rfloor - \delta_i$ errors can be corrected. Further, if η_i is reliable, but τ_i is not (and the decoder is unaware of this) then $\lfloor(\delta-\delta_i-1)/2\rfloor$ errors can be corrected.

Bibliography

[1] Murad, A., and Fuja, T., "A Generalization of Majority Logic Decoding", 1991 International Symposium on Information Theory, June 23–28, 1991, Budapest, Hungary.

[2] Massey, J.L., *Threshold Decoding*, M.I.T. Press, Cambridge, Mass., 1963.

[3] Murad A.H., "Distributed Decoding of Block Codes Through a Generalization of Majority-Logic Decoding", *M.S. thesis*, Univ. of Maryland, College Park, Mary., 1992.

†Supported in part by National Science Foundation grant NCR-8957623; also by the NSF Engineering Research Centers Program, CDR-8803012.

On the Fast Decoding of Binary BCH Codes

W T Penzhorn, Member IEEE

Department of Electrical and Electronic Engineering
University of Pretoria, 0002 PRETORIA, South Africa.

ABSTRACT

Recently, it was shown how to determine the error locator polynomial of a primitive, binary t-error correcting BCH code in a single step [3]. For this purpose it is necessary to transform the set of t syndrome polynomial equations to an equivalent set of polynomial equations, leading to an analytic expression for the error locator polynomial, $\sigma(x)$. These results facilitate decoding beyond the BCH bound, i.e. correcting more than t errors. This requires the resolving of additional syndromes coefficients, which is achieved in a simple and elegant way by means of the expression derived for the syndrome polynomial $\sigma(x)$.

SUMMARY

Primitive, binary BCH codes are attractive because of their relative simplicity and good performance. An (n, k) BCH code has k information bits per codeword of length n, is defined over $GF(2^m)$, with $n = 2^m - 1$, and can correct $t = \lfloor (d-1)/2 \rfloor$ where d is the minimum distance of the BCH code. Decoding of the received codeword requires three steps: (i) Computation of a syndrome vector whose $2t$ components belong to $GF(2^m)$,. (ii) Calculating an error locator polynomial of degree t or less over $GF(2^m)$. (ii) Finding the error locations by solving the roots of the error locator polynomial.

In this paper we are concerned with the third step, which is the time consuming one most and difficult to implement in hardware. Conventionally, this requires the use of the well-known Berlekamp-Massey algorithm, or the Euclidean algorithm [1]. The aim of this paper is to discuss methods for closed solutions to step (ii), which are easily implementable in software and hardware.

Consider a primitive, binary BCH code of length $n = 2^m - 1$ over $GF(2^m)$, with α be a primitive element of the field. The generator polynomial, which defines the code, has roots at $\alpha, \alpha^2, \alpha^3, \cdots, \alpha^{2t}$, enabling the code to correct t errors. The decoder evaluates the received codeword at α^j to determine the j-th syndrome S_j :

$$S_j = r(\alpha^j) = \sum_{k=0}^{n-1} r_k \alpha^{jk} \qquad ;j = 1, 2, \ldots, 2t \quad S_j \in GF(2^m)$$

where r_k is the k-th bit of the received vector $r(x)$. It is fairly easy to show that only t of the $2t$ syndrome components are independent [1]. The error polynomial is given as:

$$e(\alpha^j) = e_{i_1}\alpha^{j i_1} + e_{i_2}\alpha^{j i_2} + \cdots + e_{i_t}\alpha^{j i_t} = S_j \qquad ;i = 1, 2, \ldots, t$$

For binary BCH codes the error magnitudes are $e_i = 1$. For notational convenience, let $x_i = \alpha^{i_j}$, which leads to the following system of algebraic polynomial equations, which we shall refer to as F :

$$\begin{aligned} x_1 + x_2 + \cdots + x_t &= S_1 \\ x_1^3 + x_2^3 + \cdots + x_t^3 &= S_3 \\ &\vdots \\ x_1^{2t-1} + x_2^{2t-1} + \cdots + x_t^{2t-1} &= S_{2t-1} \end{aligned}$$

For the correction of one or more errors, we must solve this system of algebraic equations, to determine the error locator polynomial $\sigma(x)$, whose roots are the required error locations [1].

Following Cooper [3][4] let $x = (x_1, x_2, \ldots, x_t)$ and $K = GF(2^m)$; then $F \subset K[x]$ is the ring of polynomials in t variables. Let I be the ideal generated by F. $I(F)$ is spanned by F and is the ideal of all polynomials which vanish at a set $\{\zeta_1, \ldots, \zeta_t\}$ of points in K. By applying a reduction process it is possible to transform the set F into another set G, which is easier to solve. The resulting set G is a triangularized set of equations, which contains the required error locator polynomial $\sigma(x)$. [3]. It is noteworthy that the derived expression for the error locator polynomial is independent of a particular code or any specific finite field, making the result particularly useful for practical application.
When implemented carefully, Buchberger's algorithm [2] has complexity $O(t)$, else $O(t^2)$, which is still manageable for values of $t \leq 7$. The results for $t = 2$ and 3 are as follows [3]:

$$\begin{aligned} t = 2 &: \sigma(x) = x^2 S_1 + x S_1^2 + S_1^3 + S_3 \\ t = 3 &: \sigma(x) = x^3(S_3 + S_1^3) + x^2(S_1 S_3 + S_1^4) + x(S_5 + S_1^2 S_3) \\ &\quad + S_1 S_5 + S_3^2 + S_1^3 S_5 + S_1^6 \end{aligned}$$

Deriving an analytical expression for the error locator polynomial has several advantages. Apart from the obvious reduction of the computational complexity of the decoding algorithm, the analytical solution also allows us to decode beyond the BCH bound, since it is possible to express $\sigma(x)$ in terms of the unknown syndrome coefficient(s). The expression can then be resolved by applying the approach suggested in [5].

REFERENCES

[1] R E Blahut, **Theory and practice of error control codes.** Addison-Wesley, 1984.

[2] B Buchberger, "An algorithmic method in polynomial ideal theory", in N K Bose : **Multidimensional systems theory (Mathematics and its applications).** Reidel, Boston, pp. 184-232, 1985.

[3] A B Cooper III, "Direct solution of BCH decoding equations", in E Arikan (ed.): **Communication, control and signal processing.** Elsevier, Amsterdam, pp. 281-286, 1990.

[4] A B Cooper III, "Finding BCH error locator polynomials in one step", Electronic Letters, vol.27, no. 22, pp. 2090-2091, 1991.

[5] C R P Hartmann, "Decoding beyond the BCH Bound", IEEE trans. Inform. Theory, pp. 441-444, May 1972.

DIVERSITY SYSTEMS FOR RAYLEIGH FADING CHANNELS: AN APPLICATION OF MULTIPLE DESCRIPTION SOURCE CODES*

Shih-Ming Yang and Vinay Vaishampayan
Electrical Engineering Department
Texas A&M University, College Station, TX 77843

We consider digital transmission of a memoryless Gaussian source over a slow fading Rayleigh channel. Our objective is to demonstrate a useful application of the multiple description scalar quantizer (MDSQ) [1] and to render comparisons against a maximum ratio combiner (MRC)-based system as well as channel code based systems.

MDSQ-based system

The MDSQ is constructed for the following idealized model for a dual diversity system. Assume that two independent channels are available for transmitting information from a continuous alphabet, discrete-time source. Each channel may be in a working or non-working state; this is known in the receiver but not in the transmitter. When working, each channel can support a rate of R bits/source sample. The encoder of an N-level MDSQ maps a source sample to an index pair (i, j). Both indices are mapped to R-bit codewords, where $2^{2R} \geq N$. The codeword corresponding to index i (j) is then sent over the first (second) channel. The quantizer is designed [1] so as to minimize the mean squared-error (MSE) when both channels work, subject to constraints on the MSE when either only the first channel works or only the second channel works. Information theoretic bounds on performance are derived in [2], [3] for a memoryless Gaussian source.

The MDSQ is applied to the Rayleigh fading channel as follows. The bits corresponding to index i are temporally separated from those of index j by an interleaver operating at the R-bit word level so as to obtain two independently faded channels. Individual bits are transmitted over the channel using a BPSK modulator. Soft decision demodulation is used in the receiver and it is assumed that the receiver has perfect knowledge of the Rayleigh fading parameter (channel state information (CSI)). The fading process is assumed to be sufficiently slow so that the Rayleigh parameter remains fixed over R bits. An index is declared to be reliable if the corresponding Rayleigh parameter exceeds a certain threshold. The threshold is optimized to maximize the output SNR[1]. In Fig. 1, we plot the output SNR vs. the channel SNR for a 31-level MDSQ with $R = 4.0$ bits/sample (31-MDSQ).

Reference Systems

The first reference system is a maximum ratio combiner (MRC) based dual diversity system. A source sample is encoded by a Lloyd-Max (LM) quantizer to an R-bit codeword. Each R-bit codeword is duplicated and then temporally separated by an interleaver operating at the R-bit level and transmitted using BPSK modulation. Assuming perfect CSI about the fading parameter is available at the receiver, the output of the two channels are fed to an MRC. The recovered bits are then mapped to Lloyd-Max quantizer reconstruction levels. In order to make an equal bandwidth comparison with the MDSQ system, we consider a 16-level LM quantizer and $R = 4.0$ bits/sample. The performance of the MRC-based system is shown in Fig. 1 (16LM-MRC).

Two channel code-based systems are considered. In both, a source sample is mapped by the encoder of a 32-level LM quantizer to a 5-bit codeword. Two consecutive 5-bit codewords are concatenated, a 0 is appended and fed to an extended (16, 11) Hamming coder. The two systems differ in the interleaver that is used. In the partially interleaved system (PICC), a channel codeword is interleaved at the 8-bit level. In the fully interleaved system, interleaving is performed at the bit level. In both systems BPSK modulation is used, and perfect CSI is assumed in the soft-decision demodulator/decoder. The recovered bits are mapped to LM reconstruction levels. Note that the bandwidth requirements are identical to the MRC- and MDSQ-based systems. Performance results are shown Fig. 1.

Performance Results

First note that the MDSQ-based system significantly outperforms the MRC-based system. Improvements in output SNR at high channel SNR's of over 5 dB are obtained without any sacrifice in the usable channel SNR. Also note that the MDSQ system significantly outperforms the PICC system. The FICC system does better than all other systems. However, the price paid is an excessive increase in the interleaving delay for the FICC system as compared to the MDSQ, MRC or PICC systems. For example consider a mobile radio moving at 30 mph. If we assume a temporal separation of 5ms to obtain independent fades [5], the end-to-end delay for the MRC, MDSQ and PICC systems is roughly 5 ms. However, for the FICC system the corresponding delay is roughly 75 ms since interleaving is performed at the bit level. Thus the MDSQ system improves the performance over an MRC system without increase in the interleaving delay. This is important in two-way speech communication systems where delay must be allocated between the source coder, channel coder and interleaver so as to meet a delay budget.

References

[1]. V. Vaishampayan, "Design of Multiple Description Scalar Quantizer," accepted for publication, *IEEE Trans. Inform. Theory*.

[2]. A. A. El Gamal and T. M. Cover, "Achievable Rates for Multiple Descriptions," *IEEE Trans. Inform. Theory*, Vol. IT-28, pp. 851-857, Nov. 1982.

[3]. L. Ozarow, "On a Source Coding Problem with Two Channels and Three Receivers," *The Bell Syst. Tech. J.*, Vol. 59, pp. 1909-1921, Dec. 1980.

[4]. N. Rydbeck and C. E. Sundberg, "Analysis of Digital Errors in Nonlinear PCM Systems," *IEEE Trans. Commun.*, Vol. COM-24, pp. 59-65, Jan. 1976.

[5]. L. Wong, R. Steel, B. Glance and D. Horn, "Time Diversity with Adaptive Error Detection to Combat Rayleigh Fading in Digital Mobile Radio," *IEEE Trans. Commun.*, Vol. COM-31, pp. 378-386, Mar. 1983.

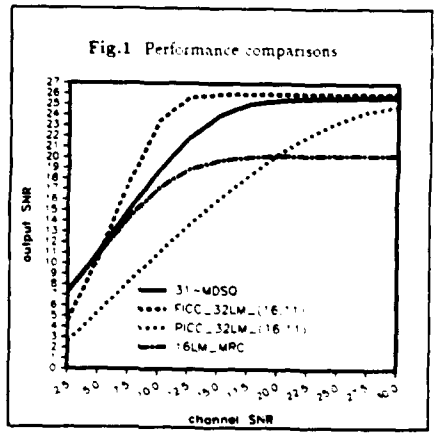


Fig.1 Performance comparisons

* This work was supported by NSF Grant NCR 9104566

[1] The output SNR is given by $10\log_{10}(\sigma^2/D)$, where D is the MSE and σ^2 is the source variance.

ERROR PERFORMANCE OVER THE UNINTERLEAVED CORRELATED RICIAN CHANNEL

Gideon Kaplan and Shlomo Shamai (Shitz)
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

1. General

The correlated Rician channel is a useful model for a slowly-fading channel, in which the complex fading process is composed of two quadrature Gaussian processes with a given normalized autocorrelation function $\rho(\tau)$, and a corresponding symmetrical power spectral density. For slow fading the correlation between adjacent symbols is relatively high and might approach 1. Single-ray model, that is flat fading, is assumed throughout.

We investigate the achievable error probabilities over the channel, employing coherent detection and ideal side information on the realization of the fading processes at the receiver. An underlying decoding delay constraint which precludes the use of (ideal) interleaving is assumed.

In [1], an upper bound on the error probability of block- or convolutionally- coded BPSK over this channel (with similar assumptions on the receiver) was presented. The fading process in [1] is assumed to be piecewise constant (p.c), that is, considered to be constant over a symbol's duration. We analyze the correlated fading channel both with and without the above mentioned p.c approximation. Coded BPSK performance, as well as the exponential behavior of the error probability are discussed. For the continuous channel it is assumed that the receiver has an accurate knowledge of the sample path of the realized fading process. We focus on obtaining the limit of ultimate performance and on verifying under what conditions the above mentioned p.c approximation is adequate. Comparisons to the block-fading model are also discussed.

2. The Piecewise-Constant Approximated Channel

2.1 Coded BPSK performance: A succession of upper bounds on the pairwise error probability ($P(c \to \hat{c})$) is reviewed (based on [1]). Two conjectures of [1] are rigorously proved.

Theorem For a channel with $\rho'(\tau) \geq \rho(\tau)$ the performance is uniformly inferior in comparison to the performance over the channel with $\rho(\tau)$, in terms of the upper bound on the pairwise error probability.

Corollary For the exponentially correlated channel ($\rho(nT) = q^n$), among all codewords of distance d from the transmitted one, the worst upper bound on the pairwise error probability happens when all the erroneous (or different) symbols are consecutive.

2.2 Exponential behavior of the error probability: A general upper bound on the average message error probability ($\overline{P}_e$) for random coding and *i.i.d.*, Gaussian inputs is presented, along with a tighter bound for the exponentially correlated channel where the fact that the fading process is Markovian is invoked [2]. Evaluation of the exponential behavior of $\overline{P}_e$ for finite code lengths is addressed. It is shown that under stringent decoding delay constraints, and a very slow fading process (compared to the user baud rate) reliable transmission of high information rates is hard to achieve. However, under mild technical conditions, when the decoding delay constraint is relaxed and when $\rho(nT) \to 0$ for $n \to \infty$, the classical Shannon capacity (that is the ultimate achievable rate) exists and it is given by the average mutual information.

3. The Continuous Channel Model

The pairwise error probability for coded BPSK depends on the squared Euclidean distance between the two faded codewords; its evaluation is pursued here.

3.1 Ideal interleaving: The Chernoff upper bound on $P(c \to \hat{c})$ is determined in terms of the *Fredholm determinant* [3] (associated with $\rho(\tau)$). The latter is evaluated based on the representation of the fading processes by a Karhunen-Loëve expansion. The Bhattacharyya distance is compared to the p.c.-approximated channel; the "inherent diversity" embedded in the continuous model (with perfect side information) is pointed out.

3.2 No interleaving: Here, the upper bound on $P(c \to \hat{c})$ depends on the Fredholm determinant for the 'filtered' fading process, namely the process multiplied by the window function $\Lambda(t)$ which is $\neq 0$ over the symbols which do not agree in c and $\hat{c}$. When the different symbols are consecutive, the evaluation is straightforward. In the general case, the Fredholm determinant of a complex Gaussian process filtered by a time-varying filter should be evaluated. To solve for the Fredholm determinant, we use a state-space representation of the system, and Collins' modification to the Riccati equation [3], and obtain for an interesting example a closed-form solution.

References

[1] F. Gagnon and D. Haccoun, "An upper bound on coded performance with non independent fading", Ecole Polytechnique de Montreal, Tech. Rep. EPM/RT-90/01. See also *IEEE Trans. on Commun.*, Vol. 40, No. 2, Feb. 1992, pp. 351–360.

[2] A.N. Trofimov, "Convolutional codes for channels with fading", Prob. of Inform. Transmission, Vol. 27, Oct. 1991, pp. 155–165.

[3] C. Helstrom, Detection and Estimation Theory, UCSD Press, Vol. II, ch. XI, 1986, see also C.W. Helstrom, Statistical Theory of Signal Detection, Pergamon Press, Oxford, ch. IV, XI, 1968.

[4] G. Kaplan and S. Shamai (Shitz), "Achievable Performance over the Correlated Rician Channel", EE Publication No. 837, Technion, June 1992, also submitted to *IEEE Trans. On Commun.*,

Error Probability Bounds for M-ary DPSK Signaling over Doubly Selective Fading Diversity Channels *

Daniel L. Noneaker and Michael B. Pursley

Coordinated Science Laboratory
1308 W. Main Street
University of Illinois, Urbana, Illinois 61801 USA

Abstract A method is described for obtaining tight closed-form bounds on the probabililty of error for M-ary differential phase-shift keying and Rician fading diversity channels. The channels exhibit doubly selective fading. In addition, the gains of the diversity channels are correlated. As an example, the results of applying this technique are given for 16-ary DPSK signaling.

I. Summary

Differential phase-shift keying (DPSK) modulation and diversity transmission have long been employed for communications over fading, multipath channels. In recent years, the desire for high bandwidth efficiency and the introduction of multi-phase coding schemes have led to consideration of M-ary DPSK with $M \geq 4$.

Previous results [1] for differential binary PSK signaling over doubly selective fading diversity channels can be extended to differential quadriphase shift keying. Performance of M-ary DPSK is analyzed in [2] for doubly selective fading diversity channels. An iterative expression for the error probability is given for arbitrary symbol set size and an arbitrary order of diversity combining. The results of [2] are applicable only to Rayleigh fading diversity channels that are modeled by independent, identically distributed random processes. In this paper we present a method for determining performance bounds when the diversity channels exhibit correlated Rician fading.

It has been shown [3] that for M-ary DPSK, the bit error probability can be bounded in terms of the probability that the complex-valued decision statistic falls within any of several specified half-planes and quarter-planes. However, for the system and channels under consideration in this paper, the probability that the decision statistic falls within an arbitrary quarter-plane cannot be expressed in closed form. Thus, we use only bounds obtained by considering half-planes.

The receiver employs differentially coherent detection and square-law diversity combining, as in [2]. The diversity channels are modeled as jointly Gaussian wide-sense-stationary uncorrelated scattering fading channels. The probability that the decision statistic falls within a specified half-plane is equal to the probability that a certain Hermitian quadratic form in jointly Gaussian random variables is less than zero. This probability is evaluated by the approach of [1]. Thus, tight bounds on the error probability are obtained for M-ary DPSK and a general class of fading diversity channels.

An application of the results of this paper is illustrated in the figure, where upper and lower bounds on the bit error probability are given for 16-ary DPSK signalling and dual diversity combining. Each diversity channel is a doubly selective fading channel. The delay spectrum of each channel is rectangular with a normalized rms delay spread of 0.1 and the time-correlation function is exponential with a normalized Doppler spread of 0.0001. The signal-to-noise ratio is the ratio of the mean received signal energy to the noise power density.

References

[1] P. A. Bello, "Binary error probabilities over selectively fading channels containing specular components," *IEEE Trans. Commun. Technol.*, vol. COM-14, pp. 400-406, Aug. 1966.

[2] D. L. Noneaker and M. B. Pursley, "M-ary differential phase-shift keying with diversity combining for communications over a time- and frequency-selective fading channel," *Proc. 1992 International Conference on Communications*, pp. 46 50, June 1992.

[3] P. J. Lee, "Computation of the bit error rate of coherent M-ary PSK with Gray code bit mapping," *IEEE Trans. Commun.*, vol. COM-34, pp. 488 491, May 1986.

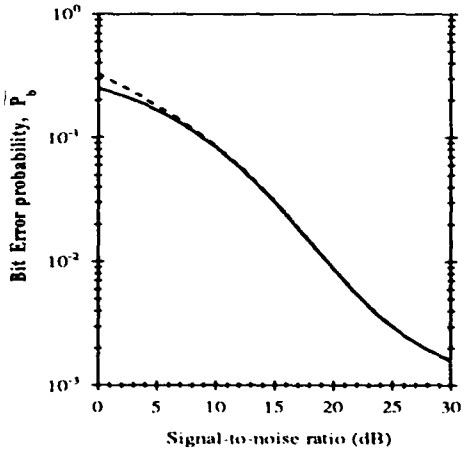


Figure. Performance of 16-ary DPSK and dual diversity combining.

*This research was supported by the Joint Services Electronics Program under grant N00014-90-J-1270.

LIMITING CUTOFF RATE FOR PHASE-ONLY MODULATION ON A SLOW-FADING RICIAN CHANNEL[†]

J.W. Modestino
Electrical, Computer and Systems Engineering Department
Rensselaer Polytechnic Institute
Troy, New York 12180

Abstract

We consider the limiting channel cutoff rate for memoryless phase-only modulation operating on the slow-fading Rician channel for large alphabet size, M. Previous work has considered evaluation of the cutoff rate, R_0, in bits/channel use, for multiple phase-shift keyed (MPSK) modulation with *fixed* M as a function of the channel parameters, as well as system implementation details such as whether or not interleaving/deinterleaving or channel state information (CSI) is used. Here we evaluate the limiting R_0 when the restriction of a fixed alphabet size, M, is removed but the transmitted signal is restricted to utilize memoryless phase-only modulation. Results are provided under the assumptions of *ideal* interleaving/deinterleaving both with *perfect* and *no* CSI. In the case of *no* CSI we show that there exists a maximum useful signaling rate which depends only upon the ratio of specular-to-diffuse energy.

Summary

In previous work [1] we have evaluated the channel cutoff rate for multiple phase-shift keyed modulation (MPSK) operating on the slow-fading Rician channel. Here, the cutoff rate, R_0, in bits/channel use, was evaluated for fixed alphabet size, M, as a function of channel parameters as well as system implementation details. The latter include whether or not interleaving/deinterleaving or channel state information (CSI) is employed. It is of some interest to determine the limiting cutoff rate performance under the same conditions when the restriction of fixed M is removed but the transmitted signal is restricted to utilize memoryless phase-only modulation.

The capacity under a memoryless phase-only constraint is described in [2] for operation on the additive white Gaussian noise (AWGN) channel. No work to the author's knowledge has considered the corresponding cutoff rate, under a phase-only constraint, on the AWGN channel let alone a fading channel. In the present paper we provide the derivation and numerical evaluation of the cutoff rate for phase-only modulation on the slow-fading Rician channel under the assumption of *ideal* interleaving/deinterleaving and for the two extremes of *perfect* and *no* CSI. The channel modeling assumptions include the AWGN as a special case. Indeed, in this case we demonstrate that for large E_b/N_0 the cutoff rate is approximately 1.68dB from the asymptotic capacity determined in [2]. Thus, at least in the AWGN case, conclusions based on capacity arguments are mimicked by the corresponding cutoff rate results. This is heartening since, as first argued by Massey [3],[4], the cutoff rate has come to be accepted as the *practical* upper limit on channel signaling rates for which arbitrarily high reliability can be expected.

The more general results for an arbitrary slow-fading Rician channel are likewise useful in assessing modulation/coding tradeoffs on representative fading channels. For example, for the case of *no* CSI we show that there exists a maximum useful signaling rate which depends only upon the ratio of specular-to-diffuse energy. Thus, regardless of the alphabet size, M, the channel throughput cannot be improved by increasing E_b/N_0 as is the case within *perfect* CSI.

References

[1] J.W. Modestino, K. Park and S.N. Hulyalkar, "Trellis-Coded MPSK Operating on the Slow-Fading Rician Channel," submitted to IEEE Trans. on Inform. Theory.

[2] R.E. Blahut, Principles and Practice of Information Theory, Addison-Wesley, Reading, MA, 1987.

[3] J.L. Massey, "Coding and Modulation in Digital Communications," Proc. 1974 International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.

[4] J.L. Massey, "The How and Why of Channel Coding," Proc. 1984 International Zurich Seminar on Digital Communications, pp. 67-79, Zurich, Switzerland, March 1984.

[†]This work was supported in part by DARPA under Contract No. F30602-92-C-0030.

Bidirectional Decoding of Convolutional Codes over Rayleigh Fading Channels[1]

Jean Belzile, David Haccoun and Serge Forest
Department of Electrical and Computer Engineering
Ecole Polytechnique de Montréal
P.O. Box 6079, station "A"
Montréal, Qc, Canada
H3C 3A7

Abstract

A suboptimal breadth-first multiple-path bidirectional decoding algorithm for convolutional codes has been shown to provide very attractive error performances over the binary symmetric channel. In this paper, new computer simulation results for bidirectional decoding of convolutional codes over soft-decision Rayleigh fading channels are presented. Using a memory length $\nu = 19$ and rate $R = \frac{1}{2}$ code, these results show that a gain near 5 dB can be achieved for a low frame error probability ($P_f < 10^{-3}$) over the Viterbi algorithm of equivalent decoding complexity ($\nu = 6, R = \frac{1}{2}$). The results also indicate that, depending on the length of the frames, a significant gain can also be achieved for low bit error probability ($P_b < 10^{-5}$).

Summary

The Viterbi algorithm is widely used for the decoding of convolutional codes [1-3]. This optimal algorithm [4] exhaustively searches all states of the trellis and delivers the most likely information sequence given the received symbols. The major drawback of this technique is that its complexity increases exponentially with the memory of the code, making its use restricted, for practical reasons, to short memory codes ($\nu \leq 7$). For the decoding of longer memory codes, suboptimal decoding procedure must be considered.

The M-Algorithm [5] and other breadth-first decoding algorithms [6,7], have a constant computational load which is guided by the number of paths explored, M, instead of the number of states in the trellis. A major drawback of these decoding techniques is the lack of resynchronization, leading to long error events when the correct path is lost.

The bidirectional decoding algorithm [8] has been shown to be very effective in reducing the length of the error events caused by correct path lost. This suboptimal algorithm, suited for long memory codes, uses a fixed number of paths, M, all of equal length, in a bidirectional breadth-first tree searching manner. By a judicious sharing of the forward and reverse explorations of the tree, this decoding technique restricts the extend of the error propagation due to correct path lost. Bidirectional decoding does not introduce any computational variability and effectively lowers the number of computations in order to achieve the same bit error rate as a Viterbi decoder of equivalent decoding complexity, that is, same number of path extensions at each decoding step.

In this paper, bidirectional decoding of a $\nu = 19$ and rate $R = \frac{1}{2}$ convolutional code with 64 path extensions at each decoding step ($M = 64$) is compared using computer simulations to Viterbi decoding of a $\nu = 6$ and rate $R = \frac{1}{2}$ code over Rayleigh fading channels. Two fading channels have been considered. First, the urban radio mobile channel with Rayleigh fading and Bessel autocovariance of the received symbols' energy is examined. Results show that a gain of 3.7 dB to 4.5 dB is achieved by the bidirectional algorithm over the Viterbi algorithm on a frame error probability $P_f < 10^{-3}$ for normalized Doppler frequencies ranging from $F_D T = 0.1$ to $F_D T = 0.005$. These last results have been obtained for a frame length of $L = 500$ information bits, but extensive simulation results indicate that the gain provided by bidirectional decoding, for frame error performances, is only slightly dependent on the frame length. However the bit error performance gains obtained by the bidirectional algorithm over the Viterbi algorithm are frame length-dependent since the bit error rate of the bidirectional algorithm is influenced by the length of the data frame. For frames of $L = 500$ information bits, results show that gains ranging from 0.2 dB to 2.3 dB for a bit error probability $P_b < 10^{-5}$ can be obtained with bidirectional decoding over the Viterbi algorithm of an equivalent computational complexity.

Computer simulation results for the Rayleigh fading channel with exponential autocovariance will also be presented. These results, over the same normalized Doppler frequency range ($F_D T = 0.1$ to $F_D T = 0.005$), show that the Viterbi algorithm performs better in this channel than in the previous one, reducing somewhat the advantage of the bidirectional algorithm over the Viterbi algorithm. Nevertheless, substantial coding gains in the frame error performances (2 to 3 dB) can still be achieved and, furthermore depending on the length of the frames, an improvement in the bit error performances (about 1 dB) can also be obtained with the bidirectional algorithm.

References

[1] BHARGAVA, V. K., HACCOUN, D. MATYAS, R. and NUSPL, P. P., *Digital Communications by Satellite*. New York: John Wiley & Sons, 1981.

[2] WU, W. W., HACCOUN, D., PEILE, R., and HIRATA, Y., "Coding for Satellite Communication," *IEEE Journal on Selected Areas in Communication*, vol. SAC-5, pp. 724-748, May 1987.

[3] PROAKIS, J. G., *Digital Communications*, McGraw-Hill, 1989.

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

[5] JELINEK, F. and ANDERSON, J. B., "Instrumental Tree Encoding of Information Sources," *IEEE Transactions on Information Theory*, vol. IT-17, pp. 118-119, Jan 1971.

[6] SIMMONS, S. J., A Reduced-computation Trellis Decoder with Inherent Parallelism, Ph.D. Thesis, Queen's University, Kingston, Ontario, Canada, June 1986.

[7] LIN, C. F., *A Truncated Viterbi Algorithm approach to Trellis Codes*, Ph.D. Thesis, ECSE Dept., Rensselaer Polytechnic Institute, Troy, N.Y., September 1986.

[8] BELZILE, J. and HACCOUN, D., "Bidirectional Breadth-first Algorithms for the Decoding of Convolutional Codes," *IEEE Transactions on Communications*, Feb. 1993.

1. This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada, the Fonds pour la formation de Chercheurs and l'Aide à la Recherche of Québec and the Canadian Institute for Telecommunication Research under the National Centers of Excellence program of the Government of Canada.

A BAYESIAN METHOD FOR DEPENDENT ERASURES IN FREQUENCY-HOP COMMUNICATION SYSTEMS WITH RAYLEIGH FADING

Carl W. Baum
Clemson University - ECE Department
207D Riggs Hall, Box 340915
Clemson, SC 29634-0915, USA

and

Michael B. Pursley
Coordinated Science Laboratory
University of Illinois
1308 W. Main St., Urbana, IL 61801, USA

Abstract

The use of block coding and errors-and-erasures decoding can enhance performance in frequency-hop communication systems, provided that a good scheme is employed to determine which symbols to erase. The problem of making erasure decisions from collections of receiver outputs is investigated in this paper. Methods to determine which received symbols to erase are derived from Bayesian decision theory. The result is a Bayesian scheme in which erasure decisions are made collectively for each codeword. The performance of this scheme is compared with the performance of another Bayesian method in which erasure decisions are made independently from symbol to symbol, and both are compared to the performance of receivers that do not erase. The Bayesian method with dependent erasures is found to provide the best performance.

Summary

The performance of frequency-hop communication systems that are subjected to wideband noise and frequency-selective fading is generally unacceptable without some form of error correction. Erasure of the least reliable symbols prior to decoding can provide significant improvement in performance if the communications receiver has some way to accurately determine the reliability of the received symbols. To accomplish this, the receiver must generate a statistic that is a measure of the likelihood that a symbol is in error, and the decoder must use this statistic effectively to decide whether to erase the symbol. Some approaches require the transmission of additional redundant symbols in order to obtain side information for determining which symbols to erase (see [1] and the references in [2]).

An alternative approach is to base erasure decisions on the envelope detector outputs of a noncoherent receiver, without transmitting additional symbols. In [2], Bayesian decision theory is used to obtain a erasure scheme that minimizes a bound on the probability of not decoding correctly. With this method, one first computes a function of the envelope detector outputs that correspond to a given code symbol (this function is essentially a reliability function). The result is then compared to a threshold to decide whether to erase the corresponding symbol.

There is one significant drawback to this Bayesian technique. Because the erasure decisions are made independently from symbol to symbol, it is possible that more symbols can be erased than the code is capable of correcting. One intuitive solution to this problem is to employ the Bayesian erasure scheme in parallel with errors-only decoding. This has been proposed with other erasure schemes[1, 3]. Unfortunately, our studies show that negligible performance improvements result for the Bayesian scheme.

In this paper, we propose the use of a Bayesian decision rule that makes *dependent* erasure decisions. The form of this rule is obtained by using a decision-theoretic approach to minimize the probability of not decoding correctly under a model that distorts the prior probabilities to make all symbol sequences equally likely. The result is a decision rule that offers significant performance improvements over the Bayesian scheme with independent erasures with only a modest increase in complexity.

The system under consideration is similar to the system described in [4]. Frequency-hop spread-spectrum transmission, noncoherent demodulation, and an $(n, \dot{\kappa})$ extended Reed-Solomon (RS) code are employed. The modulation is M-ary orthogonal signaling with $M = n$, and n is a power of two. The channel includes the effects of Rayleigh fading as well as white Gaussian noise with spectral density $\frac{1}{2}N_0$. The receiver determines a set of symbols to erase, makes hard decisions on the other symbols, and then employs errors-and-erasures bounded-distance decoding.

We let $\mathbf{Y}^j$ denote a vector of envelope detector outputs that correspond to the j-th code symbol in a codeword. The conditional density of $\mathbf{Y}^j$, given that the j-th code symbol sent was s_i, is denoted by $f(\mathbf{y}^j|s_i)$. The receiver we propose can be described as follows:

1. If the jth code symbol is not erased, choose the s_i that maximizes $f(\mathbf{y}^j|s_i)$.
2. If ℓ symbols are to be erased, then they should be the symbols with the ℓ smallest values of $L(\mathbf{y}^j)$, $1 \le j \le n$, where
$$L(\mathbf{y}^j) = \max_k f(\mathbf{y}^j|s_k) / \sum_{i=0}^{n-1} f(\mathbf{y}^j|s_i).$$
3. For each i, $1 \le i \le n$, let L_i be the ith smallest element in the sequence $L(\mathbf{y}^1), L(\mathbf{y}^2), \ldots, L(\mathbf{y}^n)$. Then ℓ symbols are erased if and only if ℓ minimizes
$$\mathrm{P}(X_{\ell+1} + X_{\ell+2} + \cdots + X_n > \lfloor (n-k-\ell)/2 \rfloor),$$
where X_i is a Bernoulli random variable with parameter L_i.

The receiver described above is quite general, and can be applied to a variety of channel and signaling models. For the channel and system described above,

$$L(\mathbf{y}^j) = \frac{\max_k \exp\{(y_k^j)^2/[2\sigma^2(1+\sigma^2/\tau^2)]\}}{\sum_{i=0}^{M-1} \exp\{(y_i^j)^2/[2\sigma^2(1+\sigma^2/\tau^2)]\}},$$

where σ and τ satisfy $\tau^2/\sigma^2 = (\log_2 n)(k/n)(E_b/N_0)$. In this expression, E_b is the average received energy per data bit, and y_k^j is the value of the envelope detector output that corresponds to s_k.

The performance of this erasure decision scheme is measured by the probability of not decoding correctly. For performance comparisons,. we consider the Bayesian method in [2] as well as simple hard-decision demodulation (no erasures). Our simulation results show that, over a wide range of error probabilities and with (32,12) and (32,16) RS codes, the method in [2] provides several dB of performance gain over errors-only decoding, and the dependent erasure scheme provides roughly an additional 0.5 dB of gain.

References

[1] K. G. Castor and W. E. Stark, "Parallel decoding of diversity/Reed-Solomon coded SSFH communications with repetition thresholding," *Proc. of the Conf. on Inform. Sci. and Syst.*, pp. 75–80, March 1986.

[2] C. W. Baum and M. B. Pursley, "Bayesian methods for erasure insertion in frequency-hop communication systems with partial-band interference," *IEEE Trans. Commun.*, vol. 40, pp. 1231–1238, July 1992.

[3] M. B. Pursley and W. E. Stark, "Performance of Reed-Solomon coded frequency-hop spread-spectrum communications in partial-band interference," *IEEE Trans. Commun.*, vol. COM-33, pp. 767–774, Aug. 1985.

[4] M. B. Pursley, "Frequency-hop transmission for satellite packet switching and terrestrial packet radio networks," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 652–667, Sept. 1986.

Acknowledgement. This work was supported in part by the Joint Services Electronics Program under Grant N00014-90-J-1270 and in part by Motorola, Inc. Carl W. Baum is the recipient of a Motorola Partnerships in Research Grant.

Performance Analysis of Frequency-Hopped Digital FM Diversity Systems

Leonard E. Miller and Jhong S. Lee
J. S. Lee Associates, Inc., Rockville, MD

The system studied is a frequency-hopping continuous-phase frequency-shift keying (digital FM) communication system (FH/CPFSK) that utilizes error-control coding, interleaving, and L hops/bit diversity to mitigate the effects of noise and jamming. In the slow-hopping transmission scheme, coded binary data symbols are repeated on L different hops in order to increase the likelihood that some of the symbols are free of partial-band jamming interference. The coded symbols are first read into a Q-symbol buffer, where Q is the number of symbols that can be transmitted in one hop period. After interleaving, L copies of the Q-symbol sequence are transmitted on L successive hops. The reception scheme uses a method for combining diversity transmissions whose performance is to be evaluated.

The conditional bit error performance of a binary FM communications system employing as a decision variable the sum of L demodulator output samples z_i ($i = 1, 2, \ldots, L$) that are independent and identically distributed can be formulated as

$$P_e(L;\beta) = \tfrac{1}{2} - \tfrac{1}{\pi}\int_0^\infty \frac{d\nu}{\nu}\,\mathrm{Im}\left\{\left[C_{z_i}(\nu;\beta)\right]^L\right\}, \tag{1}$$

where $C_{z_i}(\nu;\beta)$ is the characteristic function (CHF) of z_i under parametric conditions denoted by β. The unconditional error probability is found by averaging with respect to the parametric conditions; for example, β can represent the effects of intersymbol interference (ISI), with the averaging being taken over different adjacent-bit patterns. This formulation can be applied to FH digital FM systems in which the decision variable is a weighted sum of diversity transmission samples that are subject individually and independently to jamming (as under partial-band noise jamming of FH signals) by writing

$$C_{z_i}(\nu;\gamma;\beta) = (1-\gamma)C_x(w_0\nu;\rho_N;\beta) + \gamma C_x(w_1\nu;\rho_T;\beta). \tag{2}$$

In (2), γ is the probability that a hop is jammed (fraction of band jammed); (ρ_N, w_0) and (ρ_T, w_1) are, respectively, the combinations of SNRs and of weights multiplying the samples that pertain under non-jamming and jamming conditions; and $C_x(\nu;\rho_x;\beta)$ denotes the characteristic function for a sample when the SNR in the signal intermediate-frequency (IF) bandwidth equals $\rho_x = E_b/LN_x$ (x = N for noise-only, and x = T for noise-plus-jamming), the bit energy-to-noise power spectral density ratio, divided by the number of hops per bit.

For a differential detector, in [1] it is shown that the CHF for a demodulator output sample has the form

$$C_x(\nu;\rho_x;\beta) = \frac{1}{1+4\nu^2c_1c_2+2j\nu(c_2-c_1)} \times \exp\left\{\frac{j\nu(c_1d_1-c_2d_2)-2\nu^2c_1c_2(d_1+d_2)}{1+4\nu^2c_1c_2+2j\nu(c_2-c_1)}\right\} \tag{3a}$$

where in terms of the in-band noise variance σ^2 and its in-phase and cross-phase correlation coefficients r and λ

$$c_{1,2} = \tfrac{1}{4}\sigma^2(\sqrt{1-r^2} \mp \lambda) \tag{3b}$$

and

$$d_{1,2} = \frac{2\rho_x\{U' - rW'\cos\Delta\phi \pm \sqrt{1-r^2}W'\sin\Delta\phi\}}{\sqrt{1-r^2}(\sqrt{1-r^2}\mp\lambda)}. \tag{3c}$$

In (3), the ISI-dependent parameters are $\beta \equiv (U', W', \Delta\phi)$, where U' and W' are the values of the arithmetic and geometric means, respectively, of the SNRs at the beginning and at the end of the bit interval, when the SNR value is $\rho_x = 1$; $\pm\Delta\phi$ are the possible values of the output sample (a differential phase) in the absence of noise.

For demodulation using a limiter-discriminator, the CHF for a demodulator output sample has the form [1]

$$C_x(\nu;\rho_x;\beta) = e^{-\alpha(1-\cos 2\pi\nu)-j\alpha\sin 2\pi\nu}C_\psi(\nu;\rho_x;\beta), \tag{4a}$$

with α denoting the average number of FM "clicks" [2] and $C_\psi(\nu;\rho_x;\beta)$, the CHF for ψ, the value of the modulo-2π differential phase of the signal in noise at the end of the bit interval. An estimate of α is [1]

$$\hat{\alpha} = (\Delta\phi/2\pi)e^{-\rho U'} \tag{4b}$$

The probability density function (PDF) for ψ exists in integral form [2], and a general expression for its CHF does not exist in closed form. However, an excellent approximation for the CHF has been found to be [1]

$$C_\psi(\nu;\rho_x;\beta) \approx \epsilon(\rho)\cdot\frac{\sin\pi\nu}{\pi\nu} + [1-\epsilon(\rho)]\cdot e^{-\sigma_0^2\nu^2/2}, \tag{4c}$$

with σ_0^2 and the mixture parameter ϵ approximated heuristically by

$$\epsilon(\rho) = e^{-2\rho} \text{ and } \sigma_0^2 = \frac{\sigma_\psi^2 - \pi^2\epsilon(\rho)/3}{1-\epsilon(\rho)}. \tag{4d}$$

The formula for σ_0^2 is based on equating the actual phase noise variance, σ_ψ^2, with that of the approximate PDF corresponding to (4c). For high SNR, the differential phase PDF approaches that for a Gaussian distribution, while for low SNR it approaches that for a uniform distribution.

A form of "adaptive gain control" (AGC) combining has been proposed, under which it is assumed that the SNR ρ_x can be measured on each hop and that the L detector output samples are weighted in proportion to the value of ρ_x associated with each sample. Referenced to the absence of jamming, for partial-band jamming the two weights referred to in (2) then can be expressed as

$$w_0 = 1 \text{ and } w_1 = \rho_T/\rho_N. \tag{5}$$

This weighting scheme has the effect of improving the quality of the decision variable by de-emphasizing the samples on jammed hop transmissions, unless all hops were jammed. In the oral presentation, example comparisons will be shown of the uncoded bit error probability (*vs* E_b/N_J) that results from using an adaptive gain control combining scheme for differential detection and for limiter-discriminator detection under the assumption of selected values of L, E_b/N_0, the FM modulation index h, and IF time-bandwidth product $W_{IF}T$; and for worst-case partial-band noise jamming, in which γ is chosen to maximize the error probability. The effective jammer spectral density level is assumed to be N_J/γ, its average over the hopping band divided by γ, so that there is a tradeoff between how much of the band is jammed and the strength of the jamming in the jammed portion of the band. Also, as L increases, the energy per hop is reduced but the chances of having an unjammed hop increases; the tradeoff involved is that the noncoherent combining of the detector output samples cannot recover the total energy effectively, that is, $P_e(L) \geq P_e(1)$ without jamming.

The use of AGC diversity combining results in significant performance improvement. In one example, a 10^{-5} probability of bit error can be achieved for about 24 dB more jammer power when L is increased from 1 to 2. Generally, the amount of "diversity gain" depends on both the type of demodulator and on E_b/N_0. Comparisons of demodulator types based on having their E_b/N_0 values sufficient to produce the same probability of error for $L = 1$ and no jamming reveal that a system using the differential detector with diversity will outperform one using the limiter-discriminator; the reason for this effect is that the limiter-discriminator incurs more noncoherent combining losses than does the differential detector. In the oral presentation of this paper, additional results will be shown, including comparisons of the adaptive gain control scheme with a hard-decision combining scheme in which a majority vote is taken among bit-value decisions made on each hop.

[1] J. S. Lee Associates, Inc., "Studies of ECCM Improvements for *Frequency-Hopping CPFSK Systems*," report to US Army Research Office under contract DAAL03-89-C-0010, May 1990. (DTIC accession number AD-A222 995.)

[2] R. F. Pawula, S. O. Rice, and J. H. Roberts, "Distribution of the Phase Angle Between Two Vectors Perturbed by Gaussian Noise," *IEEE Trans. on Commun.*, vol. COM-30, pp. 1828-1841 (Aug. 1982).

ALGORITHMS FOR PARALLEL DECODING

Wayne E. Stark[1] and Amer A. Hassan[2]
[1] Electrical Engineering and Computer Science Department
The University of Michigan, Ann Arbor, MI 48109 USA
[2]GE Corporate Research and Development Center
Schenectady, NY 12301 USA

Abstract

In this paper we address the parallel decoding problem in a general formulation. The structure of the receiver consists of a bank of z demodulators each followed by an errors and erasures correcting decoders. Each demodulator has a threshold θ that determines an erasure region; we then assign a cost $f(\theta)$ to the interference for causing an erasure and a (larger) cost $f(\bar{\theta})$ for causing an error. The goal in designing the receiver is to choose the thresholds to maximize the interference cost necessary to cause a decoding error. We demonstrate that the above formulation is solvable for many channels of interest.

Problem Statement

Parallel decoding is of considerable interest to improve on the performance of coded communications systems limited by various types of interference. Early work on parallel decoding dates back to Forney in his work on generalized minimum distance decoding [1]. Parallel decoding has been used since then for decoding concatenated codes [2, 3, 4, 5]. In parallel decoding the channel output is processed by z branches; each branch consists of a demodulator connected to a decoder. The i-th demodulator is characterized by a threshold θ_i for deciding whether to erase or to output its best estimate to the decoder; the input to the decoder is then an erasure, a correct estimate, or an erroneous symbol. Then z identical bounded-distance decoders (one for each branch) are used to correct the maximum number of errors and erasures. The receiver, therefore, produces z candidate estimates of the transmitted codeword in which the most likely codeword is selected. The interference distorts the signal and there is a cost associated with each type of distortion. The cost of causing an erasure in branch i is $f(\theta_i)$. The larger cost $f(\bar{\theta}_i)$ is incurred for causing an error to the nearest code symbol. The above communication system can be characterized by the following a game with two players: the communicator and a jammer.

Communicator's Game - Choose the thresholds $\theta_1, ..., \theta_z$ to maximize the minimum cost necessary for a jammer to cause the overall decoding system to err (not decode to the correct codeword).

Jammer's Game - Chose the distortion to minimize the cost needed to force the communicator to cause an error no matter what thresholds are used.

Solution and Examples

The solution to the game above can be proven to satisfy the following set of equations:

$$f(\bar{\theta}_k) + f(\theta_{k-1}) = \alpha \quad k = 1, 2, ..., z+1$$

with the following boundary conditions

$$\begin{aligned} f(\theta_{z+1}) &= f(\bar{\theta}_z), \\ f(\bar{\theta}_{z+1}) &= f(\theta_z) + f(\theta_0) \end{aligned}$$

and with $f(\theta_0) = 0$ if the demodulation is continuous and $f(\theta_0) = 1$ if the demodulation is discrete (e.g. Hamming distance decoding). The above formulation is valid for parallel demodulation or decoding of concatenated codes in which the inner decoder of branch i is characterized by a threshold θ_i. Also, the above formulation is solvable for many channels including the simple M-ary input-output channel with the Hamming distance as the cost function and the additive channel where the cost function corresponds to Euclidean distance. The next two examples illustrates the applicability to noncoherent channel with ratio threshold like decision rules.

Noncoherent Case- Ratio Threshold

Consider the transmission of M-ary code symbols over a continuous additive white Gaussian channel, using orthogonal Frequency Shift Keying (FSK). The received signal is **noncoherently** demodulated with the resulting M matched filters energy outputs $\{Y_0, ..., Y_{M-1}\}$, each corresponding to a transmitted M-ary symbol. In conventional receivers the transmitted symbol is chosen that corresponds to the largest energy value. With no loss of generality, assume that symbol 0 is transmitted and $Y_1 = \max\{Y_1, ..., Y_{M-1}\}$. The decision rule for a decoder with Viterbi ratio threshold characterized by θ is:

$$\begin{aligned} \text{Choose 0 if } \ & \frac{|Y_1|}{|Y_0|} \leq \tan\theta; \\ \text{Erase if } \ & \tan\theta < \frac{|Y_1|}{|Y_0|} < \cot\theta; \\ \text{Error if } \ & \frac{|Y_1|}{|Y_0|} \geq \cot\theta. \end{aligned}$$

It can be shown (see [6]) that for worst case interference the cost function is

$$\begin{aligned} f(\theta_i) &= \sin^2\theta_i \quad \theta_i \in [0, \frac{\pi}{2}] \\ f(\bar{\theta}_i) &= \cos^2\theta_i. \end{aligned}$$

For arbitrary number of branches the optimal θ's and the error correcting capability of the decoding algorithm α are, respectively,

$$\sin^2\theta_k = \frac{k}{2z+1}, \qquad \alpha = \frac{2z}{2z+1}.$$

It can be shown [6] that the error correcting capability,α, for difference thresholding type of decoder is larger than that for ratio thresholding by a factor $\sqrt{2}$. This corresponds to a gain of 1.5 dB in signal-to-noise ratio.

[1] G.D. Forney, *Concatenated Codes*, MIT research monograph No. 37, The MIT press, Cambridge, Mass. 1966.
[2] I.I. Dumer, V.A. Zinovev, and V.V. Zyablov, "Cascaded decoding with respect to minimal generalized distance," *Problems of Control and Information Theory*, Vol. 10, No. 1, 1982, pp. 1-17.
[3] A. A. Hassan and W. E. Stark, "On decoding concatenated codes," *IEEE Trans. Inform. Theory*, vol. 36, no. 3, pp. 677-683, May 1990.
[4] S.I. Kovalev, "Two classes of minimum generalized distance decoding algorithm," *Probl. Peredachi Inf.*, vol. 22, No. 3 1986.
[5] V. V. Zyablov, "Optimization of concatenated decoding algorithms," *Prob. Peredachi Inf.*, vol. 9, no. 1, pp 26-32 1973.
[6] A. A. Hassan and W. E. Stark, "Parallel decoding for channels with jamming," *Proceedings of the IEEE Conference on Military Communications*, October 1992.

Exact Analysis of the Lempel-Ziv Algorithm for I.I.D. Source

Tsutomu Kawabata
Department of Communications and Systems
University of Electro-Communications

Abstract: A new analysis shows that, when we apply the Lempel-Ziv incremental parsing algorithm to i.i.d. source with probabilities:$p_i, i = 1, \ldots, m$, the expected length $E|\mathbf{W}_t|$ of the t-th parsed segment $\mathbf{W}_t$ is given by a simple formula. Following this approach we can show a VF(Variable to Fixed length) version of Ziv-Lempel universal coding theorem.

Let $A = \{a_1, \ldots, a_m\}$ be an alphabet. Denote by $A^* := \cup_{l=0}^{\infty} A^l$ the set of all strings and by $\lambda \in A^*$ a null string. The Lempel-Ziv parsing $LZ : A^\infty \to (A^*)^\infty$ is defined([1]) for an instance $LZ(\mathbf{x}) = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_t, \ldots)$ such that $\mathbf{x} = \mathbf{w}_1\mathbf{w}_2\ldots$ and for each $t \geq 0$ recursively $\mathbf{w}_{t+1}$ is determined to be the shortest string not in the set $T_t^I := \{\lambda, \mathbf{w}_1, \ldots, \mathbf{w}_t\}$. For a given T_t^I, let ∂T_t^I denote the set of possible outcomes of $\mathbf{w}_{t+1}$. (Since T_t^I can be regarded as a set of inner nodes of a tree, ∂T_t^I is a set of corresponding leaves.) Since $|\partial T_{t-1}^I| = (m-1)t + 1$, at most $\lceil \log_2 \prod_{t=1}^{t_0}\{(m-1)t+1\}\rceil$ infomation bits are sufficient to represent $\mathbf{w}_1\mathbf{w}_2\ldots\mathbf{w}_{t_0}$. Now, consider an i.i.d. process taking values on A with probability parameters $\{p_i\}_{i=1}^m$. By the renewal theory we may define a rate $R(t_0)$ of the Variable to Fixed(VF) source code, which consists of all possible concatenations of first t_0 parsing segments, by

$$R(t_0) \quad := \quad \frac{\sum_{t=1}^{t_0} \log_2\{(m-1)t+1\}}{\sum_{t=1}^{t_0} E|\mathbf{W}_t|}, \tag{1}$$

where $|\mathbf{W}_t|$ denotes the length of the string $\mathbf{W}_t$. The term $E|\mathbf{W}_t|$ is given exactly and asymptotically respectively by the following two theorems.

Theorem 1 (Main theorem)

$$E|\mathbf{W}_t| = \sum_{n=1}^{t}(-1)^{n-1}\binom{t}{n}\prod_{l=2}^{n}(1-\sum_{i=1}^{m}p_i^l), \tag{2}$$

where the null product is 1.

Remark: In [2], the sum $S(t) \stackrel{def}{=} \sum_{l=1}^{t} E|\mathbf{W}_l|$ is considered directly and the following recursion is obtained by a different approach from ours:

$$S(t) \quad = \quad t + \sum_{i=1}^{m}\sum_{k=0}^{t}\binom{t}{k}p_i{}^k(1-p_i)^{t-k}S(k-1),$$

with $S(-1) = 0$.

Theorem 2 (VF coding theorem) *For an i.i.d. source with entropy H,*

$$\lim_{t\to\infty}\frac{E|\mathbf{W}_t|}{\log t} = H^{-1}. \tag{3}$$

Remark: This shows sufficiently that $\lim_{t_0\to\infty} R(t_0) = H$.

These theorems are based on the following two lemmas. Let us consider $\mathbf{W}_{t+1}$ instead of $\mathbf{W}_t$, since the former gives more elegant discussion.

Lemma 1

$$E|\mathbf{W}_{t+1}| = \sum_{\mathbf{y}\in A^*} p(\mathbf{y})Pr\{\mathbf{y} \in T_t^I\}, \tag{4}$$

where $p(y_1^k) \stackrel{def}{=} p(y_1\ldots y_k) = \prod_{n=1}^{k} p(y_n)$, and the expectation is taken over all random generations of the parsing tree T_t^I after the t-th parsing.

The key term $Pr\{\mathbf{y} \in T_t^I\}$ in above has an alternative expression given as follows.

Lemma 2 *For any fixed $y_1^\infty \in A^\infty$, let $N_1, N_2, \ldots$ be independent (not identically distributed) geometric random variables, distributed as*

$$Pr\{N_k = n\} = (1-p(y_1^k))^{n-1}p(y_1^k) \text{ for all } n \geq 1. \tag{5}$$

Then

$$Pr\{y_1^k \in T_t^I\} = Pr\{\sum_{l=1}^{k} N_l \leq t\}. \tag{6}$$

To have Theorem 2, we have evaluated the following inequalities for a summand in Lemma 1. These may be interesting by themselves.

Lemma 3 (An upper bound)

$$\sum_{\mathbf{y}\in A^k} p(\mathbf{y})Pr\{\mathbf{y} \in T_t^I\} \quad \leq \quad E[1-(1-p(Y_1^k))^t)]. \tag{7}$$

Lemma 4 (A lower bound) *For any real $J \geq 0$,*

$$\sum_{\mathbf{y}\in A^k} p(\mathbf{y})Pr\{\mathbf{y} \in T_t^I\}$$
$$\geq \quad (1-e^{-J})^k Pr\{kJp^{-1}(Y_1^k) \leq t-k\}. \tag{8}$$

References

[1] J.Ziv and A.Lempel, "Compression of Individual Sequences via Variable-rate Coding," *IEEE Trans. on Information Theory*, vol.IT-24, no.5, pp.530-536, Sept. 1978.

[2] Y.M.Shtarkov and T.J. Tjarkens, "The redundancy of the Ziv-Lempel Algorithm for Memoryless Sources," The Preliminary Manuscript for the *International Workshop for Information Theory*, at Einthoven, Aug. 1990.

[3] T.Kawabata, "Exact Analysis of the Lempel-Ziv Parsing Algorithm for I.I.D. Source," *IEEE Trans. on Information Theory*, to appear.

On Asymptotic Optimality of a Sliding Window Variation of Lempel–Ziv Codes

HIROYOSHI MORITA AND KINGO KOBAYASHI

Department of Computer Science and Information Mathematics,
The University of Electro-Communications, Chofu, Tokyo 182, JAPAN

Abstract

Ziv and Lempel proposed two important universal coding algorithms in 1977 and 1978[1, 2]. While the second algorithm called LZ78 has been sufficiently analyzed in the literature, the first LZ77 has not yet. LZ77 parses input data into a sequence of phrases, each of which is the longest match in a fixed-sized sliding window which consists of the previously encoded M symbols. Each phrase is replaced by a pointer to denote the longest match in the window. Then a window slides to just before the next symbol to be encoded, and so on. In this paper, we modify the algorithm of LZ77 to restrict pointers to starting only at the boundary of a previously parsed phrase in a window. Although the number of parsed phrase should increase more than those in LZ77, the amount of bits needed to encode pointers is considerably reduced since the number of possible positions to be encoded is much smaller. Then we show that for any stationary finite state source, the modified LZ77 code is asymptotically optimal with the convergence rate $O(\log\log M/\log M)$ where M is the size of a sliding window.

Definition of Stationary Finite State Sources

Let $X_1^n = X_1, X_2, \ldots, X_n$ be a finite output sequence of an information source where X_i is a random variable which takes values in the finite set $\mathcal{A}$ with cardinality $|\mathcal{A}| = \alpha (< \infty)$. Also, let $S_0^n = S_0, S_1, S_2, \ldots, S_n$ be a sequence of states of the source where S_i is also a random variable which takes values in the finite set $\mathcal{S}$ with cardinality $|\mathcal{S}| = \beta (< \infty)$ where S_0 is called the initial state. We also use a bold italic letter to represent a sequence or string of symbols such as $\boldsymbol{X}$ or $\boldsymbol{S}$ if its length is given in the context. The structure of the source is described by giving the statistical dependence of X_i on states of the source: A source is said to be *finite-state* if the joint probability of $\boldsymbol{X} = X_1^n$ and $\boldsymbol{S} = S_0^n$ is given by

$$\Pr(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{S} = \boldsymbol{s}) = q(s_0) \prod_{i=1}^{n} p(x_i, s_i | s_{i-1}).$$

for any $\boldsymbol{x} = x_1 x_2 \cdots x_n \in \mathcal{A}^n$ and $\boldsymbol{s} = s_0 s_1 \cdots s_n \in \mathcal{S}^{n+1}$ where $p(a, s|t), a \in \mathcal{A}, s, t \in \mathcal{S}$ are conditional join probability mass functions and $q(\cdot)$ is a probability mass function. Since the state sequence $\boldsymbol{S}$ is not apparent, only the output sequence $\boldsymbol{X}$ is observable. The probability of $\boldsymbol{X}$ is determined by

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \sum_{\boldsymbol{s} \in \mathcal{S}^{n+1}} \Pr(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{S} = \boldsymbol{s}).$$

Both probabilities $\Pr(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{S} = \boldsymbol{s})$ and $\Pr(\boldsymbol{X} = \boldsymbol{x})$ are also denoted by $P(\boldsymbol{x}, \boldsymbol{s})$ and $P(\boldsymbol{x})$, respectively. Moreover, $\mathcal{P}_{\mathcal{S}}$ denote the class of all stationary finite-alphabet source with $\mathcal{A}$ and $\mathcal{S}$ where $|\mathcal{A}| = \alpha$ and $|\mathcal{S}| = \beta$.

A coding scheme considered here is mostly the same as the LZ77 scheme, except making pointers denote only boundaries of previously parsed phrases as follows:

A Parsing Algorithm *PARR*

Repeat the following steps until the input is exhausted:

1. Find the longest match of the current input sequence from the previously parsed sequence within the window of length M. The start position of the match must be that of the head of a previously parsed phrase. And the longest match is allowed to extent beyond the window as long as it matches the input data.
2. The longest matched sequence is parsed into a phrase W.
3. Without finding any matched sequence, the next input symbol is parsed into a phrase of length 1. □

The above algorithm to parse the input sequence will be denoted as *PARR* (*P*arsing *A*lgorithm based on *R*estricted *R*eproducibility).

Coding Methods

The encoder converts each phrase into a binary codeword under one of two separated modes. Which mode is taken depends on the status of the algorithm when parsing a phrase. One mode is corresponding to the case no match is found. The other is to the case the longest match is found. The former case is denoted as a direct mode, and the latter case as an indirect mode. In a direct mode, the encoder sends out one bit 'zero' followed by a symbol which has been parsed into a phrase length of 1 since there is no match for it. On the other hand, in an indirect mode, it sends out a codeword consisting of three parts: one bit 'one', the position p of the longest match W, and the length l of W. To encode a symbol, a matched position, and length, it is sufficient to assign $\lceil \log \alpha \rceil$ bits for symbol α, $\lceil \log M \rceil$ bits for p, and $2\lfloor \log(l+1) \rfloor$ for l.

The decoding process is quite simple. The first step of the decoder reads in a single bit. If this is a zero, the next $\lceil \log \alpha \rceil$ bits will contains a symbol. If the input bit was a one, it reads in a matched position and length instead of a symbol. Then the decoder can reproduce a phrase from the contents of the current window. Repeat the above procedure until the code sequence is exhausted.

Results

Let us suppose that $\boldsymbol{x}$ is parsed into ν phrases, $W_1, W_2, \ldots, W_\nu$ through *PARR* and encoded according to the method described above. Then the total length of a codeword $\rho_M(\boldsymbol{x})$ is given by

$$\rho_M(\boldsymbol{x}) = \nu + \sum_{j \in J_D} \lceil \log \alpha \rceil + \sum_{j \in J_I} \{ \lceil \log m_j \rceil + 2 \lfloor \log(l_j + 1) \rfloor \}$$

where m_j is the number of the previously parsed phrases to be referred when parsing W_j into a phrase, and l_j is the length of W_j. Moreover, J_D is the number of phrases which belong to the direct mode and J_I is that of those which belong to the indirect mode, that is, $\nu = |J_D| + |J_I|$. Then, we have obtained the following main result.

Theorem Let $\{X_i\}_{i=-\infty}^{\infty}$ be a stationary finite-state stochastic process with the state set $\mathcal{S}$. Let $\rho_M(\boldsymbol{x})$ be the codeword length associated with $\boldsymbol{X} = X_1, X_2, \ldots, X_n$ where M is the length of a sliding window associated with *PARR*. Then, for any $M > 0$, if n is sufficiently large, then

$$\frac{1}{n} \rho_M(\boldsymbol{X}) \leq -\frac{1}{n} \log \max_{P \in \mathbf{P}_{\mathcal{S}}} P(\boldsymbol{X}) + O(\log\log M / \log M)$$

where $\mathbf{P}_{\mathcal{S}}$ denotes the class of all stationary finite-state sources with the finite alphabet $\mathcal{A}$ and the finite state $\mathcal{S}$.

References

[1] J. Ziv and A. Lempel. A Universal Algorithm for Sequential Data Compression. *IEEE Trans. Inform. Theory*, 23(3):337 343. May 1977.

[2] J. Ziv and A. Lempel. Compression of Individual Sequences via Variable-Rate Coding. *IEEE Trans. Inform. Theory*, 24(5):530 536. September 1978.

Adaptive Multi-Dictionary Model for Data Compression*

Chia-Lun Yu and Ja-Ling Wu

Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan, R. O. C.

The main purpose of data compression is to represent source data with a compact form by applying coding techniques. It can use fewer amount of data to substitute original large volume of information. Compression techniques can be applied to data storage and data transmission applications. It can save the space needed when storing enormous data and reduce the time used when transmitting data via communication channels. Data compression plays a very important role in modern information systems. By its result, data compression can be classified as two categories: lossless and lossy compression. Lossless compression assures the original data can be exactly recovered without any distortion. We will focus on lossless case in the following discussions. Common lossless techniques include run-length coding, Huffman coding, arithmetic coding, Lempel-Ziv coding, and BSTW coding. Two major fundamental models are probabilistic model and dictionary model.

One obvious redundancy of many data sets is the repeated occurrence of substrings or patterns. Techniques that factorize common substrings are known as dictionary techniques. A dictionary of common substrings could be constructed using dictionary techniques either on the fly or in a separate pass. It may use the same dictionary for all input data sets (static) or construct a different dictionary for each data file (adaptive or semi-adaptive). Lempel-Ziv coding can be classified as one of the adaptive dictionary techniques.

Cache memories are high-speed buffers which are inserted between the processors and main memory to capture those portions of the contents of main memory which are currently in use. Some well-known management policies include: block placement, block identification, block replacement, and write strategy. The idea of fast access in cache can be applied to data compression. If we collect frequently occurring substrings (patterns) in a small cache-like dictionary and encode these patterns with fewer bits, the overall compression performance should be better.

For dictionary techniques, policies that maintain the contents of dictionaries can be adopted from those of cache management. We proposed a new adaptive multi-dictionary model to describe the behavior of compression coding by the management policies of dictionary. Parameters defined in the model include: the number of dictionaries, the sizes of dictionaries, the generate policy to define new words during encoding, the codeword representation mapping that specifies the output bit pattern of each dictionary entry, the flagbit representation mapping that specifies the flag bit pattern to point out the current used dictionary, the placement policy to decide where a dictionary word should be placed, the replacement policy to throw away old entries when dictionary fills, the update policy to control the exchange of words among dictionaries, and the adjustment policy to modify codeword mapping after each coding step.

Under the proposed model, the coding process of dictionary-based coding can be viewed as the construction, insertion, deletion, and modification of dictionary contents. The characteristics of Lempel-Ziv type methods such as LZ77, LZ78, and LZW can be exactly described by the specified management policies. Meanwhile, some other non-dictionary techniques can also be included in our model. By relating the coding procedures with the dictionary management actions, we had successfully interpreted Huffman coding and arithmetic coding as special cases under the proposed model.

The model describes the operational behavior of dictionary-based coding by nine parameters. Compression efficiency is affected greatly by those factors. The features of our proposed model include multiple dictionaries, time-varient codeword mapping mechanism, adaptive vocabulary exchange capability between dictionaries, and the placement, replacement, update policies for dictionary vocabulary.

Possible applications of the proposed coding model are: First, it provides an unified framework to interpret existent techniques. Second, it points out the possible directions to improve current techniques. Third, new coding system can be easily developed by choosing suitable management policies. The influences of different parameters on compression are the future research topics.

*This work was supported by National Science Council, Taipei, Taiwan, Republic of China, under the contract No. NSC-0408-E-002-232.

UNIVERSAL REDUNDANCY RATES DON'T EXIST

Paul C. Shields
Department of Mathematics
University of Toledo
Toledo, Ohio 43606
and
Eötvös Loránd University
Budapest, Hungary
To appear in IEEE Transactions on Information Theory.

The expected per-letter redundancy

$$R_n(C_n), \mu) = \frac{1}{n} \sum_{x_1^n} (L(x_1^n) + \log \mu(x_1^n)) \mu(x_1^n)$$

measures how far the n-block -to-variable length binary prefix code C_n, with length function L_n, is from being optimal for a given source distribution μ. If a sequence of block-to-variable length prefix codes is to be used on a class $\mathcal{S}$ of sources then a standard requirement is *universality*, namely that $R_n(C_n, \mu) \to 0$ for each member μ of the class, as n goes to infinity. The existence of universal codes for the class $\mathcal{E}$ of all ergodic sources with a fixed alphabet is well known; for example, the Ziv-Lempel code as well as other codes are known to be universal for the class of all ergodic sources. A stronger requirement is to ask that $R_n(C_n, \mu)$ go to 0 at some universal rate for each member of the class. Sequences of prefix codes which have the property that the expected redundancy per symbol is $O((\log n)/n)$ have been constructed for various classes of sources, such as the class of memorlyless sources, the class of Markov sources of a given order, and, more recently, the class of finite-state sources with a given number of states. Furthermore, such results can easily be extended to countable unions of such "nice" classes. Thus, for example, there is a sequence such that $R_n(C_n, \mu) = O((\log n)/n)$ for each μ which is Markov of some order, or for each finite-state process μ.

The purpose of this paper is to show that rates of convergence for redundancy are possible only for special classes of sources, that is, there is no universal redundancy rate for any sequence of prefix codes on the class $\mathcal{E}$ of all ergodic sources. The following is a precise statement of this result.

Theorem 1 *For each n let C_n be a prefix n-code and suppose $\lim_n \rho(n) = 0$. There is an ergodic source μ^* and a subsequence n_m such that*

$$lim_{m \to \infty} \frac{R_n(C_{n_m}, \mu^*)}{\rho(n_m)} = \infty, \text{a.s.}$$

The starting point for the construction of a counterexample μ is a simple method for selecting a periodic measure μ^n such that $R_n(C_n), \mu^n) > 1 - o(1)$, a method suggested to the author by Shtarkov. Iteration of the method leads to a sequence of periodic measures μ^{n_m} such that

$$\lim_{m \to \infty} \frac{R_{n_m}(C_{n_m}, \mu^{n_m})}{\rho(n_m)} = \infty,$$

and such that only a few substitutions and deletions are needed in a μ^{n_m} periodic sample path to produce a $\mu^{n_{m+1}}$ periodic sample path. The latter will guarantee the existence of a limit measure μ for which Theorem 1 holds.

Acknowledgements. The author was partially supported by NSF grants DMS8742630 and DMS-9024240.

A NOVEL SOURCE CODING TECHNIQUE WITH HIGH CONVERGENCE SPEED BASED ON THE LZW ALGORITHM

Junichi Kubo, Takaya Yamazato, Iwao Sasase, and Shinsaku Mori

Dept. of Elec. Eng., Keio University
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, 223 Japan

Abstract

The LZW(Lempel-Ziv-Welch) data compression method is the most popular universal coding algorithm and used in several practical systems. The LZW method, however, has following two disadvantages: the compression ratio converges too slowly and the compressibility is poor when the entropy of the information source is very high. In order to alleviate these disadvantages, we propose a novel source coding technique based on the LZW Algorithm and a splay tree. Our proposed method is superior to the LZW method in terms of universality and convergency. Especially, it is very effective to compress the high entropy information source.

Summary

In the LZW algorithm[1], at the beginning of encoding (Case 1) or in the case that the entropy of an information source is very high (Case 2), parsed strings in a string table are not used efficiently and symbols are encoded frequently. In these cases, it is not effective to map a source alphabet $A = \{a_1, a_2, \ldots, a_\alpha\}$ into fixed-length codes which are longer than $\lceil \log_2(\alpha) \rceil$ bits, where α is the number of the source alphabet A. Here, for the ith segment, L_i^{LZW}(bits), the length of the ith codeword is

$$L_i^{LZW} = \lceil \log_2(\delta) \rceil.$$

where δ is the number of entries in the last string table. As a result, this mapping cause following two problems: in Case 1, the compression ratio converges very slowly and in Case 2, the compressibility is poor when the entropy of the information source is very high.

On the other hand, binary trees are excellent in terms of convergency of the compression ratio because the codewords assigned to symbols depend on the probability of their occurrences. Jones proposes a data compression method using a splay tree[2], which is a self-adjusting binary tree. It can adjust itself quickly to a local redundancy of the information source, and is effective to the high entropy information source.

In this study, we propose a novel source coding technique based on the LZW Algorithm in order to alleviate disadvantages on the LZW method. This technique maps the source alphabet into variable-length codes by using the splay tree, and maps parsed strings into the shortest fixed-length codes, which is suitable for the number of entries in the string table. However, in this mapping, the decoder can not recognize which code is sent, variable or fixed. In order to distinguish one from another, we use a flag bit which is added to the codeword. Here, for the ith segment, $L_i^{Propose}(bits)$, the length of the ith codeword is

$$L_i^{Propose} = \begin{cases} \text{variable-length} + 1 & (0 \le i < \alpha) \\ \lceil \log_2\{\delta - (\alpha - 1)\} \rceil + 1 & (\alpha \le i \le D_{max}) \end{cases}$$

where D_{max} is the maximum dictionary size. As a result, the proposed method is expected to be superior to the LZW method in terms of convergency of the compression ratio and the compressibility in encoding the high entropy information source.

Let the number α of symbols be 256. In the proposed method and the LZW one, the dictionary size, which correspond to the maximum number of parsing segment, is restricted up to 4096. Both methods use LRU (Least Recently Used) deletion heuristic. After the symbol is encoded, the splay tree is updated by using semi-splaying, a variant of splaying.

Figs.1, 2 show the compression ratio at the beginning of encoding respectively for C program and image data which are digitized using 256 grey levels. For the low entropy information source such as C program, the proposed method gives a high compression ratio which is almost equal to that of the LZW method because parsed strings can be encoded effectively. For the high entropy information source such as image data, the LZW method are difficult to compress structurally because parsed strings cannot be encoded effectively. The proposed method, however, can further compress because it maps symbols into variable-length code. Furthermore, the proposed method can adapt itself to these files more quickly than the LZW method.

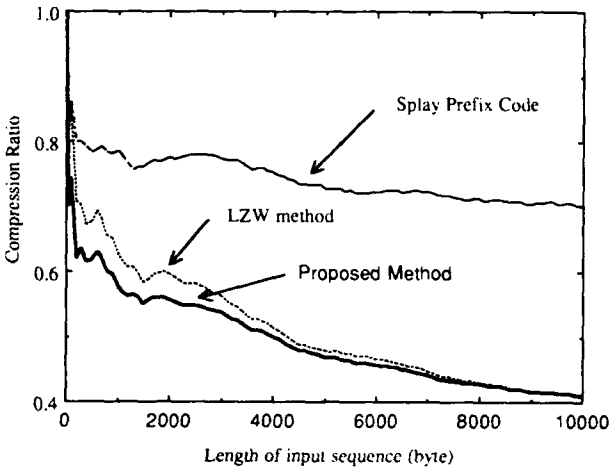


Fig.1 Compression ratio vs length of input sequence (C program)

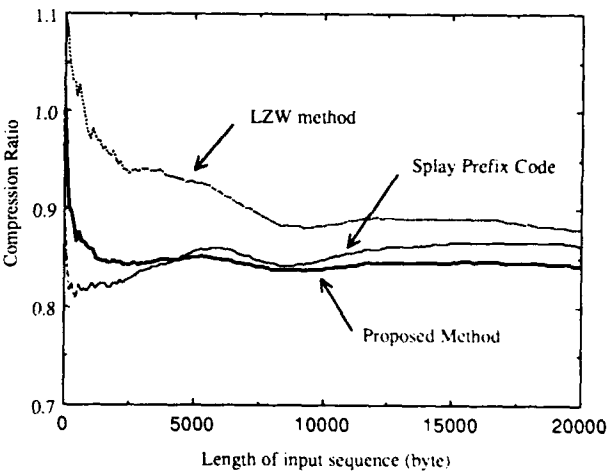


Fig.2 Compression ratio vs length of input sequence (image data)

The proposed method is not only superior to the LZW method in terms of convergency but also compresses the high entropy information source effectively. That is, the proposed method has the higher universality.

References

[1] T. A. Welch, "A Technique for High-Performances Data Compression," *IEEE Computer*, Vol.17, No.6, pp.8-19, June. 1984.

[2] D. W. Jones, "Application of Splay Trees to Data Compression," *Communications of the ACM*, Vol.31, No.8, pp.996-1007, Aug. 1988.

Finite Storage Discriminators for Ergodic Processes

A.D. Wyner
AT&T Bell Laboratories
Murray Hill, NJ 07974

and

J. Ziv
Department of Electrical Engineering
Technion
32000 Haifa
ISRAEL

Abstract

We are looking for an "essential statistic" of a finite-alphabet ergodic source, that, under a given storage (memory) constraint, will allow discrimination between the given source and any other finite-alphabet source.

In our model an encoder is given the n-th order statistics of a stationary process, and the encoder output is a binary N-vector. A discriminator, observes the n-th order statistics of a second source that is either identical to the first source or differs from it by a specified Kullback- Leibler divergence. In a sense made precise in the paper, we show that when n is large, this can be done if and only if $N > \exp(nH)$, where H is the entropy of the first source.

Block Arithmetic Coding for Markov Sources

Charles G. Boncelet Jr.
Department of Electrical Engineering
University of Delaware
Newark, DE 19716

1 Introduction

In a recent paper submitted to IEEE Transactions on Information Theory [1], we introduced BAC. BAC is a variable to fixed block coder in that the input is parsed into variable length substrings which are encoded with fixed length output strings. Assume the input is taken from an alphabet with m symbols and the codebook has K codewords. With each input symbol, the encoder splits the set of codewords into m disjoint, nonempty subsets. The recursion continues until fewer than m codewords remain. One of these is transmitted, and the encoder reinitialized. The encoding process is described in Figure 1.

```
K = K_S;
A = 1;
while ((l = getinput()) ≠ EOF){
        Compute K_1, K_2, ..., K_m;
        A = A + Σ_{j=1}^{l-1} K_j;
        K = K_l;
        if (K < m){
                Output code(A);
                A = 1;
                K = K_S;
        }
}
doeof(A, K);
```

Figure 1: Basic BAC Encoder.

The K_j satisfy the following: 1) $K_j = 1 + L_j * (m-1)$ for $L = 0, 1, 2, \ldots$ and 2) $\sum_{j=1}^{m} K_j = K$. The first condition assures two things: that $K_j > 0$ and that K_j equals the number in a complete and proper set. Let $N(K)$ denote the expected number of codewords encoded with K codewords. For i.i.d. inputs, $N(K)$ satisfies

$$N(K) = 1 + \sum_{j=1}^{m} p_j N(K_j) \tag{1}$$

The principal question is how are the K_j determined. In [1], we offered several methods. Firstly, K_j can be chosen optimally by dynamic programming. Secondly, K_j can be chosen by an arithmetic coding heuristic: $K_j = [p_j K]$, where $[p_j K]$ is a quantization such that conditions 1) and 2) above are satisfied. Thirdly, if we imagine that we can ignore the necessity that K_j be integer, then take $K_j = p_j K$. This solution results in a hypothetical entropy coder.

Denote the expected number number of input symbols encoded with K codewords by $N_o(K)$, $N_h(K)$, and $N_e(K)$, respectively. Then, for i.i.d. inputs, for all K and for some constant C, we showed the following:

$$\frac{\log K}{H(X)} = N_e(K) \geq N_o(K) \geq N_h(K) \geq \frac{\log K}{H(X)} - C \tag{2}$$

For Markov sources, let $p(i|l) = \Pr(x_j = a_i | x_{j-1} = a_l)$. Then the recursion for $N(K)$ splits into two parts. The first is for the first input symbol; the second is for all other input symbols:

$$N(K) = 1 + \sum_{l=1}^{m} p_l N(K_l|l) \tag{3}$$

$$N(K|l) = 1 + \sum_{i=1}^{m} p(i|l) N(K_i|i), \tag{4}$$

$$\tag{5}$$

where $N(K|l)$ is the number of input symbols encoded using K codewords given that the current input is a_l. ($N(K)$ does not include the current input, a_l.) The heuristic is as follows: Choose $K_i = [p(i|l)K]$. The optimal K_j can again be chosen by dynamic programming. We can state the following theorem:

Theorem: If the Markov chain is time-invariant, ergodic, and symmetrical in the following way,

$$H(X|l) = -\sum_{j=1}^{m} p(j|l) \log p(j|l) \tag{6}$$

is independent of l, then, for all l,

$$N_h(K|l) \geq \frac{\log K}{H(X|l)} - C \tag{7}$$

and

$$N(K) \geq \frac{\log K}{H(X|l)} - C' \tag{8}$$

Proof (sketched): The symmetry condition allows that the entropy solution, $K_j = p_j K$, satisfies (4). The proof that $N_h(K|l)$ satisfies (7) follows almost identically from [1]. (8) follows directly from (7).

If the encoder starts anew with each block, then some loss of efficiency occurs since the first symbol of each block is encoded with its stationary probability, nots its Markov one conditioned on the previous symbol. However, encoding blocks separately yields greater resistance to channel errors.

To get a feeling for the magnitude of C, we computed $N(K)$ for two situations. The first is a binary symmetric Markov chain with crossover probability equal to 0.05. For 65536 codewords (16 bits), $N_e = 53.4$, $N_o = 51.0$, and $N_h = 50.5$. The second is a binary asymmetric Markov chain with crossover probabilities equal to 0.05 and 0.50. Again for 16 bit codewords, $N_e = 45.3$, $N_o = 45.2$, and $N_h = 44.4$.

References

[1] C. G. Boncelet Jr. Block arithmetic coding for source compression. *IEEE Trans. on Info. Theory*, 1993. Submitted Sept. 1991.

A Positional Representation for Noiseless Compression

George H. Freeman

Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario, Canada N2L3G1
(519) 885-1211, Ext. 2876, FAX (519) 746-3077
G.Freeman@EandCE.UWaterloo.CA

Abstract—The usual representation of a random sequence on a finite alphabet is obtained by recording the value occurring in each position as the positions are scanned in some standard order. Here, we propose a representation obtained by recording the positions occupied by each value as the values are scanned in some specified order. Entropy is preserved in converting to the positional representation. Also, the unknown positions can be arbitrarily rearranged as the occupied positions are revealed. Under control of a memory model, we propose a rearrangement acting to reduce the first-order entropy. This allows better compression using an adaptive method, such as the Lempel-Ziv algorithm. Memory effects over large sample distances, multiple dimensions, or large alphabets can be directly applied in predicting positions rather than slowly learned by the adaptive coder. Some empirical results for grey-scale television-quality images (480 rows by 512 columns by 256 intensities) are included.

A Summary of the Representation

Suppose the source to be compressed is characterized as the discrete-time, discrete-valued random process $\{X_n;\ n \in \mathcal{N}\}$ where each $X_n \in \mathcal{V}$. For simplicity, we will assume here that the sequences $\mathcal{N} = \{1, \cdots, N\}$ and $\mathcal{V} = \{1, \cdots, V\}$ are finite. This model is sufficient for one video or audio frame which has already been quantized over a bounded interval.

Typically, the source would be processed by considering the values X_n under the prearranged ordering of the $n \in \mathcal{N}$. For a speech frame, we usually process the samples in time order. For an image frame, we usually process the pixels in the order they appear in a raster scan (left-to-right pixels within top-to-bottom rows, say).

Consider the indicator process $\{1_X(n,v);\ n \in \mathcal{N}, v \in \mathcal{V}\}$ derived from the source process by

$$1_X(n,v) = \begin{cases} 1, & \text{if } X_n = v \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\{1_X(n,v)\}$ is constrained to be unit-valued for exactly one $v \in \mathcal{V}$ at each $n \in \mathcal{N}$. Clearly, the mappings between the source representations $\{X_n\}$ and $\{1_X(n,v)\}$ are unique, so the entropy

$$H(1_X) = H(X)$$

is preserved in converting between them.

Let $\mathcal{M} = \{m_i = (n_i, v_i);\ i = 1, \cdots, NV\}$ be a sequence comprising an arbitrary permutation of the elements in the product space $\mathcal{N} \times \mathcal{V}$. Define the indicator process $\{\Phi(m);\ m \in \mathcal{M}\}$ by

$$\Phi(m_i) = 1_X(n_i, v_i) \text{ for } i = 1, \cdots, NV.$$

This is just an arbitrary one-dimensional ordering imposed on the elements of $\mathcal{N} \times \mathcal{V}$ so the entropy

$$H(\Phi) = H(1_X) = H(X)$$

is still preserved.

The process $\{\Phi\}$ has a strong memory effect due to the constraint on the process $\{1_X\}$. Suppose $m_i = (n_i, v_i)$ has $\Phi(m_i) = 1$. For some $j > i$, if $m_j = (n_j, v_j)$ has $n_j = n_i$, we observe that $\Phi(m_j) = 0$ must obtain. That is, once a value is determined for a particular position $n \in \mathcal{N}$, no other value can be specified for that same position. Define the modified indicator process $\{\Psi(m);\ m \in \mathcal{M}\}$ by

$$\Psi(m_j) = \begin{cases} 1, & \text{if } \Phi(m_j) = 1 \\ 0, & \text{if } \Phi(m_j) = 0 \text{ and } \Phi(m_i) = 0 \\ & \text{for all } i < j \text{ having } n_i = n_j \\ \lambda, & \text{if } \Phi(m_j) = 0 \text{ and } \Phi(m_i) = 1 \\ & \text{for some } i < j \text{ having } n_i = n_j. \end{cases}$$

Here, λ is the null symbol. Since knowledge of $\mathcal{M}$ allows the λ's to be inserted, they can be left out of the representation of $\{\Psi\}$. Again, the entropy

$$H(\Psi) = H(\Phi) = H(X)$$

is preserved, for any choice of the ordering $\mathcal{M}$ on $\mathcal{N} \times \mathcal{V}$.

We conjecture an advantage in doing lossless compression using the process $\{\Psi\}$ for three reasons. First, $\{\Psi\}$ is binary, regardless of the size of the source alphabet $\mathcal{V}$. Second, $\{\Psi\}$ is one-dimensional, regardless of the natural or usual dimensionality of the space $\mathcal{N}$ indexing the original source process $\{X\}$. Third, the freedom in choosing $\mathcal{M}$ means that the compressibility can be maximized by using memory modelling, combined with backward adaptation or side information, to determine the best permutation of $\mathcal{N} \times \mathcal{V}$.

Lossless compression methods, such as the Lempel-Ziv algorithm, work best on small alphabet sources having a low first-order entropy. We split $\mathcal{M}$ into two contiguous segments, $\mathcal{M}_0$ having the probability of obtaining $\Psi(m) = 0$ maximised, and $\mathcal{M}_1$ having the probability of obtaining $\Psi(m) = 1$ maximised. These are compressed separately. The memory model is applied to yield a good choice for $\mathcal{M}$, and the split into $\mathcal{M}_0$ and $\mathcal{M}_1$, in the sense of minimising the first-order entropies of the segments.

This work has been supported by the Natural Sciences and Engineering Research Council of Canada under Research Grant A6658 and by the Information Technology Research Centre, an Ontario Centre of Excellence.

LIKELIHOOD METHODS IN IMAGING

Donald L. Snyder
Electronic Systems and Signals Research Laboratory
and
Institute for Biomedical Computing
Washington University
St. Louis, Missouri 63130

Summary

The use of likelihood methods to treat image data has grown significantly over the past twenty years. Three forces continue to drive this evolution. The first is the rapid and continued development of instrumentation used to acquire image data, with tomographic instrumentation in nuclear medicine and radiology being an important early example. A second important development was the identification by L. Shepp and Y. Vardi in 1982 of numerical procedures making likelihood methods feasible. Lastly, the increasing power of digital computation has permitted more and more complicated likelihood methods to be used.

As in other application areas, the power of likelihood methods for imaging relies on having an accurate statistical model describing the available data and how these data are influenced by the underlying objects to be imaged. Poisson and Gaussian processes in time and space often appear in models that account for most of the major sources of noise and distortion in a wide variety of imaging modalities. Side information placing constraints on the object to be imaged can also be of major importance. U. Grenander's theory of object shapes, the introduction into imaging by U. Grenander and M. Miller of jump-diffusion processes, and the use of Markov random fields to accommodate shape constraints are all important developments strengthening the use of likelihood methods for imaging. Regularization is also significant because imaging problems are often ill posed leading to unstable solutions with unconstrained maximum likelihood. Penalty constraints, including object-model constraints, have been found useful as a form of regularization as has U. Grenander's method of sieves.

My objective in this talk is to review likelihood methods that are being used for imaging. The original motivations from single-slice tomographic imaging in nuclear medicine will be mentioned, but emphasis will be placed on more recent developments, applications, and trends.

ON THE PRINCIPAL STATE METHOD FOR RUNLENGTH LIMITED SEQUENCES.

Tjalling Tjalkens
Eindhoven University of Technology,
P.O.Box 513, 5600 MB Eindhoven,
The Netherlands.

Abstract.

We present a detailed result on Franaszek's principal state method for the generation of runlength constrained codes. We show that, whenever the constraints k and d satisfy $k \geq 2d > 1$, the set of "principal states" is $s_0, s_1, \cdots, s_{k-1}$. Thus there is no need for Franaszek's search algorithm anymore. The counting technique used to obtain this result also shows that "state independent decoding" can be achieved using not more than three codewords per message and it allows us to compare the principal state method with other practical schemes originating from the work of Tang and Bahl and also allows us to use an efficient enumerative coding implementation of the encoder and decoder.

Introduction.

Shannon [4] considers the (d,k)-constrained channel, where the only possible binary sequences that can be transmitted over the channel are those containing runs of zeros of length $d, \ldots, k$. $(d < k)$. These channels can be described by a state model where each state is indexed by the length of the current run of zeros. Shannon defines the capacity of this channel as the limit as $n \to \infty$ of the logarithm of the size of the set of all sequences of length n satisfying the (d,k)-constraint divided by n.

A runlength constrained code, (d,k)-constrained code, is a binary encoding of information such that in the code sequence successive ones are separated by at least d zeros and at most k zeros and thus is well suited for use on a (d,k)-constrained channel.

We shall consider fixed length codes for these purposes. Valid codewords follow a possible path in this state model, starting at the state where the previous codeword ended. So, a code for this state model contains several codewords sets, each containing a variable number of words, where the selected set depends on the previous codeword and is such that the concatenation of that codeword with any word in the set is permissible. Since we consider fixed length codes the size of the code is determined by the smallest set belonging to some state in the model.

Franaszek [3] noted that if we take a subset of all states in the model and require the codewords to start and end in states of this subset then an optimum subset exists. This subset is known as the set of principal states and Franaszek described an algorithm to search for these principal states.

Another approach, presented by several authors, [1, 5, 6], is to use a single set S of codewords that satisfy the (d,k)-constraint internally. A special sequence is put in between two codewords such that the (d,k)-constraint remains satisfied between codewords.

The principal state method is an optimal code for systems that can be described in the state model framework, and thus it is at least as efficient as any of the glue methods, since the glue methods can also be described in the state model framework.

The principal states.

Our goal

We start with the definition of the building blocks or basic sets $U(m)$ for the codeword sets given the (d,k)-constraint, containing all sequences that start and end with a "one" and satisfie the (d,k)-constraint internally. Let $U(m)$ denote the size of $U(m)$.

In the following we shall repeatedly use the shorthand notation $[a;b] \triangleq \{a, a+1, \ldots, b\}$.

Let $S \subset [0;k]$ denote the set of permitted channel states, (not necessarily the set of principal states). Consider the sets $V_S(n;i)$ containing all sequences starting with a run of i zeros and ending in a run of $r \in S$ zeros and satisfying the (d,k)-constraint internally. Note that $V_S(n;i)$ can be described using the basic sets as

$$V_S(n;i) = \bigcup_{j \in S} \{0^i\} * U(n-i-j) * \{0^j\},$$

where $U * V$ indicates the set containing all concatenations of the sequences $\underline{x} \in U$ with any sequence $\underline{y} \in V$.

With these sets we can make Franaszek's state depending codeword sets $W_S(n;i)$, i.e. the set of possible codewords of length n starting in state $i \in S$ and ending in any state $j \in S$. We have

$$W_S(n;i) = \bigcup_{\max\{0,d-i\} \leq j \leq \min\{n,k-i\}} V_S(n;j).$$

Now we can formulate our goal and the result:

> Given n, k, and d, (with the restriction $n \geq k \geq 2d > 0$), find the set $S^* \subset [0;k]$ such that:
>
> $$W_{S^*}(n) \triangleq \max_{S \subset [0;k]} \min_{i \in S} |W_S(n;i)| = \sum_{i=0}^{k-1} \sum_{j=0}^{k-d} U(n-d-i-j).$$

Message mapping for state independent decoding

Partition the W_{S^*} messages into sets M_i of sizes $\mathcal{M}_i \triangleq |V_{S^*}(n;d+i)|$, where $i = 0, 1, \cdots, k-d$. Let r be the number of trailing zeros in the previous codeword. We distinguish between the following cases:

- $d = 1$ **and** $r = 0$: The messages in the set M_i are assigned to the set $V_{S^*}(n;i+1)$.
- $d = 1$ **and** $r \geq 1$: The set M_0 is assigned to $V_{S^*}(n;1)$ and $M_1 \cup \ldots \cup M_{k-1}$ are assigned to $V_{S^*}(n;0)$.
- $d > 1$ **and** $r < d$: For all $i = 0, \ldots, k-d-r$ we assign to the set M_i the codewords from $V_{S^*}(n;d+i)$ respectively. For $i = k-d-r+1, \ldots, k-d$ we assign to the set M_i the codewords from $V_{S^*}(n;i+2d-k+1)$ respectively.
- $d > 1$ **and** $r \geq d$: The sets $M_0 \cup M_1 \cup \ldots \cup M_{k-d}$ are assigned to $V_{S^*}(n;0)$ and $V_{S^*}(n;1)$ in that order.

So, it is easy to see that every message is encoded into one of two or three different codewords, depending on r.

Enumerative coding.

We shall briefly indicate the application of the well-known enumerative coding technique [2] to the generation of the (d,k)-constrained sequences.

First we determine the message subset M_i of the message m that we want to transmit. Then, with the rules of the previous section we determine the set $V_{S^*}(n;j)$ and the relative index $i(\underline{x}^n; V_{S^*}(n;j))$ of our message in the set.

Finally we use the enumerative reconstruction to produce the codeword $\underline{x}^n \in V_{S^*}(n;j)$ from its index.

Let the codeword $\underline{x}^n$ be given as $\underline{x}^n = 0^{\alpha_0}10^{\alpha_1}1 \ldots 10^{\alpha_\rho}$. So $\alpha_0 = j$.

Although we will not need the (source) encoding algorithm, it is instructive to see how the index can be computed recursively as

$$\begin{aligned} i(\underline{x}^n; V_{S^*}(n;j)) &= i(\underline{x}^n_{j+1}; V_{S^*}(n-j;0)) = \\ &= i(\underline{x}^n_{\alpha(\underline{x}^n)+2}; V_{S^*}(n-\alpha(\underline{x}^n)-1;0)) + \sum_{l=d}^{\alpha(\underline{x}^n)-1} |V_{S^*}(n-l-1;0)|, \end{aligned}$$

where $\alpha(\underline{x}^n) = \alpha_j$ as given above.

Note that this computation produces a lexicographical ordering given the symbol ordering "1 < 0". Also note that in order to compute the index we only need the $n+1$ numbers $|V_{S^*}(p;0)|$ for $0 \leq p \leq n$.

Reconstructing $\underline{x}^n$ involves producing the $\alpha_0, \ldots, \alpha_\rho$ and they can be found recursively by the corresponding enumerative decoding algorithm.

References.

[1] C.F.M. Beenker and K.A. Schouhamer Immink, "A generalized method for encoding and decoding run-length-limited binary sequences," *IEEE Trans. Inform. Theory*, vol IT-29, pp. 751–754, Sept. 1983.

[2] T.M. Cover, "Enumerative source encoding," *IEEE Trans. Inform. Theory*, vol IT-19, pp. 73–77, Jan. 1973.

[3] P.A. Franaszek, "Sequence-state coding for digital transmission," B.S.T.J., vol 47, pp 143–157, Jan. 1968.

[4] C.E. Shannon, "A mathematical theory of communication," B.S.T.J. vol 27, pp. 379–423, July 1948.

[5] D.T. Tang and L.R. Bahl, "Block codes for a class of constrained noiseless channels," *Information and Control*, vol 17, pp. 436–461, 1970.

[6] J.H. Weber and K.A.S. Abdel-Ghaffar, "Methods for cascading runlength-limited sequences," *Proc. Twelfth Symposium on Information Theory in the Benelux*, Veldhoven, May 23 & 24, 1991.

Joint Runlength/Error-Control Codes Based on Set-Concatenatable Collections

Jian Gu and Tom Fuja
Department of Electrical Engineering
Systems Research Center
University of Maryland
College Park, MD 20742

Recent work by the authors [1] described a new approach for constructing fixed-length (d,k) codes; these codes are "block-decodable" – i.e., they can be decoded with no memory and no anticipation [2] – and they do not require look-ahead at the encoder. Furthermore, it is shown in [1] that the approach described therein is optimal over *all* block-decodable codes with no look-ahead. The new codes do not rely on a search, and they have a very simple structure. In this talk we shall discuss how the new approach can be combined with an error-control structure to yield combined modulation and error control coding.

The approach in [1] is based on *set-concatenatability*. Let $\mathcal{C}^n$ denote the set of binary n-tuples satisfying the (d,k) constraint. Then a set $\mathcal{S} = \{S_0, S_1, \ldots S_{M-1}\}$ of disjoint subsets of $\mathcal{C}^n$ is called a *set-concatenatable collection* (SCC) if for any $S_i, S_j \in \mathcal{S}$ and any $x \in S_i$ there exists a $y \in S_j$ such that $x * y \in \mathcal{C}^{2n}$. The block codes in [1] are based on the maximal set-concatenatable collection (MSCC), and they can encode up to M messages where M is the size of the MSCC. (In [1] it is shown that M is equal to the number of (d,k) sequences of length n with at least d leading zeroes and at most $k-1$ trailing zeroes.)

In this talk we demonstrate how the simple structure of the codes in [1] is easily incorporated into a joint runlength/error-control scheme. In this summary we shall show how the approach of Lee and Wolf [3] may be easily adapted to the new technique.

Let $\mathcal{S} = \{S_0, \ldots, S_{M-1}\}$ be a MSCC with blocklength n_1 for the (d,k)-constrained channel; assume that $\mathcal{S}$ is constructed according to the procedure described in [1]. Let $\mathcal{S}_p = \{0^*, 1^*\}$ be another set-concatenatable collection – one of size two with the smallest possible blocklength n_2; assume once again that $\mathcal{S}_p$ is constructed using the approach of [1]. It can be shown that

$$n_2 = \begin{cases} d+1, & \text{if } k > d+1; \\ d+2, & \text{if } k = d+1. \end{cases}$$

Definition: Given $\mathbf{x} = (x_1, x_2, \cdots, x_{n_1}) \in \mathcal{C}^{n_1}$ and $\mathbf{c} \in \mathcal{S}_p$, define a generalized parity check $h(\cdot,\cdot)$ as follows:

$$h(\mathbf{x},\mathbf{c}) = \begin{cases} 0, & \text{if } \sum_{i=1}^{n_1} x_i \text{ is even and } \mathbf{c} \in 0^* \\ 0, & \text{if } \sum_{i=1}^{n_1} x_i \text{ is odd and } \mathbf{c} \in 1^* \\ 1, & \text{if } \sum_{i=1}^{n_1} x_i \text{ is even and } \mathbf{c} \in 1^* \\ 1, & \text{if } \sum_{i=1}^{n_1} x_i \text{ is odd and } \mathbf{c} \in 0^* \end{cases}$$

Given $\mathcal{S}$ and $\mathcal{S}_p$, we "glue" them together in two different ways to obtain two new collections $\mathcal{S}^e = \{S_0^e, S_1^e, \ldots, S_{M-1}^e\}$ and $\mathcal{S}^o = \{S_0^o, S_1^o, \ldots, S_{M-1}^o\}$:

$$S_i^e = \{\mathbf{x} * \mathbf{c} : \mathbf{x} \in S_i, \mathbf{c} \in (0^* \cup 1^*), \mathbf{x} * \mathbf{c} \in \mathcal{C}^{n_1+n_2}, h(\mathbf{x},\mathbf{c}) = 0\}$$

and

$$S_i^o = \{\mathbf{x} * \mathbf{c} : \mathbf{x} \in S_i, \mathbf{c} \in (0^* \cup 1^*), \mathbf{x} * \mathbf{c} \in \mathcal{C}^{n_1+n_2}, h(\mathbf{x},\mathbf{c}) = 1\}.$$

Claim: $\mathcal{S}^e$ and $\mathcal{S}^o$ are disjoint collections such that $|\mathcal{S}^e| = |\mathcal{S}^o| = M$. Furthermore, $\mathcal{S}^e \cup \mathcal{S}^o$ is a set-concatenatable collection; finally, any two codewords from different elements of $\mathcal{S}^e$ (resp., $\mathcal{S}^o$) lie at distance at least two from one another.

If $M \geq 2^m$, we can construct a code with $d_{free} = 3$ that can encode 2^m messages by numbering the elements of S^e (resp., S^o) with even (resp., odd) integers; the rate of the resulting code will be $m/(n_1+n_2)$. This is done by constructing a completely connected trellis with 2^m states numbered by $\{0, 1, \cdots, 2^m - 1\}$ such that the edge from state x to state y is associated with the "codeword" – the set of (n_1+n_2)-tuples, actually – numbered with the integer $L(x,y) = x - 2y \pmod{2^{m+1}}$. This encoding rule guarantees a non-catastrophic encoder such that the outgoing edges from any state have the same "parity" – i.e., are labeled with either the elements of $\mathcal{S}^e$ or $\mathcal{S}^o$ but not both. This in turn guarantees a free distance of at least three.

Example: We shall construct a code with $(d,k) = (2,4), n = 9, R = 2/9$. Start with a $(2,4)$ code with blocklength $n_1 = 6$:

$$S_0 = \{000010, 010001\} \quad S_1 = \{000100, 010010\}$$

$$S_2 = \{001000, 100001\} \quad S_3 = \{001001, 100010\}.$$

We now use as the "parity check" $\mathcal{S}_p = \{0^*, 1^*\}$ where $0^* = \{000, 010\}$ and $1^* = \{001, 100\}$. This yields the collection

$$S_0' = \{000010001, 010001000\} \quad S_1' = \{000010010, 010001001\}$$

$$S_2' = \{000100100, 010010010\} \quad S_3' = \{000100010, 010010001\}$$

$$S_4' = \{001000100, 100001000\} \quad S_5' = \{001000010, 100001001\}$$

$$S_6' = \{001001000, 100010010\} \quad S_7' = \{001001001, 100010001\}$$

Using the completely connected trellis with four states labeled according to the rule above, we can encode data at a rate $R' = 2/9$ with free distance three. By comparison, if we use the approach in [3] – which employs codewords that can be freely concatenated regardless of encoder state – the best rate that could be achieved with blocklength $n = 9$ would be $R = 1/9$. To obtain a rate close to $2/9$ using the method of [3] would require a blocklength of $n = 14$, with the resulting increase in complexity. ■

Moreover, when $k \geq 2d+1$, we can construct quasi-systematic codes whose codewords are made up of d merging bits, $n_1 - d$ information bits, and $d+1$ checking bits – provided we revise $h(\cdot,\cdot)$ suitably.

A class of single bit-shift error detecting and/or correcting codes can also be constructed by defining an appropriate parity function. Furthermore, these codes are able to deal with bit-shift errors crossing the border of two adjacent codewords.

References

1. J. Gu and T. Fuja, "A New Approach to Constructing Optimal Block Codes for the Runlength-Limited Channel," submitted to *IEEE Transactions on Information Theory.*

2. K.Immink, "Block-Decodable Runlength-Limited Codes via Look-Ahead Technique," *Philips Journal of Research,* June 1992.

3. P. Lee and J. Wolf, "A General Error-Correcting Code Construction for Run-Length Limited Binary Channels," *IEEE Transactions on Information Theory,* November 1989.

Systematic Runlength-Limited Codes For Single Error Detection In the Magnetic Recording Channel.

by Patrick Perry
January 1, 1992
College of Micronesia

The runlength-limited codes are used in magnetic recording. A runlength-limited code is characterized by its (d,k) constraints. The d constraint being the minimum run of consecutive zeros and the k constraint being the maximum run of consecutive zeros.

Methods for mapping unconstrained binary sequences to (d,k) constrained sequences exist. Such mappings are called modulation codes. In this article, we assume the existence of a modulation code and are concerned with detecting errors which occur in magnetic recording.

Errors which occur in magnetic recording can be categorized as drop in errors, drop out errors, or shift errors. A drop in error occurs when a zero is changed to a one. A drop out error occurs when a one is changed to a zero. A shift error occurs when the pattern 01 is changed to 10 or the pattern 10 is changed to 01.

To detect single errors for an unconstrained binary symmetric channel, a single parity bit is adjoined to each information sequence. With a runlength-limited code for the magnetic recording channel, the detection of single errors is not so trivial. The parity must be chosen to maintain the runlength constraints and to detect all three types of channel errors.

The purpose of this article is to present a construction of systematic runlength-limited block codes for detecting single errors in the magnetic recording channel, whether a drop in, drop out, or shift error. The codes can be designed for any (d,k) constraints. The encoder table has 3(k+1) entries. Error detection is performed by a simple arithmetic calculation. Optimal systematic single error-detecting codes are obtained, for the (1,k) and (2,k) constraints with $k > 2(d+1)$, by truncating the constructed codes.

Reduced Complexity Encoding and Decoding Algorithms for a Class of Runlength Limited Error Control Codes

A. Popplewell and J.J. O'Reilly

School of Electronic Engineering Science
University of Wales, Bangor, UK

Summary

Recently a new class of maximum runlength limited error control codes (RLLECCs) have been identified [1-3]. They are formed by taking an appropriate coset of a linear transparent error control code and thereby inherit the error control characteristics and implementation advantages of the parent linear code. A new class of parent codes which realised optimum RLLECCs with minimum distance 4 was recently identified in [2]. These codes were defined in terms of their parity check matrices and although linear they are not cyclic. Consequently practical realisation of these schemes necessitates a network of EXOR gates for encoding, whilst more obviously for decoding storage of lists of syndromes and corresponding error patterns. However the parity check matrices of these optimum codes possess some cyclic-like properties and in this paper we exploit these features to develop simplified encoding and decoding algorithms which readily lend themselves to implementation using VLSI microcircuits and require no storage of syndromes and error patterns. These circuits can be realised simply with EXOR-gates, AND-gates, a 4-input majority gate and shift registers. Furthermore, we find that the circuits are general for a particular runlength constraint the only difference for higher rate codes being the number of shifts required to perform the encoding and decoding operations.

Encoding and decoding algorithms for three particular parent ECCs which when modified appropriately yield runlength constraints 2, 6 and 14 will be considered, although similar algorithms could also be developed for other cases. Circuits which perform the coding operations will also be presented and an error performance comparison of the new algorithms with conventional syndrome decoding will be carried out. By way of example figure 1 shows a comparison between the performance of the new algorithms and syndrome decoding for a (64,46) code with a maximum runlength constraint of 6. Whilst clearly syndrome decoding performs the best there is no significant degradation in performance for the less complex new algorithms A and B. These new algorithms are also readily adaptable to perform soft decision decoding and the potential coding gains using soft decision versions of the decoders will be considered.

References

[1] POPPLEWELL A. and O'REILLY J.J.: 'Runlength limited binary error control codes', *IEE Proceedings Part I*, 1992, **139**, pp. 349–355.

[2] POPPLEWELL A. and O'REILLY J.J.: 'A new class of runlength limited error control codes with minimum distance 4', *IEE Proceedings Part I, in press.*

[3] POPPLEWELL A. and O'REILLY J.J.: 'Runlength limited codes for random and burst error correction', *IEE Electronics Letters*, 1992, **28**, pp. 970–971.

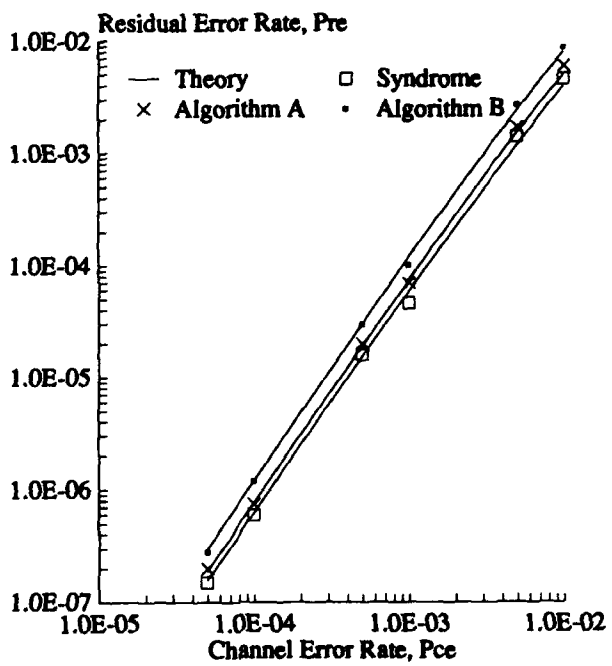


Figure 1: Error performance of (64,46) code with different decoders

A SCHEME FOR COMBINED MODULATION AND ERROR CORRECTION

Khaled A. S. Abdel-Ghaffar*
University of California
Dept. of E.E. and C.S.
Davis, CA 95616
USA

Mario Blaum
IBM Research Division
Almaden Research Center
San Jose, CA 95120
USA

Jos H. Weber
Delft University of Technology
Dept. of Electrical Engineering
2600 GA Delft
The Netherlands

Abstract

A technique for joint modulation and error correction is described. In order to construct a (d,k) modulation code with κ information bytes that corrects up to t errors, a single error-detecting (inner) block modulation code is combined with an (outer) $[\kappa+t,\kappa]$ Reed-Solomon code. The performance of this scheme is compared to the related scheme using an (inner) block modulation code and an (outer) $[\kappa+2t,\kappa]$ Reed-Solomon code, as well as to the traditional method in magnetic recording, which involves the concatenation of an error-correcting code with a sliding window modulation code.

1 Introduction

A well-known method to construct a (d,k) modulation code with error-correcting capabilities is to concatenate an inner block modulation code with an outer Reed-Solomon code [2]. To be more precise, let $\mathcal{I}$ be a binary block code of size 2^n for which the cascading of codewords gives sequences with runlengths of at least d and at most k 0's between any two consecutive 1's. For combined error protection against up to t errors and modulation of $n\kappa$ information bits, we concatenate the inner code $\mathcal{I}$ with an outer $[\kappa+2t,\kappa]$ Reed-Solomon code $\mathcal{O}$ over GF(2^n).

In this paper, we consider a modification of the above-described scheme by choosing the inner code to be single error-detecting. Since a single error in an inner codeword will thus always result in a symbol erasure for the outer code, $\mathcal{O}$ only needs to be a $[\kappa+t,\kappa]$ Reed-Solomon code in order to correct up to t bit errors.

The idea of using an error-detecting inner code is not completely new. In fact, this scheme can be considered as a special case of Ytrehus' general scheme for constructing runlength-limited codes for the mixed-error channel [6] (by choosing $s=0$ in this scheme, while Ytrehus himself accents the case $s=2$).

The single error-detecting capability of the inner code can be established by choosing (d,k) constrained sequences of either odd or even weight [4],[5]. However, by using more advanced methods like the one presented in [1], we occasionally obtain higher rates, especially when k is close to $2d$.

2 Comparisons

The traditional way to establish modulation and error protection in magnetic recording involves the concatenation of an interleaved error-correcting code with a sliding window modulation code [2]. We have compared the performance of this traditional scheme with the two block schemes from the previous section. Both analytical and simulation methods have been used. Special attention has been paid to the important case $(d,k)=(1,7)$.

We have calculated and compared for a fixed number of information bits the total redundancy for each method. Among other results, it turned out that the block scheme with the error-detecting inner code minimizes the redundancy when κ is relatively small. By establishing bounds on the rate difference for the inner block codes with and without error detection, the two block schemes could be compared.

Note that a bit error often causes a violation of the (d,k) constraints. For a block scheme, this leads to a symbol erasure, since the received (inner) word does not any longer correspond to an (outer) symbol. Hence there is a kind of interaction between the demodulator and the decoder in the block schemes, which seems to be missing in the traditional scheme. Since it is hard to measure the effect of this interaction analytically, we have run various simulations. Noise was injected based on observations made in [3]. For the $(d,k)=(1,7)$ case, the results seem to indicate that it is hard to improve upon the traditional scheme. However, it was also noticed that for the block scheme, the k-constraint can be lowered from 7 to 5 without losing rate.

References

[1] K.A.S. Abdel-Ghaffar and J.H. Weber, "Bounds and Constructions for Runlength-Limited Error-Control Block Codes," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 789-800, May 1991.

[2] M. Blaum, "Combining ECC with Modulation: Performance Comparisons," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 945-949, May 1991.

[3] T.D. Howell, "Analysis of Correctable Errors in the IBM 3380 Disk File," *IBM J. Res. Develop.*, vol. 28, pp. 206-211, March 1984.

[4] K.A. Schouhamer Immink, "Error Detecting Runlength-Limited Sequences," Eighth International IEE Conference on Video, Audio, and Data Recording, Birmingham, April 1990.

[5] P. Lee and J. K. Wolf, "A General Error-Correcting Code Construction for Run-Length Limited Binary Channels," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 1330 1335, November 1989.

[6] Ø. Ytrehus, "Runlength-Limited Codes for Mixed-Error Channels," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1577-1585, November 1991.

*This author was supported in part by NSF Grant NCR 89-08105 and by an IBM Faculty Development Award.

Resynchronizing (d, k)-Constrained Sequences in the Presence of Insertions and Deletions

Mario Blaum, Jehoshua Bruck and C. Michael Melas
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120
USA

Henk C. A. van Tilborg
Department of Mathematics and Computer Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

A typical encoding configuration for a magnetic or optical recording channel consists of encoding the information bits with an error-correcting code, generally, a Reed Solomon code, followed by a (d, k) constrained code [1].

In general, Reed-Solomon codes can handle the most common type of errors: random errors and peak shifts. A random error can be of two types: a 0 becomes a 1, denoted 0→1, or a 1 becomes a 0, denoted 1→0. Peak shifts are also of two types: 0 1→1 0 or 1 0→0 1.

However, there are other types of errors that cause a catastrophic failure due to loss of synchronization. They are, deletion of a symbol (0 or 1) and insertion of a symbol (0 or 1). Although deletions and insertions are not as common as the other types of error, if we are able to determine how many insertions or deletions occurred in an interval, by inserting or deleting a proper amount of symbols we are going to have a burst error that will either be corrected by the outer error-correcting code or, if uncorrectable, at least it will have a limited length.

Consider $(1, 7)$ sequences (the method can be generalized to any (d, k) sequence). We make the following 1-1 mapping between a $(1, 7)$ sequence and symbols in Z_7 (i.e., set of integers modulo 7): to each run of 0's, we associate the number of zeros minus one.

If we denote by L the length of the binary string, by ℓ the length of the 7-ary string and by S the sum of the symbols in the 7-ary string, these three parameters are related by $L = S + 2\ell$.

At the 7-ary level, we encode the information using an $[n, n-2]$ block code, where $n \geq 7$. The first and the last symbols in a block are redundant, while the middle $n-2$ symbols carry the information. We require that in each block, the sum of the symbols modulo 7 is 0. The last symbol in a block and the first symbol in the next block are chosen in such a way that their sum is equal to 6. Thus, we are inserting exactly 10 binary symbols between blocks in the binary sequence. It is important to have a fixed amount of redundancy while attempting to recover synchronization. Finally, we set the initial condition $a_0 = 0$.

At the receiving end, if a 7-ary sequence $b_0, b_1, b_2, \ldots$ has been received, and errors have occurred, including possible insertions and/or deletions of symbols, we show how to recover synchronization with high probability under the following conditions (that are determined by the error statistics of the channel):

1. At most 3 errors in at most λ consecutive 7-ary blocks of length n have occurred, say in blocks $m, m+1, \ldots, m+r$, where $r \leq \lambda - 1$.
2. After the last block in error, say block $m+r$, there are at least s error-free blocks.
3. The length n of each block is at least 7 (in general, $k - d + 1$).

Under these conditions, we present a method that will allow us to determine how many symbols have been deleted, allowing for recovery of synchronization.

References

[1] P. H. Siegel, "Recording Codes for Digital Magnetic Storage," IEEE Trans. on Magnetics, Sept. 1985, pp. 1344-1349.

Construction of Insertion/Deletion Correcting RLL Codes

Patrick A.H. Bours

Department of Mathematics and Computing Science,
Eindhoven University of Technology
PO-Box 513, 5600 MB Eindhoven,
The Netherlands.

Abstract

An algorithm is presented for the construction of fixed length insertion/deletion correcting RLL codes. This algorithm uses one or more fixed length q-ary codes with given Lee-distance to generate fixed length binary (d,k)-constrained codewords. This construction can be used for all possible (d,k)-constraints.

Introduction

In [1] the authors describe a way to construct peak-shift error correcting variable length RLL codes. In their construction they use a Hamming-metric based code to generate (d,k)-constrained codewords. However, due to the way the problem is stated, it is more obvious that Lee-metric based codes are used as generating codes. This was also noticed by Roth en Siegel in [2]. The main idea is to use a Lee-metric based code to encode the runlengths of an RLL code. This allows us not only to correct peak-shift errors but also insertions and deletions of zeroes.

Preliminaries

In the sequel we assume that an error has some maximal size of s bits. So, in case of a shift error, a one is shifted over at most s positions, and in case of an insertion (resp. deletion) error, at most s zeroes are inserted (resp. deleted). Now q will be defined as $2 \cdot s+1$. Furthermore, if $\mathbf{x} = 10^{\alpha_1}10^{\alpha_2}\ldots10^{\alpha_l}$ is a (d,k)-constrained word, then the Integer Representation (IR) $\boldsymbol{\beta}$ of $\mathbf{x}$ is defined by

$$\beta_i := (\alpha_i - d) \bmod q, \quad i = 1,2,\ldots,l. \tag{1}$$

The absolute weight $W_{abs}(\boldsymbol{\beta})$ of a q-ary vector $\boldsymbol{\beta}$ of length l is defined as

$$W_{abs}(\boldsymbol{\beta}) := \sum_{i=1}^{l} \beta_i, \tag{2}$$

where the sum is taken over the integers.
In the sequel C_q will denote a q-ary code of length l and minimum Lee-distance t, and C_2 will denote a binary (d,k)-constrained code of length $N_2 := n + l \cdot (d+1)$ for some $n \geq 0$.
If $t = 2 \cdot r \cdot s + 1$, then C_q is capable of correcting r errors, where each error has size at most s.

The Construction

We are now able to give a construction for the code C_2, using the code C_q.

Construction:

1. Take $\boldsymbol{\beta} \in C_q$ with $W_{abs}(\boldsymbol{\beta}) =: L$ and $(n - L) \equiv 0 \pmod q$.
2. Take all $\mathbf{x} \in \mathbb{F}_2^{n+l\cdot(d+1)}$ with IR $\boldsymbol{\beta}$, that also satisfy the (d,k) constaints. Do this as follows:
   (a) Define r_1 and r_2 such that
   $$\begin{cases} k - d = r_1 \cdot q + r_2, \\ r_1 = \lfloor \frac{k-d}{q} \rfloor, \\ r_2 = (k-d) \bmod q. \end{cases} \tag{3}$$
   (b) Find all vectors $\boldsymbol{\gamma} \in \mathbb{Z}^l$ such that
   $$\begin{cases} \sum_{i=1}^{l} \gamma_i = \frac{n-L}{q}, \\ 0 \leq \gamma_i \leq r_1 \text{ if } \beta_i \in \{0,1,\ldots,r_2\}, \\ 0 \leq \gamma_i \leq r_1 - 1 \text{ if } \beta_i \in \{r_2+1, r_2+2, \ldots, q-1\}. \end{cases} \tag{4}$$
   (c) Take as a codeword in the code C_2 the word
   $$\mathbf{x} = 10^{\beta_1+d+\gamma_1\cdot q}10^{\beta_2+d+\gamma_2\cdot q}\ldots10^{\beta_l+d+\gamma_d\cdot q}. \tag{5}$$
   (d) Repeat step 2(c) for all $\boldsymbol{\gamma}$ of step 2(b).
3. Repeat step 2 for all $\boldsymbol{\beta}$ of step 1.

Decoding is now done by taking l runs at a time, and then decoding the IR of the word they form together. Due to the fact that we have assumed that an error has maximal size s, and the alphabet of the code C_q has size $2 \cdot s + 1$, it is always possible to distinguish between insertion and deletion errors.
In general not all codewords $\boldsymbol{\beta}$ of the code C_q can be used to generate binary (d,k)-constrained codewords of length N_2. This is due to the fact that it does not always hold that

$$W_{abs}(\boldsymbol{\beta}) \equiv n \pmod q. \tag{6}$$

This can be solved by adding a parity symbol β_{l+1} to a q-ary codeword $\boldsymbol{\beta}$, such that

$$\sum_{i=1}^{l+1} \beta_i + (l+1)\cdot(d+1) \equiv N_2 \pmod q, \tag{7}$$

or equivalently

$$\beta_{l+1} = (n - (d+1) - W_{abs}(\boldsymbol{\beta})) \bmod q. \tag{8}$$

In order to increase the number of codewords of the code C_2, we can use more q-ary codes C_q, say C_q^i, for $i = 1,2,\ldots,p$, where the code C_q^i has length l_i. For the lengths l_i it most hold that

$$\begin{cases} \lfloor \frac{N_2}{d+1} \rfloor \leq l_1 \\ l_i + \lceil \frac{(t-1)/2}{q+1} \rceil < l_{i+1} - \lfloor \frac{(t-1)/2}{d+1} \rfloor \\ l_p \leq \lceil \frac{N_2}{d+1} \rceil \end{cases} \tag{9}$$

Furthermore, if we assume that only 1 run can be affected by errors, we can take q to be $s+1$. This is due to the fact that the code C_2 has a fixed length N_2.

References

[1] H.M. Hilden, D.G. Howe, E.J. Weldon Jr., *Shift error correcting modulation codes*, IEEE Trans. on Magn., Vol. MAG-27, No 6 (November 1991), pp. 4600-4605.

[2] Ron M. Roth, Paul H. Siegel, *Lee-Metric BCH Codes and their Application to Constrained and Partial-Response Channels*, preprint.

The APPLICATION of q–ary CODES for the CORRECTION of SINGLE PEAK–SHIFTS, DELETIONS and INSERTIONS of ZEROS

A.V. Kuznetsov[1], A.J. Han Vinck[2]

[1]A.V.Kuznetsov, IPPI Ermolovoy 19, Moscow.101447 Russia.

[2] A.J. Han Vinck, IEM, Ellernstr. 29, 4300 Essen, Germany.

Abstract. We construct q–ary block codes that allow correction of specific types of double errors. These codes can be used as codes for correction of peak–shifts, deletions and insertions of zeros in (d,k)–sequences applied in magnetic recording. For single peak–shifts over $t \le (k-d)/2$ positions left or right, the codes have dimension $N=q^r$, $K=q^r-(r+1)$, $q=k-d+1$. An additional condition on the structure of the code gives transparent block codes which are used to control the maximum binary length of the code words. Encoding and decoding are done by simple algorithms without using look–up tables, enumeration or denumeration procedures and therefore the code length may be large. The rate of the overall encoding approaches $(2\log_2(k-d+1))/(k+d+2)$ for large code word lengths.

A. CORRECTION of SINGLE PEAK–SHIFTS

For the transmission through (d,k)–constrained channels, q–ary code words (where $q=k-d+1$) are converted to binary sequences satisfying the (d,k)–constraint by replacing q–ary components by binary strings of i, $d \le i \le k$, consecutive zeros followed by a single one. If at most a single peak–shift of value t occurs, then the output code word of the encoder $\underline{x} \in C$ and the input word $\underline{z}$ of the decoder are related by the equation $\underline{z} = \underline{x} + \underline{e}$, where + is the componentwise addition of integers, and $\underline{e} = (e_1, e_2, \ldots, e_N)$ is an error vector with integer components e_i, that belongs to one of the following three classes:

1) $e_i = 0$ for $1 \le i \le N$ (no errors); (1)

2) $e_i = 0$ for $1 \le i \le N-1$ and $e_N \ne 0$; (2)

3) $e_j = t$, $e_{j+1} = -t$ for some $1 \le j \le N-1$, $e_i = 0$ for $i \ne j, j+1$. (3)

We have related the problem of peak–shift correction to the construction of block codes over the ring of integers modulo q correcting double errors of the type (3) and a single error in the last component of the code word (2).

Let $N = q^r$, where $q = k-d+1 \ge 3$ is a prime, and $r \ge 1$ is an arbitrary integer. For peak–shift correction we use a q–ary linear code C of length N defined by the parity check matrix $H = \|h_{ij}\|$ with two rows of following elements $h_{1,j} \in GF(q) = \{0,1,\ldots,q-1\}$ and $h_{2,j} \in GF(q^r)$:

$$h_{1,j} = j \bmod q, \qquad 1 \le j \le N; \qquad (4)$$

$$h_{2,j+1} = h_{2,j} + w_j, \qquad 1 \le j \le N-1, \qquad (5)$$

where $w_1, w_2, \ldots, w_{N-1}$ are distinct nonzero transposed r–tuples with components from GF(q), $h_{2,1}$ is the transposed r–tuple $\bar{1} = (1,0,\ldots,0)$, and + in (5) represents componentwise modulo q addition of r–tuples.

Transparency. As was pointed in [1], for the maximum length control the code C must be transparent, that is, the all–ones word $\underline{1} = (1,1,\ldots,1)$ of length N must belong to the code C. This condition can be satisfied in several ways. For example, for any prime $q \ge 3$ and $r \ge 1$ (except the case q=3 and r=1) , as an element w_i we may use the ordinary q–ary representation of its index $i = 1,2,\ldots,q^r-1$ considered as an integer.

Examples. The parity check matrices H_3 for N = 9 (q=3, r=2) and H_7 for N = 7 (q=7, r=1) constructed according to the described procedure are given below.

$$H_3 = \begin{matrix} 1\,2\,0\,1\,2\,0\,1\,2\,0 \\ 1\,2\,1\,1\,2\,1\,1\,2\,1, \\ 0\,0\,0\,1\,2\,0\,2\,1\,0 \end{matrix} \qquad H_7 = \begin{matrix} 1\,2\,3\,4\,5\,6\,0 \\ 1\,2\,4\,0\,4\,2\,1 \end{matrix}.$$

Proposition 1. The linear q–ary code defined by the parity check matrix H as given in (4)–(5), has length $N = q^r$, and $K \ge N-(r+1)$ information symbols. The code corrects peak–shifts (1)–(3) of size t, $t \le (k-d)/2$ and is transparent.

B. PEAK–SHIFTS DELETIONS and INSERTIONS of ZEROS

The proposed method can be used for the correction of other types of errors. In this section we present codes that can correct in the (d,k)–sequence a single distortion of the following type:

a) a peak–shift on $(k-d)/2$ or less positions;
b) a deletion or an insertion of $(k-d)/2$ or less zeros between adjacent one's.

The list of possible types of error vectors (1)–(3) is extended with the following

$$e_j = t \text{ for some } 1 \le j \le N,\ e_i = 0 \text{ for } 1 \le i \ne j \le N. \qquad (6)$$

In fact, errors of type (2) are particular cases of (6), and thus later we consider only three types (1),(3) and (6).

Let $N = q^r$, where $q = k-d+1 \ge 3$ is a prime, $r \ge 2$ is an arbitrary integer, and let γ be a primitive element of $GF(q^r)$ such that the element $\gamma^{q-3}(1-\gamma)$ is not an integer in $GF(q^r)$. For values of q and r such that $q^r \le 128$, Tables from [2, Chapter 10] can be used to select a primitive element γ that satisfies this condition. For the correction of errors (1)–(3) and (6) we use a q–ary linear code C defined by the parity check matrix

$$\begin{matrix} 1 & 1 & 1 & 1 & \ldots & 1 & 1 & 1 & 1 & \ldots & 1 \\ 2 & 2 & 3 & 4 & \ldots & q-1 & q-1 & 0 & 1 & \ldots & q-1 \\ 0 & \gamma^1 & \gamma^2 & \gamma^3 & \ldots & \gamma^{q-2} & \gamma^{q-1} & \gamma^q & \gamma^{q+1} & \ldots & \gamma^{N-1} \end{matrix} \qquad (7)$$

with elements $h_{1,j}$, $h_{2,j} \in GF(q)$ and $h_{3,j} \in GF(q^r)$. The code dimensions are $N = q^r$, $K = q^r-(r+2)$. The code defined by this parity check matrix is transparent. This follows from the definitions and the fact, that the summation of all elements in $GF(q^r)$ gives 0 for any $r \ge 1$. As an example for q=3 and r=2 the parity check matrix is shown below ($\gamma = (1,0)^{Tr}$, and $GF(3^2)$ is represented as in Tables of [2]),

$$H = \begin{matrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 2 & 2 & 0 & 2 & 1 & 1 & 0, \\ 0 & 0 & 1 & 2 & 2 & 0 & 2 & 1 & 1 \end{matrix}$$

Error correction uses the syndrome $\underline{S} = (S_1, S_2, S_3)^{Tr} = H \cdot \underline{z}^{Tr}$.

Proposition 2. The linear q–ary code defined by the parity check matrix H as given in (7), is transparent, has length $N = q^r$, and $K = N-(r+2)$ information symbols. The code corrects peak–shifts of size t (1)–(3) and t insertions and deletions of zeros (6) for $t \le (k-d)/2$.

REFERENCES.

[1] A. Kuznetsov and A.J. Han Vinck, "Single Peak–Shift Correction in (d,k)–sequences," ISIT 1991, Budapest, Hungary, 1991, pp. 256.

[2] R. Lidl, H. Niederreiter, "Introduction to finite fields and their applications," Cambridge University Press, 1986.

A CONSTRUCTION OF CODES WITH SPECIAL PROPERTIES

Alexander Barg
Institute for Problems
of Information Transmission
Ermolovoy 19, Moscow GSP-4
101447
E-mail: aBarg@ippi.msk.su

In this talk we present a new construction of codes with zero or constrained power of signal at zero frequency, of codes that detect synchronization failures, and of a family of sequences with low periodic correlation. The first problem has been recently given a considerable attention (see [1]) while the two latter problems have been studied since long ago [2],[3].

Let $A[k, n, M, d]$ denote a k-ary code of length n, with M words and Hamming distance d. We assume that the alphabet letters are denoted by the kth degree roots of unity.

The running digital sum of a code A is, by definition,

$$S(A) := \vec{a} \in A \rightarrow \max S(\vec{a}).$$

Following [4], let us call a family $\{A_n\}$ of codes of growing length n *dc-constrained* if

$$S(A_n) \le c(n),$$

where $c(n) = o(n)$ is a slowly growing function in n.

The capacity of a code to detect synchronization errors is determined by the *code separation* [2]. For k-ary vectors $\vec{a} = (a_0, a_1, \ldots, a_{n-1})$ and $\vec{b} = (b_0, b_1, \ldots, b_{n-1})$, let us introduce an n-vector

$$T_i(\vec{a}, \vec{b}) = (a_i, a_{i+1}, \ldots, a_{n-1}, b_0, b_1, \ldots, b_{i-1}), \ 1 \le i \le n-1.$$

The code separation is defined by

$$\rho(A) := \vec{a}, \vec{b}, \vec{c} \in A \rightarrow \ 1 \le i \le n-1 \rightarrow (T_i(\vec{a}, \vec{b}), \vec{c}).$$

Codes with $\rho(A) > 0$ are called *comma-free*.

Finally, the *periodic correlation* of two complex vectors $\vec{a}, \vec{b}$ is defined by:

$$\theta_{\vec{a},\vec{b}}(\tau) := \sum_{t=0}^{n-1} a_{t\oplus\tau} b_t^*, \quad 0 \le \tau \le n-1.$$

Consider the following code construction [4-5]. Let $q = p^m$, where p is an odd prime, and let $\chi(\cdot)$ be a multiplicative character of the field F_q of order $k|(q-1)$. Consider the set P of monic polynomials $f(x), 1 \le f \le r$, that satisfy the following restriction: in the expansion into irreducibles $f = \prod_i g_i^{e_i}$ all $e_i \le k-1$. Consider a code A with its vectors defined by

$$\vec{a}^{(f)} = \chi(f(\beta_i)), \quad 1 \le i \le q-1, f \in P \qquad (1)$$

where $(\beta_0, \ldots, \beta_{q-1})$ is some ordering of the field elements.

Theorem

(i) [5]. Let $k \ge 3$. The construction (1) defines the dc-constrained code A with the parameters $[k, q, M \sim q^r, d \ge ((k-1)/k)(q - 2r\sqrt{q}) - 2r]$ and $S(A) < srp^{s/2}(1 + \log p)$.

(ii) [5]. Let q be an odd prime and $k \ge 3$. The code A defined by (1) contains a comma-free subcode $A_1[k, q, M \sim q^{r-1}, d \ge ((k-1)/k)(q - 2r\sqrt{q}) - 2r]$ with $\rho(A) \ge ((k-1)/k)(q - 4r\sqrt{q}(1 + \log q))$.

(iii) For any two cyclically distinct vectors $\vec{a}^{(f)}$ and $\vec{a}^{(h)}$ defined by (1),

$$\theta_{\vec{a}^{(f)},\vec{a}^{(h)}}(\tau) \le (2r-1)\sqrt{q}.$$

The proof utilizes estimates of incomplete character sums similar to the Vinogradov–Polya inequality [6].

References

[1] R. Karabed and P.Siegel, Matched spectral-null codes for partial-response channels, *IEEE Trans. Inform. Theory*, **IT-37** (1991), 818–855.

[2] V. I. Levenshtein, Bounds for codes that provide error correction and synchronization, *Problemy Peredachi Inform.*,5,2 (1969), 3–13, and *Probl. Inform. Trans.* **5**, 1969.

[3] D. Sarwate and M. Pursley, Cross-correlation properties of pseudorandom and related sequences, *Proc. IEEE*, **68**, 5 (1980), 593–618.

[4] A. Barg and S. Litsyn, DC-constrained codes from Hadamard matrices, *IEEE Trans. Inform. Theory*, **IT-37**,3,Pt.2 (1991), 801–807.

[5] A. Barg, Incomplete sums, dc-constrained codes, and codes that maintain synchronization, *Designs, Codes, and Cryptography*, to appear.

[6] I. M. Vinogradov, *Elements of Number Theory*. 9th ed., Moscow, 1981, in Russian.

THE ROLE OF INFORMATION THEORY IN EMISSION TOMOGRAPHY

Larry Shepp
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

In emission tomography, useful for studying brain function, a source of radioactivity is ingested by a person, say as sugar, and a Poisson number, $n(b)$, of radioactive emissions arises in each box (pixel), b, of the brain depending on the brain activity there, and $\lambda(b) = En(b)$ is sought. Each emission (not directly observable) makes an independent Markovian transition to some detector unit, d, with probability $p(b,d)$, $\sum_d p(b,d) = 1$, where $p(b,d)$ is known from the geometry and performance of the detectors. We measure $n^*(d)$, the total number of counts in each d and wish to estimate $\lambda(b)$ to get an image of the brain activity, say during counting, speaking, or other function.

For each λ there is a likelihood (see Appendix A), $\Lambda(\lambda)$, to observe n^* and one popular approach to reconstructing or estimating λ is to seek a maximum likelihood estimator (MLE). Surprisingly enough, ideas of information theory have provided useful insight into the theoretical understanding of MLE even though entropy doesn't appear to be directly involved.

Noone knows how to produce an MLE directly but the so-called EM algorithm is used beginning with an initial λ^0 to produce ever more likely $\lambda^1, \lambda^2 \ldots$ estimates.

The only rigorous proof [1] of convergence of λ^n to a limit maximizing $\Lambda(\lambda)$ is heavily information theoretic. Unfortunately this limiting MLE was seen [2] not to be a robust estimate — due to the fact that $n(b)$ is small and hence statistically noisy — and indeed was totally useless as a practical image. If MLE were not unique then the various ML estimators could be averaged, and since $\Lambda(\lambda)$ is seen to be log concave (see Appendix A), an estimate could be obtained which is both smooth as well as maximally likely. On empirical grounds it was conjectured [2] in 1988 that MLE was, under general conditions, unique. Very recently, again using ideas of information theory, Charles L. Byrne, succeeded [3] to formulate a general and natural hypothesis on $p(b,d)$ under which the conjecture is true. This dashes all hope that smooth MLE's exist in practical emission tomography. The present approaches involve either stopping the iteration early, smoothing at each step or at the end, or maximizing posterior likelihood with a Gibbs prior.

I hope information theory will continue to shed light on emission tomography.

References

[1] Csiszàr, I. and Tusnàdy, G. (1982), Information geometry and alternating minimization procedures, Math. Inst. Hungarian Academy of Sciences.

[2] Shepp, L. A. and Vanderbei, R. J., New insights into emission tomography via linear programming, notes prepared for Nato meeting on Formulation, Handling, and Evaluation of Medical Images, 12-23 September 1988, Portugal, widely distributed.

[3] Byrne, Charles L., Iterative image reconstruction algorithms based on cross-entropy minimization, IEEE Trans. on Inf. Theory, to appear.

[4] Vardi, Y. Shepp, L. A., and Kaufman, L., A statistical model for position emission tomography, J. Amer. Stat. Assoc. 1985, vol. 80, pp. 8-20.

Appendix A

So as not to interrupt the main ideas this paragraph serves to explain to the uninitiated reader what $\Lambda(\lambda)$ is and why it is log concave.

Observe that $n^*(d)$ are independent Poisson variables since $n(b)$ are independent and Poisson and each transition from each b to some d is done independently. Thus if $n(b,d)$ is the number of emissions in b that become counts in d then by the thinning property of the Poisson law, $n(b,d)$ are all independent Poisson for different b's and for different d's. But $n^*(d) = \sum_b n(b,d)$ and so $n^*(d)$ are also Poisson and independent.

Thus $\Lambda(\lambda) = \prod_d e^{-\lambda^*(d)} \lambda^*(d)^{n^*(d)} / n^*(d)!$ where $\lambda^*(d) = En^*(d) = E\sum_b n(b,d) = \sum_b \lambda(b) p(b,d)$. We seek an MLE which maximizes $\Lambda(\lambda)$. It is easy to see from this formula that the Hessian of $\log \Lambda(\lambda)$ is negative definite and so Λ is log concave. For more details see [4].

Recursive CR-Type Bounds and the EM Algorithm: Applications to ECT Image Reconstruction[1]

A.O. Hero* and J.A. Fessler**
*Dept. of Electrical Engineering and Computer Science and **Division of Nuclear Medicine
The University of Michigan, Ann Arbor, MI 48109-2122

ABSTRACT

We give a class of iterative algorithms to monotonically approximate submatrices of the CR matrix bound on the covariance of any estimator of a vector parameter $\underline{\theta}$. A natural implementation of the iterative algorithm employs a "complete data - incomplete data" formulation similar to that underlying the EM parameter estimation algorithm. Our results make it feasible to compute CR-type bounds for previously intractible problems involving a large number of "nuisance parameters," such as arise in image reconstruction.

I. Summary

The Cramer-Rao (CR) bound on estimator covariance is an important tool for predicting fundamental limits on best achievable parameter estimation performance [5], predicting the impact of side information and constraints on estimation performance [3], and obtaining optimal experimental designs [1]. For a vector parameter $\underline{\theta} \in \Theta \subset \mathbb{R}^n$ the upper left $p \times p$ matrix of the inverse of the $n \times n$ Fisher information matrix provides the CR lower bound on the minimum achievable covariance of any unbiased estimator of $\theta_1, \ldots, \theta_p$, $p \leq n$. Equivalently, the first p rows of F_Y^{-1} provide the CR bound. The method of sequential partitioning [4] for computing the upper left $p \times p$ submatrix of F_Y^{-1} and Cholesky based Gaussian elimination techniques [2] for computing the p first rows of F_Y^{-1} are efficient direct methods for obtaining the CR bound but require $O(n^3)$ floating point operations and $O(n^2)$ memory storage. Unfortunately, in many practical cases of interest, e.g. when there are a large number of nuisance parameters, high computation and memory requirements make direct implementation of the CR bound impractical. For example, in the case of image reconstruction for a 256×256 pixelated image F_Y is $256^2 \times 256^2$ so that direct computation of the CR bound on estimation errors in a small region of the image requires on the order of 256^6 or 10^{19} floating point operations and on the order of 4GByte of memory storage!

In this paper we give a class of iterative algorithms for computing columns of the CR bound which requires only $O(n^2)$ floating point operations per column of F_Y^{-1}. These algorithms fall into the class of "splitting matrix iterations" [2]. The inverse of this splitting matrix should be sparse and simply determined. The splitting matrix is chosen based on purely algebraic or purely statistical considerations to ensure that a valid lower bound results at each iteration of the algorithm. By embedding the parameter estimation problem into a particular complete data - incomplete data setting, and applying a version of the "data processing theorem" for Fisher matrices, the Fisher matrix F_X for the complete data set can frequently be used as a splitting matrix. This complete-incomplete data setting is similar to that which underlies the classical formulation of the EM algorithm. The EM algorithm generates a sequence of estimates $\{\hat{\underline{\theta}}^k\}_k$ for $\underline{\theta}$ which successively increase the likelihood function and converge to the maximum likelihood estimator. In a similar manner, our algorithm generates a sequence of tighter and tighter lower bounds on estimator covariance which converge to the actual CR matrix bound. The iterative algorithm allows one to compute the CR bound for estimation problems that would have been intractible by exact methods due to the large dimension of F_Y.

We conclude with an implementation of the recursive algorithm for bounding the minimum achievable estimator error covariance for problems arising in emission computed tomography (ECT). For the case where the complete data is selected as the set of image pixel emission counts in each of d "detector tubes", which is the standard choice of complete data for the EM image reconstruction algorithm, F_X is diagonal. Furthermore, due to the sparseness of the tomographic system response matrix the computation of each column of the CR bound matrix recursion only requires $O(n)$ memory storage as compared to $O(n^2)$ for the general algorithm. We show that in general the rate of convergence depends on the image intensity and the tomographic system response matrix. We have applied the iterative algorithm to compute the CR bound for practical estimation tasks including: reconstruction of a small region-of-interest (ROI), estimation of total uptake in a ROI, estimation of dose distribution heterogeneity in a ROI, impact of anatomical side information on ROI reconstruction.

[1]This research was supported in part by the National Science Foundation under grant BCS-9024370, the National Cancer Institute under grant R01-CA-54362-02, and a DOE Alexander Hollaender Postdoctoral Fellowship.

REFERENCES

[1] V. Federov, *Theory of Optimal Experiments*, Wiley, New York, 1972.

[2] G. H. Golub and C. F. Van Loan, *Matrix Computations (2nd Edition)*, The Johns Hopkins University Press, Baltimore, 1989.

[3] J. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. on Inform. Theory*, vol. IT-36, pp. 1285–1301, Nov. 1990.

[4] A. R. Kuruc, "Lower bounds on multiple-source direction finding in the presence of direction-dependent antenna-array-calibration errors," Technical Report 799, M.I.T. Lincoln Laboratory, Oct., 1989.

[5] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

SIMULTANEOUS RECOVERY OF THE OBJECT AND ABERRATIONS FROM A SEQUENCE OF IMAGES DEGRADED BY ATMOSPHERIC TURBULENCE

Timothy J. Schulz

Michigan Technological University
Department of Electrical Engineering
Houghton, MI 49931
(906) 487-2754 Email: schulz@mtu.edu

Summary

Atmospheric turbulence severely limits the effective resolution of a long-integration image obtained by an uncompensated, ground-based telescope. Because of this fact, most ground-based telescopes typically collect a sequence of short-exposure images. The simplest and most widely used model for the intensity of the kth short exposure image-intensity is

$$r_k(x) = h_k(x) * s(x), \tag{1}$$

where $s(x)$ represents the light-intensity distribution of the object being viewed and $h_k(x)$ represents the point-spread function due to the telescope's finite aperture the turbulent nature of the Earth's atmosphere. The point-spread functions are commonly modeled as

$$h_k(x) = \mathcal{K}_k \left| \mathcal{F}\left\{ A(u) e^{j\Phi_k(u)} \right\} \right|^2, \tag{2}$$

where $\mathcal{K}_k$ is a constant that depends, among other factors, on the duration of the kth data-collection interval, $\mathcal{F}$ denotes the Fourier transform operator, $A(u)$ is a known, binary function that describes the telescope's aperture, and $\Phi_k(u)$ describes the turbulence-induced phase-aberrations that occur during the kth data-collection interval.

In all real situations, the intensities $\{r_k(x)\}$ are not detected perfectly. Instead, they are corrupted by some type of noise. Examples include read-out noise for charge-coupled-devices (CCDs) and photon noise for photon-counting cameras. For this talk, I will discuss the situation for which the data are corrupted by photon noise. In this case, the data collected in the kth frame are denoted as $d_k(x)$ and, conditioned on the object intensity $s(x)$ and the point-spread function $h_k(x)$, $d_k(x)$ is Poisson-process whose intensity is $r_k(x)$. Further, for $k \neq j$, the processes $d_k(x)$ and $d_j(x)$ are statistically independent.

The estimation problem I address is then one of estimating the desired, information-bearing signal $s(x)$, from the measured data $\{d_k(x)\}_{k=1}^K$. When the atmospheric phase-aberrations $\{\Phi_k(u)\}$ are known, the point-spread functions $\{h_k(x)\}$ are known and a *multi-frame deconvolution problem* must be solved. When the atmospheric phase-aberrations are not known, the problem is much more difficult. In this case, the point-spread functions are not known and a *multi-frame blind-deconvolution problem* must be solved. This is the problem I address.

The phase-aberrations $\{\Phi_k(u)\}$ can be modeled as a collection of deterministic functions or they can be modeled as a collection of random-processes that fluctuate randomly with k. In this talk, I consider the first situation. However, when sound statistical models are available for the phase-aberration processes they should be used. The estimation problem is stated as one of forming the *maximum-likelihood estimates* of the information-bearing signal $s(x)$ and the phase-aberrations $\{\Phi_k(u)\}$, from the data $\{d_k(x)\}$.

In the talk, a numerical technique based on the expectation-maximization (EM) algorithm will be presented for forming solutions numerically. Examples using both simulated data and real, telescope data will also be presented do demonstrate the usefulness of the technique.

Searching for Circumstellar Disks with Space Telescope Observations

Donald Geman and Joseph Horowitz
Department of Mathematics and Statistics
University of Massachusetts at Amherst
Amherst, Massachusetts 01003

At present, there are no known examples of planetary systems other than our solar system, in which the orbits of the planets all lie nearly in the equatorial plane of the sun. It is conjectured that, soon after its birth, the sun was surrounded by a disk composed of dust and gas, out of which the planets agglomerated, the residual material being blown away by high energy winds along the polar axis of the sun. In fact, astronomers believe that many young stars are surrounded by extended, essentially planar objects composed of dust and gas, called "circumstellar disks," and that these are the environment in which planetary systems develop. Apparently, this brief episode of stellar evolution is part of a broader scenario, only loosely understood, thought to begin when a cold, rotating protostellar core condenses inside a large molecular cloud to form a star-disk system. Eventually, the star enters the main sequence (i.e., hydrogen-burning) stage, possibly accompanied by a planetary complex and other disk remnants.

Aside from our own solar system, the direct optical evidence for the existence of circumstellar disks is sparse. An extended object, thought to be a disk, was observed in 1984 around β Pictoris. In addition, the "infrared excess" observed around some stars is thought to be starlight absorbed by dust particles in a disk and reradiated at longer wavelengths, resulting in significant energy at infrared and other frequencies. Finally, there is indirect evidence for large planets derived from perturbations in stellar trajectories and velocities. (Direct imaging of planets is beyond current technology.)

This talk concerns the problem of detecting circumstellar disks based on Hubble Space Telescope (HST) observations. We are currently analyzing images recently obtained with the Wide Field Planetary Camera (WF/PC) of several nearby, pre-main sequence stars, both single and binary. Despite the advantages of placing a telescope outside the earth's atmosphere, the images taken with the WF/PC are still considerably degraded, mainly due to the severe blurring resulting from the infamous aberration in the optical system. The point spread function (PSF) for the WF/PC has significant mass over a radius on the order of one arc-second, and exposure times are limited by its instability. In addition, there are several other factors which limit the amount of information that is readily accessible (e.g., visually evident), including the usual limitations imposed by photon-limited data, stability problems with the spacecraft (resulting in "trailing"), local variations in the point spread function, variations in detector sensitivity and cosmic ray strikes.

The usual approach to image restoration results in a single "restored image," deemed to capture the original brightness pattern without degrading effects, or at least to suppress noise and enhance resolution. This approach is non-dedicated and nonparametric: except for specific knowledge of the image formation process, it incorporates only *generic* assumptions, for example constraints on the positivity, smoothness, or entropy of the brightness pattern. Examples of such techniques include those based on pseudo- inverses, maximum entropy, maximum likelihood, and Bayesian inference with "prior" and "posterior" distributions.

In contrast, we formulate the problem of the existence of disks as one of statistical hypothesis testing. The probability distribution of the data is derived by modeling the image formation process using the semiclassical model of photodetection (which means that quantum effects are accounted for only at the detection end of the system) as well as other important factors such as bias correction, quantum efficiency, and read-out noise. Basically, we wish to test the hypothesis H_0 :*star alone* vs. H_1 : *star plus something.* The test statistic is based on the (generalized) likelihood ratio. This is less straightforward than it might appear. For one thing, nearby "calibration stars" provide only *estimates* of the PSF, and hence there is a random factor in the mean brightness pattern which has to be "subtracted" from the observed one. In addition, it is necessary to adjust for the difference in overall brightness between the calibration and target stars. Finally, it is also necessary to account for bias correction, variability of detector sensitivity, and electrical noise. Results will be reported on at least one set of observations of several young stars in the Taurus-Auriga star forming complex.

A Model-Based Approach to Magnetic Resonance Image Estimation

Timothy J. Schaewe
IBM - Federal Systems Company
Owego, NY 13827-1298

Michael I. Miller
Washington University
St. Louis, MO 63130-4899

Model-based image reconstruction methods such as maximum-likelihood (ML) or maximum *a-posteriori* (MAP) estimation require a signal model describing the relationship between the image parameters to be estimated and the measurement data. Two of the parameters of interest in magnetic resonance imaging (MRI) are the tissue spin density, A, and spin-spin decay time constant, T_2. This paper presents a mathematical model for the signals observed in standard two-dimensional MRI experiments and discusses how this model is incorporated into a maximum *a-posteriori* parameter estimation algorithm to compute image estimates of spin density and spin-spin decay time. A detailed description of this work can be found in [1].

The basic response of a magnetically sensitive population of nuclei to excitation in a magnetic resonance experiment was described by Bloch [2] as an exponentially decaying sinusoid whose frequency is proportional to the strength of the static magnetic field to which the nuclei are exposed. In an MRI experiment, magnetic fields which vary with spatial position are employed to create a relationship in the observed signal between the frequency and phase of a sinusoidal signal component and the spatial position from which the signal originated.

The parameterized signal model which forms the basis of our MAP image reconstruction algorithm is based upon three assumptions: 1) The frequency and phase encoding magnetic fields used for spatial localization vary linearly with spatial position. 2) Voxels of dimension $D_x \times D_y\ cm^2$ are small enough that the spin density and spin spin decay time constant within a single voxel are constant. 3) The loss of signal coherence due to static magnetic field inhomogeneity results in signal attenuation which can be represented as an exponential decay with time constant T_M. Under these assumptions the signal emitted from a single voxel at position (x,y) takes the form

$$s(t,\tau) = \frac{\sin(\pi c_x D_x t)}{\pi c_x t}\,\frac{\sin(\pi c_y D_y \tau)}{\pi c_y \tau}\,A e^{-t/T_{2e}-t/T_{Me}} e^{j2\pi(f_x(x)t + f_y(y)\tau)}.$$

The frequencies of oscillation $f_x(x) = c_x x$ and $f_y(y) = c_y y$ are linear functions of the encoding gradient strengths c_x and c_y and position (x,y). The sinc-function parameters c_xD_x and c_yD_y are equal to the frequency bandwidths across the (x,y) dimensions of the voxel. The full two-dimensional MRI signal is represented as a superposition of sinc-modulated, exponentially decaying sinusoids of the type above, one from each of the $M \times N$ voxels which form the image field.

Under the assumption of additive, white, Gaussian noise in the MRI measurement data, the maximum likelihood estimates of the spin density and spin-spin decay image parameters are those parameters which minimize the squared error between the measurement data and a signal estimate computed from the image parameters using the model described above. To compute these image parameter estimates, we have implemented a form of the iterative expectation maximization (EM) algorithm of Dempster, Laird, and Rubin [3]. This algorithm has the property of decomposing the $2 \times M \times N$-dimensional least-squares optimization problem stated above into $M \times N$ independent 2-dimensional minimizations at each iteration, allowing for efficient parallel implementation of the algorithm. The algorithm also incorporates a Markov random field prior constraining the roughness of the computed image estimates, similar to the technique employed by Miller and Roysam [4] for emission tomography. The MRI parameter estimates produced by the algorithm, then, are MAP estimates rather than ML estimates.

Magnetic resonance image reconstruction is typically performed using a 2-dimensional Fourier transform. Under the assumption that the signal emitted from a single voxel is simply a non-decaying sinusoid, the Fourier transform is exactly the maximum likelihood solution and further computation is unnecessary. However, the more detailed signal model stated above provides additional information about the behavior of the MRI signal that allows for improved image estimates. The model used in our MAP algorithm exploits the fact that the sinc-modulated, exponentially decaying sinusoidal signal components oscillate in phase with one another, whereas the Gaussian noise is modelled as the superposition of non-decaying sinusoids with random amplitudes and uniformly distributed, random phases. The Fourier transform approach models signal and noise components identically, while the MAP method uses the differences between signal and noise to reduce the sensitivity of image parameter estimates to distortion by noise. For this reason, the MAP algorithm produces image parameter estimates which are of higher precision (i.e., lower variance) than than those computed using Fourier transform based techniques.

References

[1] T.J. Schaewe, "Maximum Likelihood Estimation for Magnetic Resonance Image Reconstruction," D. Sc. dissertation, Washington University, St. Louis, MO, 1991.

[2] F. Bloch, W.W. Hansen, and M.E. Packard, "Nuclear Induction," Phys. Rev. Vol. 70, 1946.

[3] A.D. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society. Vol. B-39, 1977.

[4] M.I. Miller and B. Roysam, "Bayesian Image Reconstruction for Emission Tomography Incorporating Good's Roughness Prior on Massively Parallel Processors," Proc. Natl. Acad. Sci., USA. Vol. 88, Apr. 1991.

Presented at the 1993 IEEE International Symposium on Information Theory.
T.J. Schaewe was supported for this work via an NSF Graduate Fellowship. M.I. Miller was supported by the NSF under a Presidential Young Investigator Award #8552518.

MODEL-BASED MULTIRESOLUTION RESTORATION OF SPECKLE IMAGES: APPLICATION TO RADAR IMAGING

P. MOULIN

Bell Communications Research
445 South Street
Morristown, NJ 07960

Abstract

We consider the problem of restoring images corrupted by speckle noise with iid-exponential statistics. This model occurs in a variety of coherent imaging problems including diffuse-target radar imaging. Our estimation approach is obtained by statistical inference on the wavelet coefficients of the logarithm of the image. Under a large-sample approximation, the inference problem takes a very simple form. Estimates can be obtained under various noise/resolution tradeoffs.

1. Diffuse Radar-Target Imaging

Maximum-likelihood (ML) estimation methods have recently been explored for forming images of diffuse radar targets [1,2]. The diffuse-target model assumes independent scatterers and models the reflectivity of the target as an uncorrelated, Gaussian random field. This model has been described in the electromagnetics literature and used in the signal processing literature.

We denote by $c(x,y)$ the reflectivity of the target in range and cross-range coordinates and we define $S(x,y)\,dx\,dy := E[\,|c(x,y)\,dx\,dy|^2\,]$. $S(x,y)$ is the *scattering function* of the target and is the desired image of the target. The received data are a linear transform of the reflectivity.

It is known that ML estimation of a whole function such as $S(x,y)$ from finite data is an ill-posed problem and requires regularization. A possible solution consists in representing $S(x,y)$ in terms of a small number of basis functions and estimating their coefficients [2]. In this paper, we present a regularization method based on a wavelet representation for $\ln S(x,y)$. This method offers the ability to capture significant components of $\ln S(x,y)$ at different resolution levels. This capability for multiresolution estimation allows for increased flexibility over single-resolution regularization techniques such as those in [2]. There are two essential motivations for parameterizing $\ln S(x,y)$ instead of $S(x,y)$ itself. The first is the need to preserve positivity of scattering function estimates. The second is that in the log domain, the estimation problem can be set up as the problem of restoring an image corrupted by additive non-Gaussian noise. By application of statistical hypothesis-testing principles a solution to this estimation problem can be derived under various noise/resolution tradeoffs.

2. Statistical Model

Denote by N the number of data, $\mathbf{U}$ a discretization of the (x,y) domain into a set of N points, L the linear transform (assumed to be invertible) that maps the discretized reflectivity onto the data r, and define $p(x,y) := |(L^{-1}r)(x,y)|^2$. In the radar community, $p(x,y)$ is referred to as the *preimage* and is often used as an estimator for $S(x,y)$. The preimage is a sufficient statistic for the ML estimation problem and is analogous to the classical periodogram in spectrum estimation [2]. It unfortunately exhibits poor statistical properties that follow directly from the statistical model for the reflectivity,

$$p(x,y) = S(x,y)\,u(x,y)\,, \qquad x,y \in \mathbf{U}\,, \tag{1}$$

where $\{u(x,y)\,,\ x,y \in \mathbf{U}\}$ are iid exponential random variables with unit mean. In the image processing terminology, $u(x,y)$ is a *speckle noise*. Model (1) fits a broad class of coherent imaging problems, as well as power-spectrum estimation problems. We use a logarithmic transform to map the multiplicative model (1) into an additive one,

$$\ln p(x,y) - \ln 2 - \gamma = \ln S(x,y) + \varepsilon(x,y)\,, \qquad x,y \in \mathbf{U}\,. \tag{2}$$

In (2), $\gamma \approx 0.57721$ is Euler's constant, and $\{\varepsilon(x,y) := \ln u(x,y) - \ln 2 - \gamma\}$ are zero-mean iid additive noise samples with pdf denoted by $p_\varepsilon(.)$. The log-likelihood for S is given by

$$l(S) = \sum_{x,y \in \mathbf{U}} \ln p_\varepsilon[\,\ln p(x,y) - \ln 2 - \gamma - \ln S(x,y)\,]\,.$$

The unconstrained ML estimator is simply the preimage and is unsatisfactory. In the following we show how wavelets can be used to achieve regularization of the estimates.

3. Wavelet Regularization

We adopt the following discrete orthonormal wavelet representation for $\ln S(x,y)$:

$$\ln S(x,y) = \sum_{j,k,l \in \Lambda} a_{jkl}\,\psi_{jkl}(x,y)\,, \qquad x,y \in \mathbf{U}\,, \tag{3}$$

where $\{\psi_{jkl}(x,y)\}_{j,k,l \in \Lambda}$ is a two-dimensional wavelet basis for $\mathbf{U}$, j and (k,l) are the scale and location parameters in the discrete index set Λ, respectively, and $\{a_{jkl}\}$ are the wavelet coefficients for $\ln S(x,y)$. Similarly, we introduce the wavelet coefficients $\{b_{jkl}\}$ and $\{e_{jkl}\}$ for the scaled log-preimage and for the additive noise ε in (2), respectively. From (2), we obtain $b_{jkl} = a_{jkl} + e_{jkl}\,,\ j,k,l \in \Lambda$, in which $\{e_{jkl}\}$ is interpreted as an additive noise corrupting the wavelet coefficients of $\ln S(x,y)$. The estimation problem consists in estimating $\{a_{jkl}\}$ given the transformed data (sufficient statistics) $\{b_{jkl}\}$. A possible scheme is proposed in [3] and outlined below.

Under a simple technical condition on the wavelet transform used, a good large-sample approximation consists in assuming that $\{e_{jkl}\}$ are iid Gaussian [3]. Then the log-likelihood for the wavelet coefficients of the scattering function can be maximized over each wavelet coefficient independently. If $\ln S(x,y)$ is smooth enough, the wavelet coefficients $\{a_{jkl}\}$ decay rapidly at fine scales [4]. This behavior is to be contrasted with that of the wavelet coefficients for the noise $\{e_{jkl}\}$, which have scale-independent variance. This property can be used to discriminate between signal and noise components of the observations. The significance of each wavelet coefficient can be tested by application of a likelihood ratio test (LRT). By application of this classical regression technique, only significant wavelet components of $\ln S(x,y)$, regardless of their scale, are retained in the regularized wavelet representation. Various noise/resolution tradeoffs can be obtained by selecting the significance level of the LRT appropriately. The complexity of the estimation algorithm is linear in the number of pixels of the image.

References

[1] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller, "The Use of Maximum-Likelihood Estimation for Forming Images of Diffuse Radar-Targets from Delay-Doppler Data," *IEEE Trans. on Info. Theory*, Vol. 35, No. 3, pp. 536-548, 1989.

[2] P. Moulin, J. A. O'Sullivan and D. L. Snyder, "A Method of Sieves for Multiresolution Spectrum Estimation and Radar Imaging", *IEEE Trans. on Info. Theory*, Special Issue on Wavelet Transforms and Multiresolution Analysis, Vol. 38, No. 2, pp. 801-813, 1992.

[3] P. Moulin, "A Wavelet Regularization Method for Diffuse-Target Radar Imaging and Speckle-Noise Reduction," to appear in *J. of Math. Imaging and Vision*, Special Issue on Wavelets, Jan. 1993.

[4] S. G. Mallat, "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbf{R})$," *Trans. Am. Math. Soc.*, Vol. 315, No. 1, pp. 69-87, 1989.

A Markov Random Field Product Model for Complex-Valued Radar Imagery

John D. Gorman and Brian J. Thelen

Environmental Research Institute of Michigan, Box 134001, Ann Arbor, MI 48113-4001

1 Summary

The zero-mean delta-correlated complex circular Gaussian random field model is commonly used as a spatial model for the joint statistics of the complex amplitudes of the pixels in a radar or coherent optical image [1, 3 and references therein]. Two deficiencies in the model are the lack of the ability to model heavy-tailed distributions and spatial correlation. Tails of the empirical distributions of real imagery are often heavier than those predicted by the Gaussian model. Real imagery can also exhibit spatial correlation, an effect that is not captured by the classical delta-correlated Gaussian model.

We introduce a generalization of the Gaussian speckle model called the *Markov random field (MRF) product model.* This model is formed as the pixel-by-pixel product between a nonnegative spatial random process $\underline{T}$, called the *texture process*, and a spatially-white process $\underline{S}$, called the *speckle process.* An essential property of this MRF product model is that it admits the heavy tails and spatial correlation seen in real imagery.

We propose a particular MRF product model in which the texture process $\underline{T}$ is represented by a *transformed Gaussian MRF (TGMRF)* [4]. The TGMRF is a nonnegative MRF generated through a one-parameter nonlinear transformation of a Gaussian MRF. We then discuss parameter estimation in the MRF product model using an alternative criterion based upon Csiszar's information divergence.

The MRF Product Model

Let $\mathbf{K}$ denote a set of pixel indices. We will denote the radar image by $\underline{Y}$, where $\underline{Y} = \{Y_k; k \in \mathbf{K}\}$ is the lexicographically-ordered vector of complex pixel amplitudes in the image. The model we propose is: $Y = S_k T_k, k \in \mathbf{K}$, where $\underline{S}$ is a spatially-white, zero-mean compiex circular gaussian process with identity covariance and $\underline{T}$ is constructed as follows.

Define the power-law transformation:

$$\psi_\lambda(y) \stackrel{\text{def}}{=} \begin{cases} (y^\lambda - 1)/\lambda & y > 0,\ \lambda > 0 \\ \log(y) & y > 0,\ \lambda = 0 \end{cases} . \qquad (1)$$

Equation (1) is typically referred to as the *Box-Cox transformation* [2]. The term $1/\lambda$ is included to ensure the continuity in λ of $\psi_\lambda(y)$ at $\lambda = 0$.

Suppose that there exists a Gaussian MRF $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$ such that

$$\underline{X} = \psi_\lambda(\underline{T}) \sim \mathcal{N}(\mu, \Sigma), \qquad (2)$$

where the transform ψ_λ is applied on a pixel-by-pixel basis. This latter relationship then specifies a spatial model for $\underline{T}$ that is a non-Gaussian Markov random field (but includes Gaussian model as a special case when $\lambda = 1$):

$$\underline{T} \stackrel{\text{def}}{=} \psi_\lambda^{-1}(\underline{X}). \qquad (3)$$

That $\underline{T}$ is Markov follows by noting that ψ_λ is a continuous one-to-one transformation from $\mathbf{R}_+$ to $\mathbf{R}$ and that $\underline{X}$ is a MRF.

Parameters of the product model to be estimated are the mean and covariance of the Gaussian MRF, μ and Σ, and the Box-Cox parameter λ. We will denote these parameters by $\underline{\theta}$.

Parameter Estimation Approach

Let $D(f_{\underline{S}} \| f_{\underline{S}|\underline{Y}})$ be the information divergence between $f_{\underline{S}}$ and $f_{\underline{S}|\underline{Y}}$. Application of this criterion for parameter estimation in the MRF product model is based upon the following heuristic. One can view the speckle process $\underline{S}$ as a "noise" process and $f_{\underline{S}|\underline{Y}}(\underline{s}|\underline{y};\underline{\hat{\theta}})$ as an estimate of the pdf of this noise process. Minimization of the divergence then is in some sense equivalent to choosing the parameter $\underline{\theta}$ for which the residual speckle term, as predicted by $f_{\underline{S}|\underline{Y}}$, is "white", e.g., is a *best match* our for our iid speckle model, $f_{\underline{S}}$.

Maximization of the following criterion:

$$\log f_{\underline{Y}}(\underline{y};\underline{\theta}) - D\left(f_{\underline{S}}(\underline{s}) \| f_{\underline{S}|\underline{Y}}(\underline{s}|\underline{y};\underline{\theta})\right) \qquad (4)$$

results in an alternative to the maximum-likelihood estimator (MLE) that is computationally simpler to evaluate than the MLE. This alternate criterion results in an estimator $\underline{\hat{\theta}}$ that simultaneously maximizes the likelihood function and minimizes the information divergence between $f_{\underline{S}}$ and $f_{\underline{S}|\underline{Y}}$. We investigate properties of this estimator both theoretically and via simulation.

References

[1] Dainty, J. C., (editor) *Laser Speckle and Related Phenomena*, Topics in Applied Physics, Vol. 9, Springer-Verlag, New York, 1975.

[2] Hernandez, F. and Johnson, R., "The large-sample behavior of tranformations to normality," *J. of American Stat. Ass.*, vol. 75, 855-861, 1980.

[3] Snyder, D. L., O'Sullivan, J. A., and Miller, M. I., "The Use of Maximum Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay-Doppler Data," *IEEE Trans. Info. Theory*, vol. 35, no. 3. pp 536-548, May, 1989.

[4] Thelen B. J., and Gorman, J. D., "A nonnegative MRF for modeling nonnegative imagery," *Proc. Conf. Info. Sci. and Sys.*, Princeton, NJ, March 18-20, 1992.

THE NORMALIZED SECOND MOMENT OF THE BINARY LATTICE DETERMINED BY A CONVOLUTIONAL CODE

A. R. Calderbank
P. C. Fishburn
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

The output of a finite state machine is a collection of codewords that can be searched efficiently to find the optimum codeword with respect to any nonnegative measure that can be calculated on a symbol by symbol basis. One recent application of this principle is the trellis coded quantization work of Marcellin and Fischer where the measure is mean squared error (m.s.e.). A second application, closely related to the latter, is the trellis shaping work of Forney. Trellis shaping is a sequence based technique for decreasing the average transmitted signal power in a communications system, and in this application, the measure is the power of an individual signal point. Both applications involve representing a source sequence x as the sum of a codeword c and an error sequence $e = (e_i)$. In quantization, the objective is the codeword c, and the expected value $E(e_i^2)$ is the mean squared error (per dimension). In trellis shaping the objective is the error sequence e. The signal constellation will be the error sequences e that result from a suitably chosen discrete set of source sequences x. Here the expected value $E(e_i^2)$ will determine the extent to which average transmitted signal power is reduced. This correspondence between vector quantization and the design of finite dimensional signal constellations is apparent in the work of Conway and Sloane. The extension of this correspondence to sequence based methods of quantization and shaping has been described by Forney.

We calculate the per-dimension mean squared error $\mu(S)$ of the 2-state convolutional code C with generator matrix $[1, 1 + D]$, for the symmetric binary source $S = \{0, 1\}$, and for the uniform source $S = [0, 1]$. When $S = \{0, 1\}$, the quantity $\mu(S)$ is the second moment of the coset weight distribution, which gives the expected Hamming distance of a random binary sequence from the code. When $S = [0, 1]$, the quantity $\mu(S)$ is the second moment of the Voronoi region of the modulo 2 binary lattice determined by C. The key observation is that a convolutional code with 2^ν states gives 2^ν approximations to a given source sequence, and these approximations do not differ very much. It is possible to calculate the steady state distribution for the differences in these path metrics, and hence the second moment. In this paper we shall only give details for the convolutional code $[1, 1 + D]$, but the method applies to arbitrary codes.

We also define the covering radius of a convolutional code, and calculate this quantity for the code $[1, 1 + D]$.

New Constructions of $k/(k+1)$ Rate-Variable Punctured Convolutional Codes

Pisit CharnKeitKong †, Kazuhiko YAMAGUCHI ‡, Hideki IMAI †††

†Faculty of Engineering, Yokohama National University.
‡Department of Computer Science and Information Mathematics, The University of Electro-Communications.
†††Center of Function-Oriented Electronics, Institute of Industrial Science, University of Tokyo.

Abstract: In our previous study, it was shown that good high-rate punctured convolutional codes(PCCs) in the class Ξ_f (Ξ_f-PCC) can be systematically searched. Leting $\mathbf{P}_\Xi = \{P_0, P_1, \cdots, P_{n-1}\}$ be a set of n different generators for a Ξ_f-PCC, we construct a rate-variable PCC by using only the generators in $\mathbf{P}_\Xi$. For constraint lengths 7, 8, and 9, we have found new good $k/(k+1)$ rate-variable PCC systems that provide good BER performance for k =1,2,$\cdots$,7.

Introduction

Punctured Convolutional Code(PCC)[3] is a class of high-rate convolutional codes obtained by periodically puncturing the outputs of a low-rate encoder. The Viterbi decoder of a PCC is much more simple than that of the usual high-rate codes. However, because of the lack of mathematical structures good high-rate PCCs could not be efficiently searched by a systematical algorithm. This problem was solved by introduction of Ξ_f[1]. The punctured convolutional encoder for a Ξ_f code, Ξ_f-PCC, can be obtained by a systematic search algorithm.

The conventional method to construct a rate-variable PCCs is to search good PCCs for different coding rates restricting n generators to those for an optimal rate $1/n$ code. In this paper, rather than using the generators of a low-rate code, we construct rate-variable PCC by using different generators for a good high-rate Ξ_f-PCC.

Background

Yamada et al.[2] proposed a maximum likelihood decoding technique, called the YHM algorithm. This is a breakthrough over the inherent difficulties in decoding of any high-rate convolutional code. The idea of YHM algorithm is to divide the trellis diagram of the syndrome-former of a rate $k/(k+1)$ code into $k+1$ stages such that there are only two branches or less entering each state.

Ξ is defined in [1] as a class of $(k+1, k, \nu)$ convolutional codes having $\eta_\nu \leq 2$, where η_ν is the number of polynomials $H^j(D)$ having $\deg[H^j(D)] = \nu$ in the parity-check $\mathbf{H}(D)$. Ξ codes can be efficiently decoded by YHM algorithm. In general, however, the trellis of a Ξ code has time-varying branch structure. Ξ_f is a particular class of Ξ codes that can be decoded by the fixed branch structure trellis. Since Ξ is defined by the parity-check matrix, it can be efficiently constructed by a systematic algorithm. Several good high-rate Ξ and Ξ_f codes of $d_{free} = 4, 5, \cdots, 8$ have been reported in [1].

In [1], it is pointed out that the trellis in YHM algorithm of a $(k+1, k, \nu)$ Ξ_f code is exactly the same as the trellis in Viterbi algorithm for a $(k+1, k, \nu)$ PCC. Hence, the punctured convolutional encoder for Ξ_f can be derived from the trellis of a Ξ_f code. The PCCs obtained by this method is called Ξ_f-PCC.

Code Search Results

Leting $\mathbf{P}_\Xi = \{P_0, P_1, \cdots, P_{n-1}\}$ be a set of n different generators for a Ξ_f-PCC, we construct a rate-variable PCC by using only the generators in $\mathbf{P}_\Xi$.

Limited searches are conducted on the set of codes having constraint lengths ν = 7,8 and 9 to construct $k/(k+1)$ rate-variable PCC that give good BER performance for $1 \leq k \leq 7$. For $\nu = 7$ and $\nu = 8$ partial searches were performed in the set of rate 6/7 Ξ_f-PCCs achieving $d_{free} = 5$ and $d_{free} = 6$, respectively. At $\nu = 9$ a partial search was performed in the set of rate 3/4 Ξ_f-PCCs achieving $d_{free} = 8$. d_{free} and the first five terms of weight spectra b_j of the obtained rate-variable PCCs systems are listed in Table 1. To compare, the parameters of the best known rate-variable PCC systems [3][4] are also listed in Table 1. The newly obtained rate-variable coding systems give moderate BER performance at low-rate. At high-rate, these new systems give significantly better BER performance than that of the best known rate-variable systems.

TABLE 1: RATE-VARIAVLE PCC SYSTEMS

Rate	code	d_{free}	$b_{d_{free}}, b_{d_{free}+1}, \cdots, b_{d_{free}+4}$
Constraint Length $\nu = 7$			
1/2	$(P_0, P_6)^{\S}$	10	10,0,73,0,687
	$(P_0, P_1)^{L}$	10	2,23,62,165,404
2/3	$(P_1, P_6), P_0^{\S}$	7	45,206,891,4076,18052
	$(P_0, P_1), P_0^{L}$	8	395,0,6695,0,135288
3/4	$(P_0, P_6), P_0, P_2^{\S}$	6	100,585,3839,24570,155815
	$(P_0, P_1), P_3, P_2^{L}$	6	67,651,4008,24638,153642
4/5	$(P_1, P_6), P_1, P_3, P_4^{\S}$	6	1899,0,130944,0,8065820
	$(P_0, P_1), P_2, P_2, P_1^{L}$	5	93,873,7017,59170,482219
5/6	$(P_0, P_2), P_6, P_1, P_3, P_5^{\S}$	5	329,3834,38819,385064,3716879
	$(P_1, P_2), P_3, P_3, P_1, P_1^{L}$	5	366,4287,4436,423337,4009089
6/7	$(P_5, P_6), P_0, P_1, P_2, P_3, P_4^{\S *}$	5	723,10310,123861,1459133,16602878
	$(P_0, P_1), P_1, P_2, P_3, P_2, P_1^{L}$	4	17,1008,12651,152171,1780906
7/8	$(P_0, P_4), P_0, P_6, P_2, P_5, P_3, P_4^{\S}$	4	39,1863,31388,444036,5963198
	$(P_0, P_1), P_3, P_3, P_3, P_0, P_0, P_1^{L}$	4	77,2122,32024,455479,6099937
Constraint Length $\nu = 8$			
1/2	$(P_1, P_4)^{\S}$	12	67,0,472,0,3363
	$(P_0, P_1)^{L}$	11	3,26,53,150,379
2/3	$(P_1, P_5), P_0^{\S}$	8	128,0,3490,0, 62693
	$(P_0, P_1), P_2^{L}$	8	109,0,2966,0,56458
3/4	$(P_0, P_4), P_2, P_1^{\S}$	6	27,610,3196,17838,110761
	$(P_0, P_3), P_3, P_1^{L}$	6	52,490,2902,17935,109020
4/5	$(P_1, P_5), P_5, P_6, P_3^{\S}$	6	749,0,61866,0,3838837
	$(P_0, P_3), P_0, P_0, P_1^{L}$	5	12,196,2413,20874,169543
5/6	$(P_0, P_4), P_2, P_6, P_5, P_2^{\S}$	6	2750,0,312103,0,28775304
	$(P_0, P_1), P_0, P_1, P_2, P_3^{L}$	5	152,1688,18092,182519,1778286
6/7	$(P_5, P_6), P_0, P_1, P_2, P_3, P_4^{\S *}$	6	9758,0,1437137,0,183683846
	$(P_0, P_2), P_3, P_3, P_0, P_0, P_1^{L}$	4	4,462,6229,73501,879798
7/8	$(P_0, P_3), P_4, P_1, P_0, P_5, P_6, P_2^{\S}$	5	931,14309,200512,2749601,35974195
	$(P_0, P_3), P_3, P_3, P_3, P_1, P_0, P_2^{L}$	4	13,1572,23200,317333,4268249
Constraint Length $\nu = 9$			
1/2	$(P_1, P_5)^{\S}$	12	20,0,170,0,1116
	$(P_0, P_1)^{H}$	12	14,26,74,257,496
2/3	$(P_0, P_3), P_3^{\S}$	8	54,0,1720,0,30595
	$(P_0, P_1), P_0^{H}$	7	3,70,207,836,4411
3/4	$(P_2, P_3), P_0, P_2^{\S *}$	8	2118,0,78546,0,2915853
	$(P_0, P_1), P_1, P_1^{H}$	6	38,270,1640,10554,63601
4/5	$(P_2, P_3), P_1, P_2, P_2^{\S}$	6	291,0,25235,0,16118071
	$(P_0, P_1), P_1, P_1, P_1^{H}$	4	6,3,298,2604,19132
5/6	$(P_0, P_3), P_0, P_3, P_1, P_3^{\S}$	6	2180,0,226105,0,20626620
	$(P_0, P_1), P_1, P_0, P_0, P_0^{H}$	5	201,2104,22183,217194,2041494
6/7	$(P_2, P_3), P_0, P_3, P_1, P_2, P_0^{\S}$	5	94,4027,36019,511214,5033081
	$(P_0, P_1), P_0, P_0, P_0, P_1, P_1^{H}$	5	267,5105,56285,656627,7433871
7/8	$(P_2, P_3), P_0, P_3, P_1, P_2, P_2, P_0^{\S}$	5	662,10320,150932,2033984,26512597
	$(P_0, P_1), P_1, P_0, P_1, P_1, P_0, P_0^{H}$	4	3,690,10528,150237,2007749

§: Best code selected from $\mathbf{P}_\Xi = \{P_0, P_1, P_2, P_3, P_4, P_5, P_6\}$={337,227,221 ,207,215,327,255}, {461,563,537,575,673,671,613}

§: Best code selected from $\mathbf{P}_\Xi = \{P_0, P_1, P_2, P_3\}$={1147,1317,1037,1725 }

L: Best code selected from code generator sets $\{P_0, P_1, P_2, P_3\}$={337,251,237,235}, {765, 473,463,457},found in [4].

H: Best code selected from code generator set $\{P_0, P_1\}$={1167,1545} found in[2]

*: Ξ_f-PCC.

References

[1] P. CharnKeitKong, K. Yamaguchi, H. Imai , "New high-rate convolutional codes and their decoding techniques," *Paper of Technical Group on Inform. Theory IEICE*. IT92-, May 1992.

[2] T. Yamada, H. Harashima, and H. Miyakawa, "A new maximum likelihood decoding of a high rate convolutional codes using a trellis,"(in Japanese), *Trans. IEICE*. **J66-A**, pp611-616, July. 1983.

[3] D. Haccoun, G. Bégin, "High-rate punctured convolutional codes for Viterbi and sequential decoding" *IEEE Trans. Commun.*, vol. 37, pp. 1113-1125, Nov. 1989.

[4] P. J. Lee, "Construction of rate (n-1)/n punctured convolutional codes with maximum required SNR criterion," *IEEE Trans. Commun.*, vol. 36, pp. 1171-1174, Oct. 1988.

A New Bound on the Row Distance of Rate 1/n Convolutional Codes *

Y. Levy
D.J. Costello, Jr.

Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Abstract

A new formula is derived to compute the codeword weights of rate 1/n convolutional codes. This formula allows us to derive a new asymptotic lower bound on the row distance that surprisingly reaches the upper bound on free distance in the limit of large memory m. This formula also leads to a new approach for constructing finite constraint length convolutional codes.

Summary

Costello [1] and Zigangirov and Massey [2] have derived lower and upper bounds on the free distance of fixed and time-varying convolutional codes by deriving bounds on ensemble averages. In this paper, we investigate a new approach to deriving an asymptotic lower bound on the row distance of order l of rate 1/n convolutional codes, i.e., the lowest weight of codewords generated by information sequences of length less than or equal to $l+1$. Although this bound is valid only asymptotically, it suggests that a similar bound might also be found for finite constraint lengths, and it leads to a new approach for constructing finite constraint length convolutional codes.

Suppose (x, y) are two elements from the binary field $F = \{0, 1\}$, $\oplus$ denotes addition in the binary field, and $+$ denotes addition in the integer field I. Then

$$x \oplus y = x + y - 2xy. \tag{1}$$

In order to derive our new formula on the row distance of rate $1/n$ convolutional codes, we need the following definitions:

Definition 1. *Let $\mathbf{a}(X)$ be a polynomial with coefficients a_i. Then, we define the k^{th} correlation coefficients as*

$$R_{k\mathbf{a}}(j_1, j_2, \ldots, j_k) = \sum_{i=0}^{i=\infty} a_i a_{i+j_1} \cdots a_{i+j_1+\ldots+j_k}, \tag{2}$$

where $j_1, j_2, \ldots, j_k$ are k integers strictly greater than 0.

Definition 2. *Let $g^{(1)}(X), g^{(2)}(X), \ldots, g^{(n)}(X)$ be the n degree m generator polynomials of a rate 1/n convolutional code C. Then, let $\mathbf{G}(X)$ be the composite generator polynomial*

$$\mathbf{G}(X) = g^{(1)}(X^n) + Xg^{(2)}(X^n) + \ldots + X^{n-1}g^{(n)}(X^n). \tag{3}$$

Then, for any information sequence $\mathbf{u}(X)$, the code sequence $\mathbf{v}(X)$ is generated by

$$\mathbf{v}(X) = \mathbf{u}(X^n)\mathbf{G}(X). \tag{4}$$

Using the previous definitions, we can obtain the following theorem on computing the row distance of rate 1/n convolutional codes.

Theorem 1. *Let C be a rate 1/n convolutional code with composite generator polynomial $\mathbf{G}(X)$. Then the row distance of order l of C can be computed as:*

$$d_l = \min_{\substack{\mathbf{u}(X)\neq 0 \\ \deg \mathbf{u}(X) \le l+1}} \Big(R_{0\mathbf{u}}R_{0\mathbf{G}} - 2\textstyle\sum_{j=0}^{j=\infty} R_{1\mathbf{u}}(j)R_{1\mathbf{G}}(nj) + 4\sum_{j,k=0}^{\infty} R_{2\mathbf{u}}(j,k)R_{2\mathbf{G}}(nk, nj) - \ldots \Big) \tag{5}$$

(5) gives a general formula for computing the row distance of order l of a rate 1/n convolutional code. This formula can be simplified when m goes to infinity. Specifically, let us construct our generator polynomial by randomly selecting its coefficients from F, that is:

$$\mathbf{G}(X) = \sum_{i=0}^{i=n(m+1)-1} g_i X^i, \tag{6}$$

where $g_i \in F = \{0, 1\}$ and $Pr(g_i = 0) = Pr(g_i = 1) = \frac{1}{2}$ for any integer $i \geq 0$. For these randomly constructed codes, the following theorem can be derived:

Theorem 2. *Let $\mathbf{G}(X)$ of degree $n(m+1)-1$ be the composite generator polynomial of a randomly constructed rate 1/n convolutional code with memory order m. Then, for any finite order l, with probability 1,*

$$\lim_{m\to\infty} \frac{d_l}{n(m+1)} = \frac{1}{2}. \tag{7}$$

Thus, by taking a random generator $\mathbf{G}(X)$, with probability 1 the row distance of any finite order l is on the order of $n(m+1)/2$ as m goes to infinity. Since there exists a large number of randomly generated codes, (7) represents a lower bound on the row distance of rate 1/n convolutional codes. This bound is significant since it implies that there exists codes for which the row distance of any finite order l reaches the asymptotic upper-bound on free distance derived by Costello [1]. Since the bound is only valid for $l < m$, however, it is not a bound on $d_{free} = lim_{l\to\infty} d_l$. But, for almost all known codes, the row distance reaches the free distance within the first constraint length, suggesting that it may be possible to strengthen existing lower bounds on the free distance.

Although the bound derived in Theorem 2 is only valid as m goes to infinity, it is possible to use Theorem 1 for constructing rate 1/n convolutional codes. We note that a code with large free distance requires the optimization of the functional given by (5). Thus, different algorithms for seeking optimization of a functional can be used to construct generators of rate 1/n convolutional codes. A large number of convolutional codes constructed in this way have free distances close to optimal codes, and the algorithms allow us to construct codes with much higher constraint lengths than previously constructed codes.

*This work was supported by NSF Grant NCR89-03429 and by NASA Grant NAG5-557.

References

[1] D.J. Costello, Jr., "Free Distance Bounds for Convolutional Codes," *IEEE Trans. Inf. Theory*, IT-20, No.3, pp.356-365, May 1974.

[2] K.Sh. Zigangirov and J.L. Massey, "Fixed convolutional codes achieve the same bounds as time-varying codes, even at small branch lengths", *Proc. Third. Soviet-Swedish Workshop on Information Theory*, Sochi, May 1987, pp. 335-338.

BLOCK CODE BASED ANALYSIS OF CONVOLUTIONAL CODES

Øyvind Ytrehus

University of Bergen, Department of Informatics, HiB, N-5020 Bergen, Norway. E-mail: Oyvind.Ytrehus@ii.uib.no
Supported by the Norwegian Research Council for Science and the Humanities (NAVF).

Abstract

Well known techniques from linear block codes analysis are used to study convolutional codes. It is demonstrated how these methods can be used to determine bounds on convolutional codes of certain parameters.

Introduction

An $(n, k, \nu, d_{\text{free}})$ code is a convolutional code of block length n, dimension k, constraint length ν, and free distance d_{free}. The code can be generated by a $k \times n$ matrix $\mathrm{G}(D)$ (an *encoder* matrix), in which each entry is a polynomial in D. The ith constraint length ν_i is the maximum degree of a polynomial of the ith row of $\mathrm{G}(D)$, and $\nu = \sum_{i=1}^{k} \nu_i$. Without loss of generality we can assume $\nu_i \leq \nu_{i+1}$ for $1 \leq i < k$. We call the vector $(\nu_1, \ldots, \nu_k)$ the constraint length vector.

In [1], $N(r, \nu, d_{\text{free}})$ is defined as the largest n such that an $(n, n-r, \nu, d_{\text{free}})$ code exists, and bounds on this function are developed.

A particularly useful upper bound on $N(r, \nu, d_{\text{free}})$, originally due to Heller [2], arises from the fact that a set of convolutional code words of bounded length is a linear block code, called a *terminated* code [3].

Lemma ([2]). *Let $\mathcal{C}$ be an (n, k, ν, d) convolutional code with constraint length vector $(\nu_1, \ldots, \nu_k)$. Then for all $j \geq 1$, $\mathcal{T}_j$ is a $[jn, \kappa(j), d]$ linear block code, where $\kappa(j) = \sum_{i:j>\nu_i} (j - \nu_i)$.*

In many cases the methods in [1] are insufficient to determine $N(r, \nu, d_{\text{free}})$ exactly. The purpose of this talk is to demonstrate that a detailed analysis of the terminated block codes sometimes provides either proofs of nonexistence of, or suggestions on how to construct, the associated convolutional codes. This analysis employs standard techniques from block code analysis.

An Example

In [1], an (8,6,5,6) code was obtained by computer search. Below follows a proof for the nonexistence of (9,7,5,6) codes.

Lemma 1. *Suppose $\mathcal{T}$ is an $[18, 9, 6]$ linear block code. Then the weight distribution of $\mathcal{T}$ is*

$$A_0 = A_{18} = 1, A_6 = A_{12} = 102, A_8 = A_{10} = 153. \qquad (1)$$

Proof. Using well-known techniques, it is possible to show that the minimum distance of $\mathcal{T}^{\perp} = 6$, and that no code word in $\mathcal{T}$ has weight 7. Then (1) is the only positive integer solution to the MacWilliams identities.

□

Theorem 2. *There is no (9,7,5,6) convolutional code.*

Proof. Suppose $\mathcal{C}$ is a (9,7,5,6) convolutional code, and that $\mathrm{G}(D)$ is a minimal encoder for $\mathcal{C}$. We can always assume that the constraint length vector is ordered. Hence, the only possible constraint length vector is $(0,0,1,1,1,1,1)$, otherwise the parameters of $\mathcal{T}_1$ are $[9, k(\geq 3), 6]$, but this is impossible by the Griesmer bound. This implies that $\mathcal{T}_1$ is a $[9,2,6]$ code, and it is easy to show that the upper left corner 2×9 submatrix of the binary encoder $\mathrm{G}(D)$ is equivalent (except for column permutations) to

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix};$$

thus $\mathcal{T}_1$ does not contain the all-one code word. Further, $\mathcal{T}_2$ is an $[18,9,6]$ code. From Lemma 1, it follows that $\mathcal{T}_2$ contains the all-one code word, so a sequence of suitable row operations will transform $\mathrm{G}(D)$ into another encoder $\mathrm{G}^*(D)$ which (i) has the same or smaller constraint length as $\mathrm{G}(D)$, and (ii) contains a row on the form $(1+D)(1,1,1,1,1,1,1,1,1)$. (i) implies that $\mathrm{G}^*(D)$ is also a minimal encoder. However, (ii) contradicts this (see, for instance, [4]).

□

In other cases, similar methods have made it possible to actually construct convolutional encoders of certain parameters. Thus, for instance, it is possible

- to construct a (9,5,4,8) code through analysis of [18,6,8] linear block codes, and
- to construct a (6,2,6,16) code through analysis of [30,4,16] linear block codes.

These results determine $N(r, \nu, d_{\text{free}})$ exactly in the respective cases.

References

[1] Ø. Ytrehus, "A note on high rate convolutional codes," Department report 68, Department of Informatics, University of Bergen, August 1992.

[2] J. A. Heller, "Sequential decoding: Short constraint length convolutional codes," space programs summary 37-54, Jet Propul. Lab., Calif. Inst. Tech., Pasadena, Dec. 1968.

[3] E. Paaske, "Short convolutional codes with maximal free distances for rates 2/3 and 3/4," *IEEE Trans. on Information Theory*, vol. IT-20, pp. 683–689, Sept. 1974.

[4] P. Piret, *Convolutional Codes - An Algebraic Approach*. The MIT Press, 1988.

COVERING PROPERTIES OF CONVOLUTIONAL CODES AND ASSOCIATED LATTICES

A. R. Calderbank
P. C. Fishburn
A. Rabinovich
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

This talk describes methods for analyzing the expected and worst-case performance of sequence based methods of quantization. We suppose that the quantization algorithm is dynamic programming, where the current step depends on a vector of path metrics, which we call a metric function. Our principal objective is a concise representation of these metric functions and the possible trajectories of the dynamic programming algorithm.

We shall consider quantization of equiprobable binary data using a convolutional code. Here the additive group of the code splits the set of metric functions into a finite collection of subsets. The subsets form the vertices of a directed graph, where edges are labelled by aggregate incremental increases in mean squared error (mse). Paths in this graph correspond both to trajectories of the Viterbi algorithm, and to cosets of the code. For the rate 1/2 convolutional code $[1 + D^2, 1 + D + D^2]$, this graph has only 9 vertices. In this case it is particularly simple to calculate per dimension expected and worst case mse, and performance is similar to the binary $[24, 12]$ Golay code.

Our methods also apply to quantization of arbitrary symmetric probability distributions on $[0, 1]$ using convolutional codes. For the uniform distribution on $[0, 1]$, the expected mse is the second moment of the "Voronoi region" of an infinite dimensional lattice determined by the convolutional code. It may also be interpreted as an increase in the reliability of a transmission scheme obtained by nonequiprobable signalling. For certain convolutional codes we obtain a formula for expected mse that depends only on the distribution of differences for a single pair of path metrics.

The Extended Invariant Factor Algorithm with Application to the Forney Analysis of Convolutional Codes

Robert J. McEliece and Ivan Onyszchuk
California Institute of Technology
Pasadena, California 91125, USA

Summary.

In his celebrated paper on the algebraic structure of convolutional codes, Forney [1] showed that by using the invariant-factor theorem, one can transform an arbitrary polynomial generator matrix for an (n,k) convolutional code C into a basic (and ultimately a minimal) generator matrix for C. He also showed how to find a polynomial inverse for a basic generator matrix for C, and a basic generator matrix for the dual code $C^\perp$. In this paper, we will discuss efficient ways to do all these things. Our main tool is the "extended invariant factor algorithm," which we introduce here.

1. The Extended Invariant Factor Algorithm.

The goal of the *invariant factor algorithm* (see e.g. [2, Sec. 6.2.4], [3, Sec. 6.3.3], or [4, Sec. 12.2]) is to take an arbitrary $k \times n$ matrix G (with $k \le n$) over a Euclidean domain R, and by a sequence of elementary row and column operations, to reduce G to a $k \times n$ diagonal matrix $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_k)$, whose diagonal entries are the invariant factors of G, i.e., $g_i = \Delta_i/\Delta_{i-1}$, where Δ_i is the gcd of the $i \times i$ minors of G. The goal of the *extended invariant factor algorithm*, which we introduce in this paper, is to take the same input, and not only find Γ, but also to find a $k \times k$ unimodular matrix X, and an $n \times n$ unimodular matrix Y, such that $XGY = \Gamma$.

To describe the extended invariant factor algorithm, we need to take a closer look at the original invariant factor algorithm. Formally, it can be described as follows. Beginning with the matrix $G_0 = G$, it produces a sequence of $k \times n$ matrices G_i, where G_{i+1} is derived from G_i by either an elementary row operation or an elementary column operation. We can represent this algebraically as

$$G_{i+1} = E_{i+1} G_i F_{i+1}, \tag{1.1}$$

where E_{i+1} and F_{i+1} are $k \times k$ and $n \times n$ elementary matrices, respectively. If G_{i+1} is obtained from G_i via a row operation, then $F_{i+1} = I_n$, but if G_{i+1} is obtained from G_i via a column operation, then $E_{i+1} = I_k$. After a finite number N of steps, we obtain $G_N = \Gamma$. (The details of which elementary row and column operations to perform, and in which order, are of central importance, of course, but for reasons of space, we refer the reader to [2, Sec. 6.2.4], or [3, Section 6.3.3] for them)

The extended invariant factor algorithm builds on the invariant factor algorithm. In addition to the sequence $G_0, G_1, \ldots, G_N$, the extended invariant factor algorithm also works with a sequence of unimodular $k \times k$ matrices $X_0, \ldots, X_N$, and a sequence of unimodular $n \times n$ matrices $Y_0, \ldots, Y_N$. The sequences (X_i) and (Y_i) are initialized as $X_0 = I_k$, $Y_0 = I_n$, and updated via the rule (cf. Eq.(1.1))

$$X_{i+1} = E_{i+1} X_i \tag{1.2}$$

$$Y_{i+1} = Y_i F_{i+1}. \tag{1.3}$$

It is a simple matter to prove by induction that

$$X_i G Y_i = G_i \qquad \text{for } i = 0, 1, \ldots, N, \tag{1.4}$$

so that specializing (1.4) with $i = N$, we have

$$X_N G Y_N = \Gamma, \tag{1.5}$$

which is the desired "invariant-factor" diagonalization of G. A rough analysis of this algorithm shows that it requires $O(dnk^2)$ polynomial divisions, or $O(d^3nk^2)$ field operations (addition, subtraction, multiplication, or division in F), where d denotes the maximum degree of any polynomial in G.

2. Application to the Analysis of Convolutional Codes.

We define an (n,k) convolutional code C over a field F to be a k-dimensional subspace of $F(D)^n$, where $F(D)$ is the field of rational functions in the indeterminate D over F. A *generator matrix* for C is a $k \times n$ matrix with entries in $F(D)$ whose rows form a basis for C. Given an arbitrary generator matrix G for C, we can easily transform G to a generator matrix with polynomial entries by multiplying the ith row of G by the lcm of the denominators of its components. In this section, we will see how the extended invariant factor algorithm introduced in Section 1 can be used to transform an arbitrary *polynomial* generator matrix for C into a *basic* generator matrix for C. (The transition from a basic to a minimal generator can, if desired, then be done by the simple algorithm originally described in [1], or perhaps more lucidly in Kailath [3, Sec. 6.3.2], where the process is described as "row-reducing" a polynomial matrix). We will see that the extended invariant factor algorithm also produces, more or less for free, a polynomial inverse for the basic generator matrix, and a basic generator matrix for the dual code $C^\perp$.

Assume then that G is a $k \times n$ polynomial generator matrix for a convolutional code C over a field F. Since the ring of polynomials over F is a Euclidean domain, we may apply the extended invariant factor algorithm described in Section 1, thereby obtaining a decomposition of the form (1.5). In what follows, the matrices X_N and Y_N produced by the extended invariant factor algorithm will be denoted simply by X and Y.

The matrices, X, Y, and Γ, contain much valuable information about the code C and the generator matrix G. To extract this information, however, we need to define several useful "pieces" of these matrices, which we call Γ_k, Γ'_k, K, and H:

$$\Gamma_k = \text{leftmost } k \text{ columns of } \Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_k). \tag{2.1}$$

$$\Gamma'_k = \gamma_k \cdot \Gamma_k^{-1} = \mathrm{diag}(\gamma_k/\gamma_1, \ldots, \gamma_k/\gamma_k). \tag{2.2}$$

$$K^T = \text{leftmost } k \text{ columns of } Y. \tag{2.3}$$

$$H^T = \text{rightmost } n-k \text{ columns of } Y. \tag{2.4}$$

Here then are useful "outputs" of the extended invariant factor algorithm, when applied to G.

- *A basic generator matrix for C:* $G_{\text{basic}} = \Gamma_k^{-1} X G$. (That is, G_{basic} is obtained by dividing the ith row of XG by the invariant factor γ_i, for $i = 1, \ldots, k$.)
- *A polynomial inverse for G_{basic} :* K^T
- *A polynomial pseudo-inverse for G, with factor γ_k:* $K^T \Gamma'_k X$.
- *A basic generator matrix for $C^\perp$, i.e., parity-check matrix for C:* H.

References.

[1] Forney, G. D., "Convolutional Codes I: Algebraic Structure." *IEEE Trans. Inform. Theory vol. IT-16 (November 1970), pp. 720–738.*

[2] Gantmacher, F. R., *The Theory of Matrices, vol. I.* New York: Chelsea Publishing Co., 1977.

[3] Kailath, T. *Linear Systems.* Englewood Cliffs, N. J.: Prentice-Hall, 1980.

[4] van der Waerden, B. L., *Algebra, vol. 2.* New York: Frederick Ungar, 1970.

Acknowledgements.

The contibution of Ivan Onyszchuk, and a portion of the contribution of Robert J. McEliece, to this paper, was carried out at Caltech's Jet Populsion Laboratory, under contract with the National Aeronautics and Space Administration. A portion of McEliece's contribution was also carried out at Caltech's Electrical Engineering department, and supported by AFOSR grant no. 91-0037

THE PERFORMANCE OF CONVOLUTIONAL CODES ON THE BLOCK ERASURE CHANNEL WITH VARIOUS FINITE INTERLEAVERS

Amos Lapidoth
Technion—Israel Institute of Technology and
Stanford University, Information Systems Laboratory, Stanford, CA 94305-4055

Abstract

Consider the transmission of a finitely interleaved rate $\frac{1}{n}$ convolutionally encoded message over a non-memoryless channel having two internal states Ξ_0 and Ξ_1 where, when in state Ξ_0, the channel resembles a noiseless Binary Symmetric Channel (BSC), whereas when in state Ξ_1, the channel is totally blocked and is well approximated by a Binary-Input-Single-Output channel. Assume that the channel's internal state is drawn at random once every h channel uses, and then remains constant for the following h channel uses. Further assume that the message is short in comparison to h, and that due to delay constraints, the message must be decoded within Nh channel uses, where N need *not* be large in comparison to the code's constraint length.

The probability of a message error, the normalized expected number of bits in error, and the Bit Error Rate (BER) are analytically computed for the periodic $N \times h$ chip and word interleavers, where a chip refers to a binary code symbol, and a word refers to a n-tuple of consecutive chips.

An analytic expression for the BER is also given for pseudo-random word and chip interleavers and for the corresponding limiting cases of infinite interleaving *i.e* $N \to \infty$.

Summary

Let $\mathcal{U} = (U_0, \ldots, U_{N-1})$ denote an *erasure pattern* i.e. an array of N elements $U_p \in \{0,1\}$ $0 \le p \le N-1$. Let $\mathcal{C} = (C_0, \ldots, C_{N-1})$ be an array of N channels where for each $0 \le p \le N-1$ the channel C_p is a noiseless binary-input binary-output channel (*"noiseless"*) if $U_p = 0$ and otherwise, if $U_p = 1$, C_p is *"erased"* i.e. a binary-input single-output channel. Thus, if $U_p = 0$ the channel transition probabilities satisfy $P(1|1) = P(0|0) = 1$, and otherwise, if $U_p = 1$ we have $P(\text{"?"}|0) = P(\text{"?"}|1) = 1$.

Consider the transmission of an L-length binary message over the array of channels $\mathcal{C}$ using a constraint length K, rate $\frac{1}{n}$ convolutional code with zero padding. To fix notations, let $D = (d_0, \ldots, d_{L-1})$ $d_j \in$ GF(2) be the message of length L which is produced by the source. Using the GF(2) arithmetic, the encoder's output can be written as

$$c_{jl} = \sum_{\nu=0}^{K-1} g_\nu^{(l)} \tilde{d}_{j-\nu} \quad j = 0, \ldots, L+K-2, \quad l = 1, \ldots, n, \qquad (1)$$

where $\tilde{d}_j = d_j$ unless $j < 0$ or $j \ge L$ in which case $\tilde{d}_j = 0$, and where $g_\nu^{(l)}$ $\nu = 0, \ldots, K-1$ are the coefficients of the l-th generating polynomial of the convolutional code. We shall use the term *word* for an n-tuple $(c_{j1}, \ldots, c_{jn})$ for some $j \in \{0, \ldots, L+K-2\}$. Similarly we shall use the term *chip* for a binary code symbol c_{jl} for some fixed $j \in \{0, \ldots, L+K-2\}$ and $l \in \{1, \ldots, n\}$.

A *periodic chip (word) interleaver* transmits the chip c_{jl} via the channel C_p where $p = n(j-1) + l \bmod N$ (resp. $p = j \bmod N$). A *random chip interleaver* transmits c_{jl} via a channel which is selected at random uniformly from $\mathcal{C}$. A *random word interleaver* ensures that all chips of a common word are transmitted via the same channel (which is selected at random).

We consider a receiver which after de-interleaving uses maximum likelihood sequence estimation (MLSE), *i.e.* Viterbi decoding, to estimate the transmitted message. We assume that ties are resolved randomly. Denote the decoder's estimate of the message by $\hat{d}_j$ $j = 0, \ldots, L-1$.

We show how given any erasure pattern $\mathcal{U}$, one can compute the probability of a message error,

$$P_{\text{MSG}}(\mathcal{U}) = \Pr\{\exists 0 \le j \le L-1 \text{ s.t. } \hat{d}_j \ne d_j\}, \qquad (2)$$

the expected number of bits in error normalized by the message length,

$$P_L(\mathcal{U}) = \frac{1}{L} \sum_{j=0}^{L-1} \Pr\{\hat{d}_j \ne d_j\}, \qquad (3)$$

and the bit error rate

$$\text{BER}(\mathcal{U}) = \lim_{L \to \infty} P_L(\mathcal{U}), \qquad (4)$$

for the periodic word and chip interleavers.

In practical applications the erasure pattern $\mathcal{U}$ is a random variable and one is then interested in the weighted averages of (2) (3) or (4) over all 2^N erasure patterns or in some other availability criterion which can be similarly computed.

The asymptotic bit error rates (4) for chip and word interleavers in the limit of infinite interleaving depth, *i.e.* as $N \to \infty$ are also computed. Notice that in the limit of infinite interleaving (periodic or random), the channel resembles a descrete memoryless erasure channel. Our approach to the analysis of this case has benefitted from the work of Burnashev and Cohn on the performance of convolutional codes on the BSC[1]. The error rate associated with nonideal *random* chip and word interleavers is also found.

Some numeric examples for scenarios based on the European cellular phone system (GSM) [2], are provided.

The analysis of finite messages transmitted using deterministic interleavers is mostly combinatoric in nature (once an erasure pattern has been fixed). For this situation we give a set of linear recursion equations which enable the computation of the probability of a message error and the expected number of bits in error. The study of the limit of the normalized expected number of bits in error (BER) for this situation involves some algebra.

References

[1] M. V. Burnashev, and D. L. Cohn, "Symbol error probability for convolutional codes," *Problemy Peredači Informacii*, vol. 26, No. 4, pp. 3-15, 1990.

[2] GSM Recommendations series 05, especially 05.03 .

Using a Modified Transfer Function to Calculate Unequal Error Protection Capabilities of Convolutional Codes *

D.G. Mills and D.J. Costello, Jr.

Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Abstract

This paper proposes a modified transfer function analysis that yields the individual bit error probability for any specified input bit position of an (n,k,m) convolutional encoder. The method is useful for analyzing the unequal error protection (UEP) capabilities of codes.

Summary

Unequal error protection (UEP) codes are of interest in several environments, e.g. packet switched networks and multi-user environments. A modified transfer function is now described that can be used for analyzing the UEP capabilities of convolutional codes.

We employ the method of determining a transfer function from an augmented state diagram [1]. To modify the augmented state diagram so that the individual bit error probabilities are determined, each branch is assigned the new label $X^i Y_1^{j_1} Y_2^{j_2} \cdots Y_k^{j_k}$, where j_k is equal to the input bit in the k^{th} position, and i is the Hamming weight of the branch output. Obviously, the sum of the j_k's is the Hamming weight of the input vector. Mason's gain formula is then applied. The resulting UEP transfer function is

$$T(X, Y_1, \cdots, Y_k) = \sum_{d=dfree}^{\infty} \sum_{j=0}^{j_d} C_{d,j} X^d Y_1^{b_{1,j}} \cdots Y_k^{b_{k,j}},$$

where $C_{d,j}$ is the number of paths associated with the j^{th} input sequence distribution of 1's that generates code vectors of weight d, j_d is the number of distinct input sequence distributions that generate code vectors of weight d, and $b_{1,j}, \cdots, b_{k,j}$ represents a particular input sequence distribution of 1's. The bound for the individual bit error probabilities is then

$$P_b^{(i)}(E) < \sum_d B_d^{(i)} P_d, 1 \le i \le k,$$

where $P_b^{(i)}$ is the probability that a bit located in the i^{th} position of the input vector is decoded incorrectly, $B_d^{(i)} = \sum_{j=0}^{j_d} b_{i,j} C_{d,j}$ is the total number of 1's in bit position i contained in all input vectors that generate code vectors of weight d, and $P_d = 2^d [p(1-p)]^{\frac{d}{2}}$. (For simplicity, we assume a binary symmetric channel with crossover probability p.)

The modified state diagram for a particular (3,2,1) code is shown in Figure 1. The generator vectors, the UEP transfer function, and the bit error bounds are shown in the first entry in Table 1.

Results for a number of other codes are also presented in Table 1. It can be seen that several factors affect the bit error probability for a specific input position. The first term of the error probability expression is the dominant term. Obviously, different exponents in the first terms result in large differences in the error protection given to the input bits. The lowest distance in the i^{th} individual bit error bound is defined to be the effective free distance, $d_{eff(i)}$. For example, for the (4,2,3) code in Table 1, $d_{eff(0)} = 6$ and $d_{eff(1)} = 4$. The effective free distance for the i^{th} position is the lowest Hamming weight among all code vectors that are generated by input sequences with at least one 1 in the i^{th} position. The individual effective free distances are lower bounded by the overall free distance, d_{free}.

In addition to the individual effective free distances, two other important factors affecting $P_b^{(i)}(E)$ are the number of low weight code vectors and the number of 1's in position i that belong to input vectors corresponding to the low weight code vectors. That is, in addition to the traditionally important minimum codeword Hamming weight and multiplicity, the distribution of 1's in the input vectors is important. The number of 1's in a particular position is related to the length and to the Hamming weight of the entire input vectors, but the exact relationship has not been completely determined.

Using the insights gained from this new UEP analysis technique, we can design new codes with different individual effective free distances, and therefore, different levels of unequal error protection.

Reference

[1] S. Lin and D.J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall,Inc., 1983.

(n,k,m) and generator vectors	K	UEP Transfer Function	$P_b^{(i)}$
(3,2,1) {3} $g_0^{(0)}=3$ $g_0^{(1)}=1$ $g_0^{(2)}=3$ $g_1^{(0)}=1$ $g_1^{(1)}=2$ $g_1^{(2)}=2$	2	$X^3(Y_1+Y_0Y_1^2)$ $+X^4(Y_1^2+Y_0^2Y_1^2+Y_0^3Y_1^3+Y_0Y_1^3+Y_0Y_1)$ $+\cdots$	$P_b^{(0)} \le 3P_3+7P_4$ $P_b^{(1)} \le 3P_3+11P_4$
(3,2,3) {3} $g_0^{(0)}=15$ $g_0^{(1)}=06$ $g_0^{(2)}=15$ $g_1^{(0)}=06$ $g_1^{(1)}=15$ $g_1^{(2)}=17$	6	$X^3(Y_1^2+Y_0^2Y_1^3+Y_1^3+Y_0^4Y_1^4+Y_0^2Y_1^4$ $+Y_0^3Y_1^3+Y_0^5Y_1^6+2Y_0^3Y_1^4+Y_1^4+2Y_0^2Y_1$ $+Y_0^2+Y_0^3Y_1^2+Y_0^3Y_1^3+Y_0Y_1^2)+\cdots$	$P_b^{(0)} \le 29P_7$ $P_b^{(1)} \le 55P_7$
(6,2,1) {1} $g_0^{(0)}=2$ $g_0^{(1)}=2$ $g_0^{(2)}=2$ $g_1^{(0)}=2$ $g_1^{(1)}=3$ $g_1^{(2)}=3$ $g_0^{(3)}=2$ $g_0^{(4)}=2$ $g_0^{(5)}=2$ $g_1^{(3)}=2$ $g_1^{(4)}=3$ $g_1^{(5)}=3$	1	$X^4(Y_1+2Y_0Y_1+Y_0+Y_0^2Y_1)$ $+X^7(Y_0Y_1+2Y_0^2Y_1+Y_0^2Y_1)+\cdots$	$P_b^{(0)} \le 5P_6+8P_8$ $P_b^{(1)} \le 4P_6+4P_8$
(4,2,3) $g_0^{(0)}=15$ $g_0^{(1)}=17$ $g_0^{(2)}=00$ $g_1^{(0)}=00$ $g_1^{(1)}=10$ $g_1^{(2)}=14$ $g_0^{(3)}=00$ $g_1^{(3)}=04$	4	$X^4Y_1+2X^5Y_0Y_1+\cdots$	$P_b^{(0)} \le 2P_6+\cdots$ $P_b^{(1)} \le P_4+\cdots$

Table 1

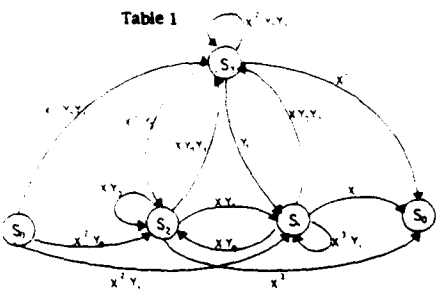


Figure 1

*This work was supported by NSF Grant EID90-17558, NSF Grant NCR89-03429, and NASA Grant NAG5-557.

THE MACWILLIAMS-SLOANE CONJECTURE ON THE TIGHTNESS OF THE CARLITZ-UCHIYAMA BOUND AND THE WEIGHTS OF DUALS OF BCH CODES

Oscar Moreno
University of Puerto Rico, Rio Piedras, PR 00931 USA
(O_MORENO@UPRENET.BITNET, MORENO@SUN386-GAUSS.UPRR.PR)

Carlos J. Moreno
Baruch, CUNY, Box 545, N. Salem, NY 10560 USA
(CARLOS@KRONECKER.BARUCH.CUNY.EDU)

Research Problem 9.5 of MacWilliams and Sloane's book *The Theory of Error Correcting Codes* asks for an improvement of the minimum distance bound of the duals of BCH codes, defined over $\mathbf{F}_{2^m}$, m odd. The objective of this talk is to give a solution to the above problem by: (i) obtaining an improvement to the Ax theorem, that we prove is best possible for many classes of examples, (ii) establishing a sharp estimate for the relevant exponential sums which implies a very good improvement for the minimum distance bound, (iii) providing a doubly infinite family of counterexamples to Problem 9.5 where both the designed distance and the length increase independently, (iv) verifying that our bound is tight for some of the counterexamples, and (v) in the case of even m we give a doubly infinite family of examples where the Carlitz-Uchiyama bound is tight, and in this way determine the exact minimum distance of the duals of the corresponding BCH codes.

More specifically we have the following results:

Theorem A: Let $F(x_1, x_2, \ldots, x_n)$ be a polynomial in n variables with coefficients in $\mathbf{F}_q$, $q = 2^m$. Let $\sigma(d)$ be the binary weight of d and

$$s = \max_{(d_1,\ldots,d_n)} \{\sigma(d_1) + \cdots + \sigma(d_n)\},$$

where the maximum is taken over the degrees of all the monomials in F. We then have that the exponential sum

$$S(F) = \sum_{x_1,\ldots,x_n \in \mathbf{F}_q} (-1)^{Tr(F(x_1,\ldots,x_n))}$$

is divisible by 2^b, where $b = \lceil mn/s \rceil$ is the smallest integer $\geq mn/s$.

Remark: The above is an improvement to a theorem of Adolphson-Sperber, where the conclusion is that $S(F)$ is divisible by $2^{\lceil mn/r \rceil}$ where r is the degree of F. To compare with our result in a concrete example, take a polynomial with 54 variables over a finite field with $q = 8$ elements, and assume that there is a term $x_1^{17}x_2^5$ with degree 22 and weight 4, and suppose there is no other term with $s > 4$. Then from theorem 2 we obtain that $2^{\lceil 3\cdot 54/4 \rceil} = 2^{41}$ divides $S(F)$ and in Adolphson-Sperber we get divisibility by $2^{\lceil 3\cdot 54/22 \rceil} = 2^8$, which is certainly smaller.

Theorem B: Let $q = 2^m$, and let

$$f(x) = \sum_j a_j x^{d_j}$$

be a polynomial with coefficients in $\mathbf{F}_q$. Suppose the maximum binary weight of the exponents is

$$t = \max_j \{\sigma(d_j)\}.$$

Let a be the smallest positive integer $\geq m/t$. We then have

$$\left| \sum_{x \in \mathbf{F}_q} (-1)^{Tr(f(x))} \right| \leq \frac{(\deg f - 1)}{2} 2^{a-1} \left[2^{2-a} \sqrt{q} \right].$$

Theorem C: Let $q = 2^{3m}$, m odd and let $f(x)$ be a polynomial, with coefficients in $\mathbf{F}_q$, of degree r, with r equal to 7 or 9. We then have

$$\left| \sum_{x \in \mathbf{F}_q} (-1)^{Tr(f(x))} \right| \leq (r-1) \cdot 2^{m-1} \left[2^{1-m} \sqrt{q} \right].$$

Remark: The inequality in the above theorem is tight for $q = 2^9$ and $q = 2^{39}$.

Our doubly infinite family of counterexamples to the MacWilliams and Sloane question is given by the following result.

Theorem D: For each prime p for which 2 is of order odd exactly $(p-1)/2$, let $q = 2^{((p-1)/2)m}$. Let $\epsilon > 0$. Then for infinitely many odd m, as well as for infinitely many even m, we have

$$\left| \sum_{x \in \mathbf{F}_q} (-1)^{Tr(x^p)} \right| \geq (p-1)\sqrt{q}(1-\epsilon).$$

The exact minimum distance of several classes of BCH codes have been previously computed, but as far as we know, no exact computation has been done in the case of their duals. We have the following result in that respect:

Theorem E: Let $\ell \mid 2^a + 1$ and let us further assume that a is the least integer with this property. Then for any b we have:

$$\sum_{x \in \mathbf{F}_{2^{2ab}}} (-1)^{Tr(x^\ell)} = (-1)^{b+1} 2^{ab} (\ell - 1)$$

Corollary 1: Polynomials x^ℓ for $\ell \mid 2^a + 1$ and for the pairs (a, b) provide a doubly infinite family of examples for which the Carlitz-Uchiyama bound is tight over fields which are an even power of 2, and of the form $\mathbf{F}_{2^{2ab}}$.

Corollary 2: The dual of the BCH code with designed distance $\ell + 2$ where $\ell \mid 2^a + 1$, and for any odd b, has minimum distance exactly $2^{2ab-1} - (\ell - 1)2^{ab-1}$ over the finite field $\mathbf{F}_{2^{2ab}}$.

The fundamental theorem of Chevalley-Warning has gone through several improvements in the work of Ax, Katz, Mazur and Adolphson-Sperber. What evolves in their work is the role played by the degree of equations. It is remarkable that the techniques developed to solve the Research Problem 9.5 of MacWilliams-Sloane, a deep question which reflects the behavior of the weights of BCH codes, have provided a new insight into the important role played by the p-adic weight of the degrees in the study of the divisibility properties of the number of solutions of a system of equations.

Our p-adic version of Serre's archimedean bound for the sum of the roots of an L-function and the improvement of the theorem of Ax are ample proof of the utility of the new techniques, both in coding theory and number theory.

Since the results in this paper are important to mathematicians and were previously unsuspected by them, they are another example where the theory of error correcting codes has been influential in the development of the new mathematical insights.

Coset weight enumerators of three Conway-Pless extremal self-dual binary codes of length 32

Paul Camion* Bernard Courteau†
André Montpetit†

*INRIA, Rocquencourt, BP. 105, 78153 Le Chesnay Cedex, France
†Département de mathématiques et d'informatique, Université de Sherbrooke, Sherbrooke (Québec) J1K 2R1.This work has been supported by CRSNG grant no. A5120

1 Introduction

Conway and Pless have enumerated in [6] the 85 non-equivalent self-dual doubly-even codes of length 32. From these, the five codes having minimum distance equal to 8 are called extremal. Two extremal codes were already known: the second order Reed-Muller code RM32 and the extended quadratic residue code QR32 of length 32. The other three discovered by Conway and Pless were new. In [8], Koch has given another more direct construction of the three Conway-Pless codes denoted by him F, U and G. Since the five extremal codes, though non-equivalent, have the same weight enumerator they also have the same classical parameters and it is natural to ask for some parameters which may distinguish them.

In [3] we have introduced a new parameter, the regularity number $\overline{r}$ of a code, which is related to other fundamental parameters [7] by the inequalities $\epsilon \leq \rho \leq t \leq \gamma \leq \overline{r}$ where ϵ is the error correcting capacity, ρ the covering radius, t the external distance (in the linear case t is the number of non-zero weight of the dual code) and γ is the number of distinct proper coset weight enumerators of the given code. In [4, 2, 9] we have developed theoretical tools based on the notion of partition design that permit us to calculate (in principle) the coset weight enumerators once a subgroup of the automorphism group of the code is given, more precisely when the orbit space of this group is computable. The situation appeared for example in [5] where the necessary notions and theorems are stated.

In this work we obtain the coset weight enumerators, and the parameters γ and $\overline{r}$ for the Reed Muller code RM32, the quadratic residue code QR32 of length 32 and the Conway Pless code $16f_2$ denoted by G in [8]. The coset weight enumerators for QR32 have already been calculated by Assmus and Pless in [1].

2 Combinatorial matrix and partition designs

In [4, 2, 3] we have introduced the *combinatorial matrix* A of a code C which is related to the distance distribution matrix B [7] (having as rows the coset weight enumerators of C) by the equality $A = BS$ where S is an easily computed nonsingular triangular matrix related to the Krawtchouk matrix. So the coset weight enumerators are easily computable once we know the combinatorial matrix A.

We have also introduced the concept of partition design admitted by a code. In the case where C is a binary linear code of length n and dimension $n-k$, let Ω be the set of columns of a parity check matrix H of C. Then a partition $\pi = \{\Omega_0, \Omega_1, \ldots, \Omega_s\}$ of the syndrome space $\mathbb{F}_2^*$ is said to be a *r partition design* admitted by the code C if $\Omega_0 = \{0\}$, $\Omega_1 = \Omega$ and if for $u, v \in \{0, 1, \ldots, r\}$ $m_{uv} = \mathrm{card}((h - \Omega) \cap \Omega_v)$ is a constant for all $h \in \Omega_u$. The matrix $M = (m_{uv})$ is said to be the *associate matrix* of π. The least possible number r such that C admits a r-partition design is called the *regularity number* of C, and the associate matrix of this minimal partition is called the *regularity matrix* of the code C.This matrix doesn't depend on the choice of the parity check matrix.

One important example of a partition design admitted by a code is the set of orbits under any subgroup G of the automorphism group of the code acting on the syndrome space.

We have also proved in an earlier work that the combinatorial matrix A of a code C is completely determined by *any* partition design admitted by the code via a linear recurrence relation. In the theoretical part of this work, we give an algorithm that computes the regularity matrix of a code from the knowledge of any given partition design admitted by the code.

3 Results

To apply the above theory, we take a permutation group $G \subseteq S_n$ letting the code $C = \ker H$ invariant and we let it act on the syndrome space $\mathbb{F}_2^*$ as follows.If $\sigma \in G$ and $h \in \mathbb{F}_2^*$, let x be any element in $\mathbb{F}_2^n$ such that $h = Hx^T$ is the syndrome of x. Then the element $\sigma(h) = H(\sigma x)^T$ where $\sigma x = \sigma(x_1, \ldots, x_n) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ is uniquely determined by h.

We have written programs in the computer algebra system Maple that implement the above theorems.To obtain the orbit partition designs for the extended quadratic residue code and the second order Reed-Muller code of length 32, we have taken the full automorphism groups $PSL_2(31)$ and $GA(5,2)$ respectively. For the code $G = 16f_2$ we have taken a subgroup of its automorphism group generated by 12 automorphisms fixing the set of glue components [6, p. 49] which has given 316 orbits in the syndrome space. Then we have determined the combinatorial matrix A and, by solving a triangular linear system, the distance matrix B, thus obtaining the coset weight enumerators. Finally, applying our last algorithm, we have obtained the regularity matrices and the regularity numbers of the considered codes. We have observed that all enumerators of the cosets of weight 1,2,3,5 and 6 are the same for the three codes.

References

[1] E.F. Assmus, V. Pless, *On the covering radius of extremal self-dual codes*, IEEE Trans. on Inform. Theory, **29** (3) (1983) 359–363.

[2] P. Camion, B. Courteau, P. Delsarte, *On r-partition designs in Hamming spaces*, Tech. Rep. 626, INRIA, 1987.

[3] P. Camion, B. Courteau and P. Delsarte, *On r-partition designs in Hamming spaces*, Applicable Algebra in Engineering, Commun. and Comput., **2** (1992) 147–162.

[4] P. Camion, B. Courteau, G. Fournier and V. S. Kanetkar, *Weight distributions of Translates of linear codes and generalized Pless identities*, J. Inform. Optim. Sci. **8** (1987) 1–23.

[5] P. Camion, B. Courteau and A. Montpetit, *Weight distribution of cosets of 2-error-correcting binary BCH codes of length* 15, 63 *and* 255, IEEE Trans. on Information Theory **38** (1992).

[6] J.N. Conway, V. Pless, *On the enumeration of self-dual codes*, J. Combin. Theory, Ser. A **28** (1980) 26–53.

[7] P. Delsarte, *Four fundamental parameters of a code and their combinatorial significance*,Inform. and Control,**23** (1973) 407–438.

[8] H. Koch, *On self-dual, doubly even codes of length 32*, J. Combin. Theory, Ser. A **51** (1989) 63–76.

[9] A. Montpetit, *Codes dans les graphes réguliers*, Ph. D. thesis, Université de Sherbrooke, Canada, 1987.

Weight hierarchies of binary linear codes of dimension 4

Torleiv Kløve, *Department of Informatics, University of Bergen, Høgteknologisenteret, N-5020 Bergen, Norway*

For any linear code D, $Supp(D)$, the *support of* D, is the set of positions where not all the codewords of D are zero. Let $w_S(D)$, the *support weight* of D, be the size of $Supp(D)$. For an $[n,k]$ code C and any r, where $0 \leq r \leq k$, the *r-th minimum support weight* (also known as the r-th generalized Hamming weight) is defined by

$$d_r(C) = \min\{w_S(D) \mid D \text{ is an } [n,r] \text{ subcode of } C\}.$$

The *weight hierarchy* of C is the set $\{d_1(C), d_2(C), \cdots, d_k(C)\}$.

For $k \leq 4$ we give explicit description of the possible weight hierarchies.

MACWILLIAMS IDENTITIES AND COORDINATE PARTITIONS

Juriaan Simonis
Delft University of Technology
Faculty of Technical Mathematics and Informatics
P.O. Box 5031, 2600 GA Delft, The Netherlands

Any partition of the coordinate set of a binary linear code is shown to correspond to a set of generalized MacWilliams identities. Thus, a well–chosen partition yields a promising method to settle existence and uniqueness problems. A short proof of a generalization of the Assmus–Mattson theorem is given. In the nonlinear case, a generalization of the Delsarte inequalities is obtained.

The main result

Coordinate partitions

Let C be a binary linear code, with coordinate set S, and let $\mathcal{T} := \{T_1, T_2, \ldots, T_p\}$ be a partition of S into sets T_u of size $n_u := |T_u|$. Then the *weight distribution* $\underline{A}(\mathcal{T})$ of the code C *with respect to the partition* $\mathcal{T}$ is the set of nonnegative integers

$$A_{\underline{i}}(\mathcal{T}) := |\{X \in C \mid |X \cap T_u| = i_u \ \forall u\}|.$$

The generalized MacWilliams identities

We show that for each index vector $\underline{i}$ the weight distribution $\underline{A}(\mathcal{T})\}$ of C and the weight distribution $\underline{B}(\mathcal{T})$ of the dual code $C^{\perp}$ satisfy the equation

$$A_{\underline{i}}(\mathcal{T}) = 2^{k-n} \Sigma_{\underline{j}} \left(\prod_{u=1}^{p} P_{i_u}(j_u; n_u) \right) B_{\underline{j}}(\mathcal{T}), \qquad (*)$$

where

$$P_i(x; \nu) := \sum_{m=0}^{i} (-1)^m \begin{bmatrix} x \\ m \end{bmatrix} \begin{bmatrix} n-x \\ i-m \end{bmatrix}$$

is the Krawtchouk polynomial of degree i. The equations (*) will be called the *MacWilliams identities* of C *with respect to the partition* $\mathcal{T}$. They generally give more information about the existence and the uniqueness of the code, but the price is a steeply increasing calculation effort. Nevertheless, a happy combination of a powerful computer and additional information on the codes under consideration should settle quite a few open problems. The number of equations is substantially reduced if C has a large minimum weight or if C is a doubly even selfdual code.

The exact weight distribution

The extreme case is the *exact weight distribution* of C, i.e. its weight distribution with respect to the partition

$$\mathcal{E} := \{\{1\}, \{2\}, \ldots, \{n\}\}$$

of the coordinate set S into its one–element subsets. Cf. [1]. Clearly, a code is completely determined by its exact weight distribution. Conversely, any nontrivial (0,1) solution of the MacWilliams identities with respect to the partition $\mathcal{E}$ will be shown to correspond to a binary linear code.

The Assmus–Mattson theorem

This famous result in [1] states sufficient conditions under which the words of fixed weight in a code form a t–design. We give a simple proof of the following extended version.

Proposition: Let $\mathcal{T}$ be a (t,n–t)–partition, let δ be an integer greater than t and let $G \subset \{0, 1, \ldots, t\} \times \{0, 1, \ldots, n-t\}$ be a subset for which the "row weights" $\nu_u := |\{\underline{x} \in G \mid x_1 = u\}|$ form a permutation of $\{\delta, \delta-1, \ldots, \delta-t\}$. Then the weight distributions $\underline{A}(\mathcal{T})$ and $\underline{B}(\mathcal{T})$ of a binary linear [n,k] code C can be calculated from the $A_{i,j}(\mathcal{T})$ with $i + j < \delta$ and the $B_{u,v}(\mathcal{T})$ *with* $(u,v) \notin G$.

The covering radius

The following reformulation of this notion in terms of coordinate partitions might be of some use.

Proposition: Let C be a binary linear code of length n and covering radius ρ. Then $\rho \geq t$ if and only if a (t,n–t)–partition $\mathcal{T}$ exists such that $A_{i,j}(\mathcal{T}) = 0$ for all i, j for which $i > j$.

Generalized Delsarte inequalities

Finally, if we define the *inner distribution* of the *nonlinear* code $C \subset \mathbb{F}_2^n$ *with respect to the partition* $\mathcal{T}$ to be the set of nonnegative rational numbers

$$A_{\underline{i}}(\mathcal{T}) := |C|^{-1} |\{(X, Y) \in C \times C \mid w_{\mathcal{T}}(X - Y) = \underline{i}\}|,$$

we can derive the following generalization of the *Delsarte inequalities* (cf. [2]):

$$\Sigma_{\underline{j}} \left(\prod_{u=1}^{p} P_{i_u}(j_u; n_u) \right) A_{\underline{j}} \geq 0 \quad \forall \underline{i}.$$

References

[1] E.F. Assmus and H.F. Mattson, New 5–Designs, J. Comb. Theory, vol. 6, 1969, pp. 122–151.

[2] P. Delsarte. An Algebraic Approach to the Association Schemes of Coding Theory, Philips Research Reports Supplements, No. 10, 1973.

[3] F.J. MacWilliams and N.J.A. Sloane, The Theory of Error–Correcting Codes. New York: North Holland, 1983.

On the Weight Distribution of Certain Primitive Binary Cyclic Codes

Jacques Wolfmann, G.E.C.T.

Université de Toulon et du Var, B.P. 132
83957 La Garde Cedex, France

The finite field of cardinality q is denoted by $\mathbb{F}_q$. Let n, s, k integers such that $ns = 2^k - 1$. Let $g(x)$ be a divisor of $x^s - 1$ over $\mathbb{F}_2$, and let $\pi(x)$ be a *primitive* polynomial of degre k over $\mathbb{F}_2$. We consider the following binary codes :

The cyclic code C, of length $N = 2^k - 1$, generated by

$$\frac{x^N - 1}{g(x)\pi(x)}$$

The cyclic code Γ, of length s, generated by

$$\frac{x^s - 1}{g(x)}$$

Theorem 1. *If $k = 2t$, $s = 2^t + 1$, then the set of non-zero weights of C is the set of all integers 2^{2t-1}, $(2^t - 1)w$, $2^{2t-1} - w$, $2^{2t-1} + 2^t - w$ such that w is a non-zero weight of Γ.*

Theorem 2. *If $k = 2t$, $s = 2^t + 1$, and if $g(x)$ is a primitive divisor of $x^s - 1$ over $\mathbb{F}_2$, then the set of non-zero weights of C is the set of all integers 2^{2t-1}, $(2^t - 1)w$, $2^{2t-1} - w$, $2^{2t-1} + 2^t - w$ such that w is even and*

$$\left| w - \frac{(2^t + 1)}{2} \right| \leqslant 2^{\frac{t}{2}}.$$

Theorem 3. *If $k = 2t$, and if $g(x) = \sum_{i=0}^{2^t} x^i$ then the set of non-zero weights of C is the set of all the integers 2^{2t-1}, $(2^t - 1)w$, $2^{2t-1} - w$, $2^{2t-1} + 2^t - w$ such that w is even and $2 \leqslant w \leqslant 2^t$.*

Theorem 4. *If*

a) $\mathbb{F}_{2^k}$ is the splitting field of $x^n - 1$ over $\mathbb{F}_2$;

b) $k = 2t$, and there exists a divisor r of t such that $2^r \equiv -1 \pmod{s}$,

then the set of non-zero weights of of all the integers

$$2^{2t-1}, \quad \left(\frac{2^{2t-1}}{s}\right) w,$$

$$2^{2t-1} + \left(\frac{\varepsilon 2^t - 1}{s}\right) w,$$

$$2^{2t-1} + \left(\frac{\varepsilon 2^t - 1}{s}\right) w - \varepsilon 2^t.$$

where $\varepsilon = (-1)^{t/r}$, and such that w is a non-zero weight of Γ.

A threshold property of linear codes

Gilles Zémor, and Gérard D. Cohen
ENST
46 rue Barrault
75634 Paris Cedex 13
France

email: zemor@res.enst.fr, cohen@inf.enst.fr

Abstract

We define and estimate the threshold probability θ of a linear code, using a theorem of Margulis originally conceived for the study of the probability of disconnecting a graph. We then apply this concept to the study of the erasure and Z-channels, for which we propose linear coding schemes that admit simple decoding. We show that θ is particularly relevant to the erasure channel since linear codes achieve a vanishing error probability as long as $p \leq \theta$, where p is the probability of erasure. Binomial codes have highest possible θ (and achieve capacity). As for the Z-channel, a subcapacity is derived with respect to the linear coding scheme. For a transition probability in the range $]\log(3/2); 1[$, we show how to achieve this subcapacity. As a by-product we obtain improved constructions and existential results for intersecting codes (linear Sperner families) which are used in our coding schemes.

Summary

We investigate and apply a seldom studied property of linear codes, namely the fact that they tend to display the following "threshold" phenomenon. Let us consider a binary linear code C, of parameters $[n, k, d]$, and let us choose randomly a vector v of length n such that every coordinate is given independently the value "1" with probability p and the value "0" with probability $1 - p$, $0 \leq p \leq 1$. Call $f_C(p)$ the probability with which v "covers" some non-zero codeword of C (i.e. is such that the support $supp(v)$ of v contains the support of some codeword c). In other words

$$f_C(p) = \sum_{v \in W(C)} p^{|v|}(1-p)^{n-|v|}$$

where $|v|$ denotes the weight of v and

$$W(C) = \{v \mid supp(v) \supset supp(c),\ c \in C,\ c \neq 0\}.$$

The behaviour we focus on is that whenever C has a large enough minimal distance, the (non-decreasing) function $p \mapsto f_C(p)$ jumps suddenly from almost zero to almost one, around a "threshold" probability θ. We will show how this fact stems from a theorem of Margulis, originally designed to prove a threshold phenomenon for the probability $f(p)$ of disconnecting a graph, when every edge is severed with probability p.

Threshold phenomena have been studied extensively in the context of random graphs. We have tried to apply those techniques to the coding context, and draw some consequences.

We will first place ourselves in the context of the erasure channel, and show that the threshold probability is a particularly relevant parameter for measuring the efficiency of a linear code.

We will also discuss at some length an application of the threshold phenomenon to the problem of devising efficient codes for the asymmetrical channel (the so-called Z-channel) where every 0 can be transformed into a 1 with a given probability p, while 1's are always correctly received. In this setting, decoding of a received vector is unambiguous whenever the latter covers no codeword apart from the one that was initially sent. The idea, broadly speaking, is to use linear codes with properly chosen threshold properties : the point is, the probability that the received vector covers some parasite codeword should be very small whenever the proportion of $0 \rightarrow 1$ faulty transitions stays under a threshold value.

We will show why *highly intersecting* codes are a good choice, provide some constructions, and discuss their behaviour relative to the capacity of the Z-channel. It will turn out that for high error probabilities (e.g. $0.586 \leq p \leq 1$) our schemes perform quite acceptably.

The automorphism group of double-error-correcting BCH codes

T.Berger
Département de Mathématiques
UFR des Sciences de Limoges, 123 av. A.Thomas,
87060 Limoges Cedex, France.

Summary

Primitive cyclic codes in a multiplicative group algebra.

A primitive cyclic code over a finite field $K = GF(q)$ is a cyclic code of length $n = q^m - 1$.

Let G^* be the multiplicative group of the finite field $G = GF(q^m)$. We consider such a code as an ideal of the modular algebra $M = K[G^*]$. An element of M is a formal sum

$$x = \sum_{g \in G^*} x_g(g), \qquad x_g \in K.$$

A primitive cyclic code C of defining-set $T \subset \{0, ..., n-1\}$ is the set

$$C = \{x \in M \,/\, \rho_s(x) = 0, \forall s \in T\}$$

Where $\rho_s(\sum_{g \in G^*} x_g(g)) = \sum_{g \in G^*} x_g g^s$, the sum $\sum_{g \in G^*} x_g g^s$ being not a formal sum, but calculated in $GF(q^m)$.

This definition is equivalent to the usual definition into the algebra $R = K[X]/(X^{q^m-1})$. If α is a primitive root of G, then the isomorphism is:

$$\begin{array}{rccc} \phi: & R & \longrightarrow & M \\ & \sum_{i=0}^{n-1} \lambda_i X^i & \longrightarrow & \sum_{i=0}^{n-1} \lambda_i (\alpha^i) \end{array}$$

Permutation groups of cyclic codes.

The permutation group of a cyclic code C is the group $Per(C)$ of permutations of the support G^*, which lets C globaly invariant.

It is known (cf. [2]) that each permutation $\sigma \in S(G)$ admits a unique representation polynomial of degree less than p^m:

$$f(X) = \sum_{i=0}^{p^m-1} \lambda_i X^i, \ \lambda_i \in G, \ and \ \sigma(g) = f(g).$$

Theorem 1 *Let C be a primitive cyclic code, and T its defining-set.*
A permutation $\sigma \in S(G^*)$, with associated polynomial $f(X) = \sum_{i=1}^{n} \lambda_i X^i$ is a permutation of C if and only if, for all $s \in T$, the polynomial $f(X)^s \bmod X^{p^m} - X$ has exponents in T, i.e. $f(X)^s = \sum_{j \in T} \mu_j X^j$, for some $\mu_j \in G$.

Automorphism groups of the binary double-error-correcting BCH codes

The binary double-error-correcting BCH code (cf.[4]) is the BCH code over $GF(2)$ of designed distance 5 and length $2^m - 1$ ($m > 2$), its defining-set is:

$$T = \{2^i, 2^i + 2^{i+1} \,/\, i \in \{0, ..., m-1\}\}$$

For $m = 3$, this code is trivial: it is the repetition code of length 7. We suppose $m > 3$.

Let B_m denote the binary double-error-correcting BCH code of length $2^m - 1$. Let σ be a permutation of B_m, with associated polynomial $f(X)$. Applying the criterion of theorem 1, for $s = 1$ we deduce

$$f(X) = \sum_{i=0}^{m-1} a_i X^{2^i} + \sum_{i=0}^{m-1} b_i X^{2^i + 2^{i+1}}$$

and for $s = 3$, we obtain

$$\begin{aligned} f(X)^3 = & \sum_{i,j \in \{0,...,m-1\}} a_i a_{j-1}^2 X^{2^i + 2^j} \\ & + \sum_{i,j \in \{0,...,m-1\}} (a_i b_{j-1}^2 + a_{i-1}^2 b_j) X^{2^i + 2^j + 2^{j+1}} \\ & + \sum_{i,j \in \{0,...,m-1\}} b_i b_{j-1}^2 X^{2^i + 2^{i+1} + 2^j + 2^{j+1}} \end{aligned}$$

The permutation σ is in $Per(B_m)$ if and only if its associated polynomial $f(X)$ is a permutation polynomial and $f(X)^3$ is a polynomial with exponents in T.

Using these conditions, we deduce the following theorem:

Theorem 2 *For $m > 4$, the automorphism group of the BCH code B_m is the semi-linear group of $GF(2^m)$ over $GF(2^m)$. For $m = 4$, the automorphism group of the BCH code B_4 is the semi-linear group of $GF(16)$ over $GF(4)$.*

Corollary 1 *For $m > 4$, the automorphism group of the extended binary double-error-correcting BCH code of length 2^m is the semi-affine group of $GF(2^m)$. For $m = 4$, its automorphism group is the semi-affine group of $GF(16)$ over $GF(4)$.*

References

[1] F.Laubie *Définition intrinsèque de certains codes cycliques et de leur extension* Rapport de recherche, département de Mathématiques, Université de Limoges, France 1991.

[2] R.Lidl, H.Niederreiter *Finite Fields* Cambridge University Press 1983.

[3] F.J.MacWilliams *Codes and ideals in group algebras* R.C.Bose and T.A.Dowling eds., Combinatorial Mathematics and its applications, Univ. of North Carolina Press, Chapel Hill (1969).

[4] F.J.MacWilliams N.J.A.Sloane *The theory of error correcting codes* North Holland, Amsterdam (1977).

A BOUND ON THE ZERO-ERROR LIST CODING CAPACITY*

Erdal Arıkan

Department of Electrical Engineering
Bilkent University, Ankara 06533, Turkey

Abstract

We present a new bound on the zero-error list coding capacity, and using which, show that the list-of-3 capacity of the 4/3 channel is at most 6/19 bits, improving the best previously known bound of 3/8. The relation of the bound to the graph-entropy bound of Körner and Marton is also discussed.

The Bound

Consider a discrete memoryless channel $K = (\mathcal{I}, \mathcal{J}, P)$ where $\mathcal{I}$ denotes the input alphabet, $\mathcal{J}$ the output alphabet, and $P(j|i)$ the probability that $j \in \mathcal{J}$ is received given that $i \in \mathcal{I}$ is transmitted. A set $S \subset \mathcal{I}^N$ is called *independent* if for every $y \in \mathcal{J}^N$

$$\prod_{x \in S} \prod_{n=1}^{N} P(y_n|x_n) = 0.$$

A set $\mathcal{C} \subset \mathcal{I}^N$ is called a zero-error list-of-L code, $L \geq 1$, if every $S \subset \mathcal{C}$ with $|S| = L + 1$ is an independent set. Zero-error list-of-L capacity is defined by

$$C_L = \limsup_{N \to \infty} \frac{1}{N} \log M(N, L)$$

where $M(N, L)$ is the maximum possible size for a list-of-L code of length N. (All logarithms are to to base 2.)

We call a channel k-uniform if k is the smallest integer for which $C_k > 0$. The new bound is as follows.

Theorem 1 *The rate R of any list-of-k code $\mathcal{C}$ on a k-uniform channel K satisfies*

$$R - \epsilon \leq \min_{1 \leq m \leq k} \min_{x_{m+1}, \ldots, x_k} \min_{P'} \frac{1}{mN} \sum_{n=1}^{N} I(X_{1n}, \ldots, X_{mn}; Y_n | x_{(m+1)n}, \ldots, x_{kn})$$

where P' ranges through all conditional probability assignments such that whenever $\{i_1, \ldots, i_m, i'_1, \ldots, i'_m, i_{m+1}, \ldots, i_k\}$ is independent in K

$$P'(j|i_1, \ldots, i_m, i_{m+1}, \ldots, i_k) P'(j|i'_1, \ldots, i'_m, i_{m+1}, \ldots, i_k) = 0$$

for all j. The mutual information term is computed using the probability assignment

$$Pr\{X_{1n} = x_{1n}, \ldots, X_{mn} = x_{mn}, Y_n = y_n\} = Q_n(x_{1n}) \cdots Q_n(x_{mn}) P'(y_n|x_{1n}, \ldots, x_{kn})$$

where Q_n is the empirical distribution of the nth coordinate of the codewords in $\mathcal{C}$, i.e., $Q_n(i)$ equals the fraction of codewords $x \in \mathcal{C}$ with $x_n = i$, $i \in \mathcal{I}$. The number ϵ goes to zero as N increases for any fixed $R \geq 0$.

For comparison, the Körner-Marton graph-entropy bound [3] states (in the above notation) that

$$R - \epsilon \leq \min_{m, P'} \frac{|\mathcal{C}|^{-(k-m)}}{mN} \sum_{x_{m+1}, \ldots, x_k} \sum_{n=1}^{N} I(X_{1n}, \ldots, X_{mn}; Y_n | x_{(m+1)n}, \ldots, x_{kn})$$

where the outer summation is over all possible choices of distinct codewords $x_{m+1}, \ldots, x_k \in \mathcal{C}$. Thus, the Körner-Marton bound upperbounds the rate R by (essentially) the average of the quantity $\sum_{n=1}^{N} I(X_{1n}, \ldots, X_{mn}; Y_n | x_{(m+1)n}, \ldots, x_{kn})$, whereas here R is bounded by the minimum of the same quantity.

The bound here may also be seen as a generalization of the Shannon bound on zero-error capacity [1], [2]. Shannon's bound is obtained by looking at the zero-error code through a single user channel; here we look at the code through a multiaccess channel.

The 4/3 Channel

The 4/3 channel has a four letter input and output alphabet $A = \{0, 1, 2, 3\}$, and the transition probabilities $P(j|i) = 1/3$ for all $i, j \in A$, $i \neq j$. The bound $C_3 \leq 6/19$ is obtained (after some manipulation) by applying the above theorem using the following P'. (i) For any $i, i_1, j \in A$, $P'(j|i_1, i, i) = \delta_{ij}$. (ii) For any $i_1, i_2, i_3, j \in A$ with $i_2 \neq i_3$,

$$P'(j|i_1, i_2, i_3) = \begin{cases} 0 & \text{if } j \in \{i_1, i_2, i_3\}; \\ (4 - |\{i_1, i_2, i_3\}|)^{-1} & \text{otherwise.} \end{cases}$$

References

[1] C.E. Shannon, 'The zero error capacity of a noisy channel,' *IEEE Trans. Inform. Theory*, vol. IT-2, no. 3, pp. 8-19, 1956.

[2] P. Elias, 'Zero error capacity under list decoding,' *IEEE Trans. Inform. Theory*, vol. IT-34, No. 5, pp. 1070-1074,ept. 1988.

[3] J. Körner and K. Marton, 'On the capacity of uniform hypergraphs,' *IEEE Trans. Inform. Theory*, vol. IT-36, No.1, pp. 153-156, Jan. 1990.

*This work has been supported by TÜBİTAK under project TBAG 1053.

APPROXIMATION THEORY OF OUTPUT STATISTICS

Te Sun Han
Dept. Information Systems
Senshu University
Kawasaki 214, Japan

Sergio Verdú
Dept. Electrical Eng.
Princeton University
Princeton, NJ 08544

Abstract

Given a channel and an input process we study the minimum randomness of those input processes whose output statistics approximate the original output statistics with arbitrary accuracy. We introduce the notion of resolvability of a channel, defined as the number of random bits required per channel use in order to generate an input that achieves arbitrarily accurate approximation of the output statistics for any given input process. We obtain a general formula for resolvability which holds regardless of the channel memory structure. We show that, for most channels, resolvability is equal to Shannon capacity.

By-products of our analysis are a general formula for the minimum achievable (fixed-length) source coding rate of any finite-alphabet source, and a strong converse of the identification coding theorem, which holds for any channel that satisfies the strong converse of the channel coding theorem.

There are situations of practical interest where a random process needs to be generated with some specified statistics. In order to generate a random process we assume that a primary random source with an equiprobable distribution is available (e.g. a stream of independent fair coin flips). A key measure of the complexity of a random process is the rate at which its most efficient generator requires random bits, in order to generate every sample-path of the random process. This question becomes particularly interesting when rather than requiring the exact reproduction of the desired statistics, we require an arbitrarily accurate approximation of the finite-dimensional distributions. This requires the introduction of a measure of distance between the desired and generated distributions; in this paper we focus most of our attention on the variational or l_1 distance. We prove that for any random process the minimum complexity required to approximate its statistics is equal to its minimum achievable fixed-rate (noiseless) source coding rate, and that this rate is equal to the *sup-entropy rate* of the random process. The Asymptotic Equipartition Property plays no role in the proof of this result, not only because it is not powerful enough to yield an approximation result in the sense of variational distance, but because the result holds for processes that are not necessarily ergodic or stationary. The proof uses a new technique we refer to as *repetition*.

Some practical situations such as system simulation or the remote artificial generation of random processes such as speech sounds or image textures, suggest an important generalization of the foregoing setup: Given an input process and a channel, we want to approximate the resulting output process. However, this problem does not boil down to the previous setup when the approximation has to be accomplished by generating the input. We define the *resolvability* of a channel as the number of random bits per input sample required to achieve arbitrarily accurate approximation of the output statistics regardless of the actual input process. Intuitively, we can anticipate that the resolvability of a system will depend on how "noisy" it is. A coarse approximation of the input statistics whose generation requires comparatively few bits will be good enough when the system is very noisy, because, then, the output cannot reflect any fine detail contained in the input distribution.

Although the problem of approximation of output statistics involves no codes of any sort or the transmission/reproduction of information, its analysis and results turn out to be Shannon theoretic in nature. In fact, our main conclusion is that (for most channels) resolvability is equal to Shannon capacity.

More concretely we show that the resolvability of an arbitrary channel is equal to the supremum of the input-output *sup-information rate*, and that this quantity coincides with the Shannon capacity if and only if the channel satisfies the strong converse.

In addition to the abovementioned connections with the theories of source coding and channel coding, the approximation of output statistics is related to the problem of identification via channels introduced by Ahlswede and Dueck [1]. Although a completely general direct identification coding theorem is known [1,2], its converse had been shown only in a so-called soft version in [1] and in the strong sense in [3], but always within the context of discrete memoryless channels. Here, we show a general strong converse to the identification coding theorem which follows as a simple consequence of the achievability part of the resolvability theorem.

The paper also investigates the effect of replacing the worst-case complexity measure by the average number of random bits required for approximation, as well as the replacement of variational distance by normalized divergence. In the cases considered, the foregoing conclusions remain valid.

We conclude with another result within the approximation theory of output statistics which formalizes a folk-theorem in channel coding: the output distribution due to any good channel code (a code with rate close to capacity and vanishing error probability) must approximate the output distribution due to the input that maximizes mutual information, and thus achieves capacity.

The journal version of this paper is to appear in [4].

References

1. R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Information Theory*, vol. IT-35, pp. 15-29, Jan. 1989.
2. S. Verdu and V. K. Wei, "Explicit Construction of Optimal Constant-Weight Codes for Identification via Channels," *IEEE Trans. Information Theory*, vol. IT-39, Jan. 1993.
3. T. S. Han and S. Verdu, "New Results in the Theory and Applications of Identification via Channels," *IEEE Trans. on Information Theory*, vol. IT-38, pp. 14-25, Jan. 1992.
4. T. S. Han and S. Verdu, "Approximation Theory of Output Statistics," *IEEE Trans. Information Theory*, vol. 39, May 1993.

This work was supported in part by the Office of Naval Research under Grant N00014-90-J-1734.

THE SPERNER CAPACITY OF LINEAR AND NONLINEAR CODES FOR THE CYCLIC TRIANGLE

A. R. Calderbank
R. L. Graham
L. A. Shepp
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

P. Frankl
CNRS
15 Quai Anatole France
75007 Paris, France

W.-C. W. Li
Mathematics Department
Penn State University
University Park, PA 16802

Shannon introduced the concept of zero-error capacity of a discrete memoryless channel. The channel determines an undirected graph on the symbol alphabet, where adjacency means that symbols cannot be confused at the receiver. The zero-error or Shannon capacity is an invariant of this graph. Gargano, Körner, and Vaccaro have recently extended the concept of Shannon capacity to directed graphs. Their generalization of Shannon capacity is called Sperner capacity. We resolve a problem posed by these authors by giving the first example (the two orientations of the triangle) of a graph where the Sperner capacity depends on the orientations of the edges.

Sperner capacity seems to be achieved by nonlinear codes, whereas Shannon capacity seems to be attainable by linear codes. In particular, linear codes do not achieve Sperner capacity for the cyclic triangle. We use Fourier analysis or linear programming to obtain the best upper bounds for linear codes. The bound for unrestricted codes are obtained from rank arguments, eigenvalue interlacing inequalities and polynomial algebra.

The statement of the cyclic q-gon problem is very simple: what is the maximum size $N_q(n)$ of a subset S_n of $\{0, 1, \ldots, q-1\}^n$ with the property that for every pair of distinct vectors $x = (x_i), y = (y_i) \in S_n$, we have $x_j - y_j \equiv 1 \pmod{q}$ for some j? For $q = 3$ (the cyclic triangle), we show $N_3(n) \simeq 2^n$. If however S_n is a subgroup, then we give a simple proof that $|S_n| \leq \sqrt{3}^n$.

SECRECY ENHANCEMENT VIA PUBLIC DISCUSSION

Alon Orlitsky* Avi Wigderson†

Abstract

(X, Y, Z) is an ensemble of independent random triples, each distributed according to some probability distribution $p(x, y, z)$. Two legitimate users, P_X having X and P_Y having Y, communicate in order to agree on a joint key while keeping it almost unknown to an eavesdropper P_Z who knows Z. Communication is conducted over a noiseless channel according to a predetermined protocol. P_Z hears all transmissions over the channel and knows the protocol used. We show: (1) The legitimate communicators can agree on the secret if and only if they can find one using just two messages. (2) There are cases where a secret can be found, but one message does not suffice. (3) Similar results hold whether the legitimate communicators are required to agree on the secret with probability one or just with high probability.

Summary

The following problem was introduced by Maurer [1] and further investigated by Ahlswede and Csiszár [2]. It concerns two parties with some common information conversing publicly to agree on a secret key that is unknown to an eavesdropper listening to their discussion.

Let (X, Y, Z) be a sequence $(X_i, Y_i, Z_i)_{i=1}^n$ of independent random triples, each distributed according to a probability distribution $p(x, y, z)$. Two *legitimate users*, P_X having X and P_Y having Y, communicate over a noiseless channel according to a predetermined protocol in order to agree on a joint key. An *eavesdropper* P_Z who knows Z and the communication protocol, and has access to all the bits transmitted over the channel, tries to determine the value of the key.

The probability distribution p is said to achieve a *secrecy rate* s if for every $\epsilon > 0$ there exists n, a (possibly randomized) communication protocol Φ defined on X and Y, and a random *key* K, such that: (1) P_X and P_Y know the key: $H(K|X, \Phi(X, Y)) < \epsilon$ and $H(K|Y, \Phi(X, Y)) < \epsilon$; (2) P_Z does not know the key: $H(K|Z, \Phi(X, Y)) \geq H(K) - \epsilon$; (3) K has a per-letter entropy of at least s: $\frac{1}{n}H(K) \geq s$. The *secrecy capacity* $C(p)$ *of* p is the largest achievable secrecy rate.

Determining the secrecy capacity of a given distribution, or even whether this capacity is positive, seems difficult and only weak general bounds are known [1]. For that reason, [2] considered the simpler one-way version of the problem. The legitimate users are allowed to transmit only one message (say from P_X to P_Y). For this restricted case, [2] determined the secrecy capacity of a probability distribution p in term of its single-letter entropies. Yet interaction introduces "memory" to the problem, and similar results appear unlikely.

In this paper we introduce a gradation of measures ranging from the one-way capacity of p to its secrecy capcity. A communication protocol is *m-message* if it always calls for at most m transmitted messages. For example, a two-message protocol may require P_X to transmit a message and then call on P_Y to respond. We define the achievable *m-message secrecy rates* and the *m-message secrecy capacity* $C_m(p)$ of p as we did before, except that the protocols allowed must be m-message. In particular, $C_1(p)$ is the one-way secrecy capacity considered by [2], and $C(p) = \lim_{m\to\infty} C_m(p)$.

Using communication-complexity techniques we find a necessary and sufficient condition for the existence of a secret key (i.e., $C(p) > 0$). We use this condition to show that $C(p) > 0$ implies $C_2(p) > 0$, hence that secrecy can be achieved if and only if it can be achieved using just two messages. We then show that there are cases where a single message cannot achieve a secret key ($C_1(p) = 0$), but two or more messages can ($C(p) > 0$). Therefore, there is a gap between one message and two or more. Potentially, one could use the necessary and sufficient condition to improve the general bounds, but so far we have not been able to do so.

We also consider the *unambiguous secrecy capacity* of p where P_X and P_Y must know the key with probability 1. Again, we show that a secret key exists if and only if it can be achieved with just two messages, and we give a simple necessary and sufficient condition. Additionally, we examine the more general case where (X, Y, Z) is an arbitrary triple of random variables (rather than an ensemble). We show that when X and Y are uniformly distributed over their support set and are independent of Z the (appropriately modified) capacity is between $I(X;Y) - \log\min\{H(X|Y), H(Y|X)\}$ and $I(X;Y)$.

References

[1] U. Maurer. Perfect cryptographic security from partially independent channels. In *Proc. of the 23rd Annual ACM Symposium on Theory of Computing*, pages 561-571, May 1991.

[2] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography part I: Secret sharing. In *Proc. of the IEEE International Symposium on Information Theory*, June 1991.

*Rm. 2C-361, AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974

†Computer Science Department, Hebrew University, Jerusalem 91904, Israel

TOWARDS COMBINING SHANNON'S THEORY ON SECRECY SYSTEMS AND THE THEORY OF AUTHENTICATION IN THE CASE OF MULTIPLE CHANNEL USE

Ben Smeets
Department of Information Theory
Lund University, Box 118
S-221 00 Lund, Sweden

Abstract

In this paper we consider cipher systems that provide both secrecy *and* security for a given number of (subsequent) transmissions. We show that there exists a broad class of situations in which we can do better (less key requirement) than just concatenating a perfect secrecy cipher and an authentication code.

Summary

One of the important results in Shannons seminal paper on secrecy systems is that if the cipher is a perfect group-operation cipher, i.e., a group-operation cipher in which the keys are uniformly, independently distributed, then the system is unconditionally perfect, [1]. The situation in which the eavesdropper is active was considered by Simmons who first to give bounds on the probability of a successful imitation attack, P_I, and the probability of a successful substitution attack, P_S. The results of Simmons deal with the situation in which the (legal) sender sends only one message. His results were improved and generalized to the case of multiple use of the channel, [3], [4].

A naive solution for obtaining security against the active eavesdropper's actions in the multiple-use case would be to select a new key for every new transmission. It follows from [2] that in the case of L transmissions the total amount of key is bounded from below by

$$H(K) \geq -L \log_2(P_I P_S) \geq -2L \log_2 P_d, \qquad (1)$$

where $P_d = \max(P_I, P_S)$ and where $H(K)$ denotes the average total key uncertainty. However, in general, it was shown in [4] that one actually has the improved bound

$$H(K) \geq -(L+1) \log_2 P_d. \qquad (2)$$

It is well-known that one has a certain amount of security in a secrecy system if the message source exhibits redundancy. Using this idea we can obtain security even if the source exhibits no redundancy by introducing redundant dummy source letters before the encryption. In the sequel we assume for simplicity that the source is a M-ary symmetric memoryless source, i.e., $H(M) = \log_2 M$. We can show the following result

Theorem 1: Suppose the source letters are encoded by an error-control code with rate $R < 1$ and subsequently are encrypted by a perfect group-operation cipher. Then this system provides perfect secrecy and the probability of a successful impersonation attack, P_I, satisfies $P_I \geq M^{-1/R}$. Moreover, the same performance can be obtained for L channel uses using a key whose uncertainty satisfies $H(K) \geq L\frac{1}{R}\log_2 M$. ■

This theorem suggests that the design of systems with secrecy and security would be a simple one. Unfortunately it can be shown that it may happen that we have $P_S = 1$.

Let us for simplicity assume that we want $P_I = P_S$ for every transmission. We can prove the following

Theorem 2: Suppose we have an A-code that provides perfect secrecy and $P_d = P_I = P_S$ for one transmission. If the A-code allows for an encoding rule updating scheme as discussed in [5], then we have a cipher system that provides perfect secrecy for a discrete memoryless M-ary source, and for which $P_d = P_I = P_S$ is equal for L subsequent transmissions such that the key uncertainty $H(K)$ satisfies

$$H(K) \geq L(H(M) + \log_2 \frac{1}{P_d}) - (H(M) - \log_2 \frac{1}{P_d}).$$

■

Note that if one naively concatenates a prefect secrecy code with the A-code we would obtain $H(K) \geq LH(M) + (L+1)\log_2 \frac{1}{P_d}$. Note also that Theorem 3, albeit for a special case, in some sense combines Shannon's result on the key entropy for secrecy systems and the key requirements induced by the security demands.

References

[1] C.E. Shannon, "Communication theory of secrecy systems, Bell Systems Tech. J., Vol. 28, 1949, pp. 656-715.

[2] Simmons, G.J., "Authentication theory/coding theory", in Advances in Cryptology, Proceedings of CRYPTO 84, G.R. Blakley and D. Chaum, Eds. Lecture Notes in Computer Science, No. 196. New York, NY: Springer, 1985, pp. 411-431.

[3] R. Johannesson and A. Sgarro, " A strengthening of Simmons' bound on impersonation", IEEE Trans. on Information Theory, Vol. IT-37, (1991), pp. 1182-1185.

[4] M. Walker, "Information-theoretic bounds for authentication schemes", J. Cryptology, Vol. 2, No. 4, 1990, pp. 131-143

[5] G.J. Simmons, B. Smeets, "A paradoxical result in unconditionally secure authentication codes - and an explanation", Proceedings IMA Conference on Cryptography and Coding, Dec. 18-20, 1989, Cirencester, England.

This work was supported by the TFR grant 222 92-662

POSITIONING AND COMMUNICATION SYSTEMS

C. R. Drane
School of Electrical Engineering
University of Technology, Sydney
PO Box 123, Broadway NSW 2007, Australia

Introduction

Positioning Systems are devices that measure the position of remote objects. Examples include radars, sonars, the Global Positioning Systems (GPS) [4] and vehicle tracking systems [2]. A recently published monograph [1] describes a unified approach to the analysis of positioning systems. The major element of this approach is Shannon theory. In that monograph, it is shown that this approach can be used to establish a performance measure for positioning systems (based on the average mutual information), a limit to that performance (using a generalisation of Shannon's capacity theorem), derive general theorems about positioning systems, calculate a source information rate for the objects being monitored and optimise aspects of the system performance. The analysis presented in that monograph covers both multi-link and multidimensional [3] channels for the case of additive, white gaussian noise (AWGN). Although the AWGN assumption does allow insight into the functioning of positioning systems, it does limit the applicability of the analysis.

Information theoretic analysis of conventional communications systems also started using the AWGN assumption, however much work has been carried out to derive results for more realistic channel models. This paper shows how a multi-link and multidimensional positioning system can be partioned into its various component parts, so allowing the results from conventional communication theory to be directly applied to positioning systems.

Analysis

The essence of the paper is prove a theorem which establishes that the average mutual information of a multi-link or multidimensional channel is equal to the sum of the average mutual information of each individual channel minus a term which represents the degree of interaction between the channels (the following symbols are defined in [1]) i.e.

$$I(x1,x2;y1,y2)=I(\xi1;\phi1)+I(\xi2;\phi2)-I(\phi1;\phi2) \quad \ldots . (1)$$

This equation is easily generalised to more than two dimensions, though the form of the last term changes somewhat.

Equation (1) holds under the condition that the noise is independent between channels e.g. the noise on one link is independent of the other links. It can then shown that provided the measurement error is small and the co-ordinate system is not highly correlated that the degree of interaction is dependent almost entirely on the geometric nature of the positioning system's co-ordinate system i.e.

$$I(\phi1;\phi2) \approx I(\xi1;\xi2) \quad \ldots . (2)$$

This means that the overall performance of a system can be estimated from the individual single channel performance (already well explored for conventional communication systems) together with this geometric term. This term will be denoted S_g i.e.

$S_g \equiv I(\xi1;\xi2)$.

Results

The paper goes on to give examples of the nature of this geometric term for radial-radial, angle-angle and polar positioning systems. A radial-radial system measures position by calculating the intersection of two circles. These circles are often derived from ranging measurements. An angle-angle system measures position by calculating the intersection of two lines. Two direction finding stations perform this type of operation. Simple radar systems operate with a polar co-ordinate system.

The geometric term, S_g, was calculated for a radial-radial, angle-angle and polar system. In each case a rectangular a priori p.d.f of width 2a and height a, centred around the y-axis with the bottom edge aligned along the x-axis. The radial-radial and angle-angle systems had reference sites at (-.5,0) and (+.5,0). For each of these three systems, S_g was calculated as a function of a, using Mathematica. This involved using a combination of numerical and analytical integration. The result is shown in Figure 1.

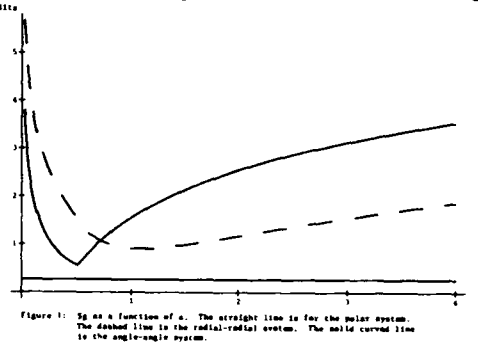


Figure 1: S_g as a function of a. The straight line is for the polar system. The dashed line is the radial-radial system. The solid curved line is the angle-angle system.

The system with the lowest value of S_g was the polar system. This is to be expected as the radial and angular co-ordinates in a radial system are almost independent. Indeed, if the a priori p.d.f is circular then S_g=0. On the other hand both the radial-radial and angle-angle systems are clearly not independent and have large values of S_g. In the case of the radial-radial system this is because the determination of the first radial measurement constrains the second measurement to those radii which will intersect with the circle prescribed by the first measurement. Similar reasoning applies to the angle-angle system.

At first sight the values of S_g for the radial-radial and angle-angle system do not seem to be large, but it should be remembered that as a rule of thumb, a one bit reduction in performance will be translated to a halving in system accuracy, so a three bit loss will cause an eight fold reduction in accuracy. Note that both the angle-angle and radial-radial systems have very large values of S_g when a is either small or large. This result should be treated with caution because a large value of S_g means that the co-ordinates are becoming highly correlated. This means that one of the assumptions used in deriving the basic equation will not be satisfied.

This geometric term can also contrasted with the Geometric Dilution of Precision (GDOP) for these systems. Given a properly selected point at which to calculate the GDOP the overall trends shown in Figure 1 are confirmed. A brief example is also provided as to how a systems engineer might use the results of this paper in the analysis of a positioning system.

Overall, the paper presents a result which will allow systems engineers to directly apply results from realistic single-link channels to the analysis of multi-link channels used in positioning systems. As well the analysis allows a deeper understanding of the comparative performance of different types of systems.

References

[1] C.R. Drane. Positioning Systems - A Unified Approach. A 170 page monograph to be published by Springer Verlag in second half of 1992.

[2] G.K. Hurst. Quiktrak: a new AVL system developed in Australia. Proceeding of IREECON '89, Melbourne, 1():78-80, 1989.

[3] T. Kailath. On multilink and multidimensional channels. IRE transactions on Information Theory, IT-8():260-261, April 1962.

[4] R.J. Milken and C.J Zoller. Principles of Operation of navstar and system characteristics. Navigation: Journal of the Institute of Navigation. 25(2):95-106. Summer 1978.

COMMUNICATING OVER A CHANNEL CONSTRAINED TO A FIXED CODE

Aaron B. Kiely and John T. Coffey
Electrical Engineering and Computer Science Department
The University of Michigan, Ann Arbor, MI 48109

***Abstract*— We examine the problem of decoding a linear block code used over a binary symmetric channel when the goal is to minimize the average information bit error probability. For fixed crossover probability p, the optimal decoder can be implemented by standard array. We present optimal strategies for choosing coset leaders in the very quiet ($p \to 0$) and very noisy ($p \to 1/2$) limits.**

SUMMARY

We consider the problem of decoding a binary (n,k) linear block code $\mathcal{C}$ used over a binary symmetric channel (BSC) with error probability $p < 1/2$. We assume that an information vector $\mathbf{u}$ is chosen at random from $\mathcal{U}$, the set of all binary k-tuples, with each element of $\mathcal{U}$ having equal probability of being chosen. The encoder transmits the codeword $\mathbf{c} = \mathbf{uG}$ across the BSC, where $\mathbf{G}$ is a generator matrix of the code.

If the goal of the decoding is to minimize the average probability of a codeword error, then the well known solution is to use a standard array decoder with minimum weight coset leaders. In some applications, however, we might be more interested in minimizing the average information bit error probability P_{inf}, given by

$$P_{\text{inf}} := \frac{1}{k} E[|u + \hat{u}(r)|]$$

where $\hat{\mathbf{u}}(\mathbf{r})$ denotes the estimate of $\mathbf{u}$ given the received vector $\mathbf{r}$, and $|\mathbf{x}|$ denotes the Hamming weight of $\mathbf{x}$.

The decoding problem in the very quiet limit $p \to 0$ has been examined before when $\mathbf{G}$ is systematic, for example in [3], [4]. The problem of choosing an optimal generator matrix under certain constraints is discussed in [2], [5]. Here we make no particular assumptions about the optimality of $\mathbf{G}$.

In general, for a particular code there will be several different strategies, each corresponding to the optimal decoding rule over some range $p \in [p_{i-1}, p_i]$. As pointed out in [1], each of these strategies can be implemented by standard array. For example, Figure 1 shows $P_{\text{inf}}(p) - p$ for the (15,7) BCH code with a systematic generator matrix and optimal decoding. Each p_i is a root of a polynomial of the form

$$\sum_{\mathbf{x} \in C_i^{(0)}} \left(\frac{p}{1-p}\right)^{|\mathbf{r}+\mathbf{x}|} - \sum_{\mathbf{x} \in C_i^{(1)}} \left(\frac{p}{1-p}\right)^{|\mathbf{r}+\mathbf{x}|} = 0$$

where $C_i^{(0)}$ ($C_i^{(1)}$) is the set of codewords that can be transmitted when the ith bit of $\mathbf{u}$ is a zero (one).

After some manipulations, we find that the optimal P_{inf} is

$$P_{\text{inf}}^*(p) = \frac{(1-p)^n}{k} \sum_{\mathbf{q} \in \mathcal{Q}} \min_{\hat{\mathbf{u}} \in \mathcal{U}} \sum_{\mathbf{u} \in \mathcal{U}} |\mathbf{u} + \hat{\mathbf{u}}| \left(\frac{p}{1-p}\right)^{|\mathbf{q}+\mathbf{uG}|}$$

where $\mathcal{Q}$ is any choice of coset representatives. For fixed p, the optimal estimate of the information vector is

$$\hat{\mathbf{u}}(\mathbf{r}) = (\mathbf{r} + \mathbf{l})\mathbf{G}^{-1}$$

This work was supported in part by the National Science Foundation under Grant NSF-NCR-9105832

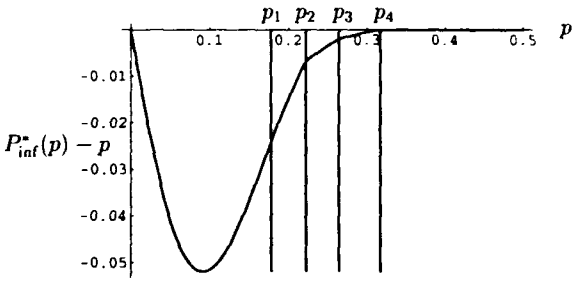


Figure 1: $P_{\text{inf}}^*(p) - p$ for the (15,7) BCH code. The optimal decoding rule may be one of five different strategies depending on the value of p.

where $\mathbf{G}^{-1}$ is the right inverse of $\mathbf{G}$ and $\mathbf{l}$ is the optimal coset leader, which is the element of the coset $\mathbf{r} + C$ that minimizes

$$\sum_{\mathbf{t} \in \mathbf{r}+\mathcal{C}} \left|(\mathbf{t} + \mathbf{l})\mathbf{G}^{-1}\right| \left(\frac{p}{1-p}\right)^{|\mathbf{t}|}.$$

Theorem 1 *For all $p \in [0, p_1]$, the following coset leader selection strategy minimizes P_{inf}:*

1. *Let $\mathbf{t}_1, \ldots \mathbf{t}_j$ denote the minimum weight elements in $\mathbf{r} + \mathcal{C}$. For each such $\mathbf{t}_i$, compute $\mathbf{g}_i := \mathbf{t}_i\mathbf{G}^{-1}$.*
2. *Let $\mathbf{y}$ be a binary vector of length k. If the majority of the $\mathbf{g}_i$ have a 1 (0) in the ith position, then set $\mathbf{y}_i$ equal to 1 (0). If no majority exists in the ith position, then repeat steps 1-2 for the ith position using the coset elements of next higher weight.*
3. *The optimal coset leader is $\mathbf{l} = \mathbf{z}(\mathbf{r}) + \mathbf{yG}$, where $\mathbf{z}(\mathbf{r})$ is the element in $\mathbf{r} + C$ having all zeros in the first k positions.*

The above strategy is different from that presented in [4]. If a coset has a unique minimum weight vector, then this vector is the optimal coset leader when $p < p_1$.

Theorem 2 *Suppose that the generator matrix is systematic, i.e., $\mathbf{G} = [\mathbf{I}_k | \mathbf{P}]$, where $\mathbf{I}_k$ is the $k \times k$ identity matrix. The following strategy minimizes P_{inf} in the very noisy region $p \in [p_m, 1/2]$: If the ith column of $\mathbf{I}_k$ occurs j times in $\mathbf{G}$, then decode the ith information bit by treating each of these j positions as a repetition code and ignoring all other positions of the received vector.*

Thus if no column of $\mathbf{P}$ is equal to a column of $\mathbf{I}_k$, then the decoding strategy that minimizes P_{inf} when $p > p_m$ is to ignore all parity check bits. I.e., the optimal coset leader will have all zeros in the first k positions. For example, in Figure 1, where a systematic encoding of the (15,7) BCH code was used, we see that for $p > p_4 \approx .32$, we get $P_{\text{inf}} = p$.

REFERENCES

[1] P. Delsarte, *IEEE Trans. Info. Th.* **24** (1978), 70-75.
[2] L. A. Dunning, *IEEE Trans. Info. Th.* **33** (1987), 91-104.
[3] M. Elia, G. Prati, *IEEE Trans. Info. Th.* **31** (1985), 518-520.
[4] B. L. Montgomery, B.V. K. V. Kumar, *IEEE Trans. Info. Th.* **34** (1988), 880-881.
[5] G. Seguin, *IEEE Trans. Info. Th.* **32** (1986), 319-322.

ABOUT CODING WITHOUT RESTRICTIONS FOR THE AWGN CHANNEL

Gregory Poltyrev

Dept. of Electrical Engineering - Systems, Faculty of Engineering
Tel-Aviv University, Tel-Aviv, 69978, ISRAEL

Abstract

Many coded modulation constructions are obtained as some restricted subset of an infinite constellation (IC) of points in the n-dimensional Euclidean space, for example, lattice code. We shall consider an IC as a code without restrictions employed for the AWGN channel. We construct exponential upper and lower bounds for the decoding error probability of an IC as functions of generalized SNR $\mu=\gamma^{-\frac{2}{n}}/\sigma^2$, where γ is the density of IC (the number of points on the unit of volume) and σ^2 is the dispersion of the AWGN. The upper bound is obtained by means of a random coding method and it is very similar to the usual random coding bound for the AWGN channel. The exponents of these upper and lower bounds coincide for lower values of μ. We show also that the exponent of the random coding bound for the ensemble of all possible IC's with the fixed density γ coincides with the exponent for the ensemble of linear IC's - lattices. We conclude from this fact that lattices have the same meaning with respect to an AWGN channel as linear codes have with respect to a discrete symmetric channel without memory.

Summary

During the last years several efficient codes for a channel with additive white Gaussian noise (AWGN) were constructed by means of coded modulation methods. Many coded modulation constellations were obtained as some restricted subset of an infinite constellation (IC) of points in the n-dimensional Euclidean space, for example lattice codes [1], [2]. Obviously, a good code can be attained only from a good IC. Furthermore the decoding error probability of the code is often estimated by means of the parameters of the IC from which this code is obtained.

Any countable set $S=\{s_1, s_2, \dots\}$ of points in the n-dimensional Euclidean space E_n will be called an infinite constellation (IC) of length n. Let $V_n(r,s)$ be the n-dimensional sphere of the radius r centered at the point s. Denote $V_n(r)=V_n(r,0)$. Let $M_n(S,r)=|S\cap V_n(r)|$, where $|A|$ is the cardinality of the set A. The limit $\lim_{r\to\infty}\frac{M_n(S,r)}{|V_n(r)|}=\gamma$, if exists, is called the density of S (here $|V_n(r)|$ is the volume of the sphere $V_n(r)$). An IC for which the density γ exists is called a regular IC. We shall consider a regular IC as a code without restrictions for AWGN channel. The value $\mu=\gamma^{-\frac{2}{n}}/\sigma^2$, where σ^2 is a dispersion of AWGN, is called a generalized SNR.

Using the random coding arguments [2], [3], we derive the following theorems.

Theorem 1. *Let*

$$E_U(\mu)=\begin{cases}\frac{\mu}{16\pi e}, & 8\pi e\le\mu,\\ \frac{1}{2}\ln\frac{\mu}{8\pi}, & 4\pi e\le\mu<8\pi e,\\ \frac{\mu}{4\pi e}-\frac{1}{2}\ln\frac{\mu}{2\pi}, & 2\pi e\le\mu<4\pi e;\end{cases}$$

$$E_L(\mu)=\frac{\mu}{4\pi e}-\frac{1}{2}\ln\frac{\mu}{2\pi}, \qquad 2\pi e\le\mu;$$

and $o(n)$ *is a sequence of reals such that* $\lim_{n\to\infty}=0$. *Then*

i: *there is a sequence of infinite constellations* S_n, n=1,2,... (n *is dimension of the Euclidian space) such that average decoding error probability of* S_n *satisfies the following asymptotical inequality* $-\frac{1}{n}\ln\lambda(S_n)\ge E_U(\mu)+o(n)$,

ii: *for any infinite constellation* S_n $-\frac{1}{n}\ln\lambda(S_n)\le E_L(\mu)+o(n)$, $\mu\ge 2\pi e$.

iii: *for any sequence of* IC S_n, n=1,2, ... *such that* $\mu<2\pi e$, $\lim_{n\to\infty}\lambda(S_n)\ge 0.5$.

Theorem 2. *There is a sequence of lattices* G_n, n=1,2,... (n *is the dimension of the Euclidian space) such that the decoding error probability* λ *of* G_n *satisfies the following asymptotical inequality* $-\frac{1}{n}\ln\lambda(0)\ge E_U(\mu)+o(n)$.

It is well known [4] that in the case of discrete symmetric channels without memory the random coding exponents for the ensemble of all codes and for the ensemble of linear codes coincide. It follows from Theorem 2, the same fact take place also in the AWGN channel case for codes without restrictions. The question, *whether the random coding exponents for the ensemble of all codes and for the ensemble of linear codes coincide also for any additive noise continues channel without memory*, remains open.

References

[1] J.H. Conway and N.J. Slone, "Sphere Packings, Lattices and Groups", *New York: Springer*, 1988.

[2] G.D. Forney, Jr., "Coset Codes - Part 1: Introduction and Geometrical Classification," *IEEE Trans. Information Theory*, vol.IT-34, no.5, pp. 1123-1151, Sept. 1988, Part II.

[3] C.E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol.38, pp. 611-656, May 1959.

[4] R.G. Gallager, "Information Theory and Reliable Communication," *New York: J.Wiley*, 1968.

NONCOHERENTLY DEMODULATED CONVOLUTIONAL CODES

Y. Kofman, E. Zehavi and S. Shamai (Shitz)
Department of Electrical Engineering
Technion—Israel Institute of Technology, Haifa 32000, Israel

Noncoherent detection schemes are extensively used when it is difficult to establish or maintain an accurate carrier phase [1]-[2]. We present a noncoherent coded system based on BPSK modulated convolutional codes which bridges the performance gap with respect to coherent coded systems by making use of a noncoherent decoding metric which incorporates an observation interval of several channel signals. The discrete time channel model considered in this paper is given by

$$Y_i = X_i e^{j\theta} + W_i \,, \quad i \in \mathbb{Z}_+ \tag{1}$$

where $X_i = \pm\sqrt{E_s}$ and Y_i are the transmitted and the received signals, respectively. The noise W_i is a sample of an independent and identically distributed sequence of complex Gaussian random variables with zero mean and variance $N_0/2$ in each dimension. The carrier phase θ is assumed to remain constant over L channel signals and to be uniformly distributed in the interval $[-\pi, \pi)$. For a rate k/n convolutional code, the suboptimal noncoherent branch metric calculated for a subsequence of $L = Jn$ signals is given by

$$\eta = \left| \sum_{j=1}^{L} Y_j X_j \right|^2 \,. \tag{2}$$

The parameter L is referred to as the length of the observation interval. The metric of an entire code sequence is given by the sum of metrics of its constituent L-long subsequences. Since L is a multiple of n, the metric is calculated over an integral number (J) of branches in the trellis diagram of the code. Therefore, for an arbitrary J and for a given number of states, decoding is easily accomplished by using a conventional rate $\frac{Jk}{Jn}$ Viterbi decoder with the same number of states and a branch metric given by (2). Note that since the metric (2) is calculated separately for each subsequence without any regard to previous subsequences, the error performance of the system would be the same whether θ changes arbitrarily once every L signals or remains constant forever.

The Chernoff bounding technique is employed to obtain upper bounds on the pairwise error probability and the average bit error probability, and a simple expression for the generalized cut-off rate [3]. Large deviations techniques are used to find the exponential rate of the error probability of the proposed system. This parameter leads to the definition of the equivalent free distance of the underlying convolutional codes of the noncoherent system. Upper bounds on the free distance are provided as well.

The metric in (2) raises the problem of phase ambiguity since it is invariant to a 180° rotation of an L-long subsequence. Conventionally, this problem is resolved by using a reference signal and differential encoding and decoding [1]-[2] In our approach, however, the phase ambiguity problem is resolved as an inherent part of the coding system in a general framework of catastrophic error propagation. Nevertheless, there are close relations, depending on the carrier phase model, between both approaches. It is shown that for a model of carrier phase changing arbitrarily every L channel signals, the proposed system is equivalent to appropriate differential systems, and for a constant carrier phase, the proposed system constitutes the natural framework for analyzing and synthesizing standard differentially encoded systems [2]. In particular, it is concluded that known optimal codes for coherent detection, namely those codes which achieve large Hamming distance, are not necessarily optimal for various differential systems as long as the observation interval is longer than two. This fact is demonstrated by the bounds on the pairwise and bit error probability and verified by the equivalent free distance of specific codes found by a computer search.

References

[1] D. Divsalar and M.K. Simon, "Multiple-Symbol Differential Detection of MPSK", *IEEE Trans. on Commun.*, Vol. **38**, No. 3, pp. 300-308, March 1990.

[2] D. Divsalar, M.K. Simon, and M. Shahshahani, "The Performance of Trellis-Coded MDPSK with Multiple Symbol Detection", *IEEE Trans. on Commun.*, Vol. **38**, No. 9, pp. 1391-1403, September 1990.

[3] M.K. Simon, J.K. Omura, R.S. Scholtz and B.K. Levit, Spread Spectrum Communications, Computer Science Press, Rockville, MD., 1985

SELECTION AND SQUARE-LAW COMBINING FOR NCFSK WITH CORRELATED BRANCH DIVERSITY

P.J. McLane and C.S. Chang
Department of Electrical Engineering
Queen's University, Kingston, ON, K7L 3N6*

A theoretical performance analysis has been conducted for the reception of noncoherent frequency shift keying (NCFSK) over a fading channel. The receiver is a bank of energy detectors, one energy detector for each frequency in the NCFSK signal set. The tones used in this set are assumed to be orthogonally spaced. The key aspect of the study is that antenna diversity is considered and the fading on the diversity branches is assumed to be correlated. A general number of diversity branches is considered. The signal set is 2-, 4- or 8-Ary NCFSK. The fading is assumed to be flat and to vary slowly in time. Both the Rayleigh and Rician fading models are treated.

Two diversity combining rules are considered. In square-law combining the outputs of the various diversity branches for each energy detector are weighted and summed. In selection diversity the diversity branch with the largest signal-to-noise ratio is the branch chosen for NCFSK detection.

We first discuss our results for square-law combining. It is well known that a Rayleigh random variable can be regarded as the magnitude of a complex Gaussian random variable. In our correlation matrix the real parts of these variables on different branches are assumed to be correlated. The same is true of the imaginary terms. However, the real and imaginary parts are always assumed to be uncorrelated. The detection random variable is a sum of squares of these variables. This sum is then transformed to another sum of squares, but now the terms are independent. The probability of error is then expressed as up to a two-dimensional integral in the transformed random variables. This diagonalization technique works with any correlation matrix, Rician or Rayleigh fading, and with any practical order of diversity L.

Samples of our calculations are shown in Figures 1 and 2, below. In Figure 1 we show how Rician fading degrades with correlated branch diversity and square-law combining. Figure 2 repeats the situation for the Rayleigh case. The loss due to correlation is greater for Rician than for Rayleigh. Thus, in a correlated diversity environment, performance in Rayleigh fading can be better than in Rician fading. This can never occur for uncorrelated fading.

In the result just discussed the correlation coefficient between diversity branches was the same. The weighting on branches was equal. No matter what the distribution of correlation coefficients between branches, equal weighting per diversity branch was always found to be best.

We have also considered selection diversity combining and compared it to square-law combining. It was found that selection combining is inferior to square-law combining in correlated diversity situations. A report has been written on our research and is referenced as [1] below. A list of previous research on correlated diversity branches is given in [1]. Finally, Mazo's matched filter lower bound [2] for two-beam, frequency-selective fading involves a quadratic form. We have applied our diagonalization method to it and rederived Mazo's result. The derivation is given in [1].

References

[1] P.J. McLane and C.S. Chang, "Selection and Square-Law Combining for NCFSK with Correlated Branch Diversity, Final Progress Report, Part II", "A Study of Space Communications Spread-Spectrum Systems", The Department of Communications, DSS Contract No. 36001-0-3505/01-SS.

[2] J.E. Mazo, "Exact Matched Filter Bound for 2-Beam Rayleigh Fading", *IEEE Trans. on Comm.*, Vol. 39, pp. 1027-1030, July 1991.

* The research contained herein was funded by the Department of Communications, Contract #36001-0-3505/01-SS.

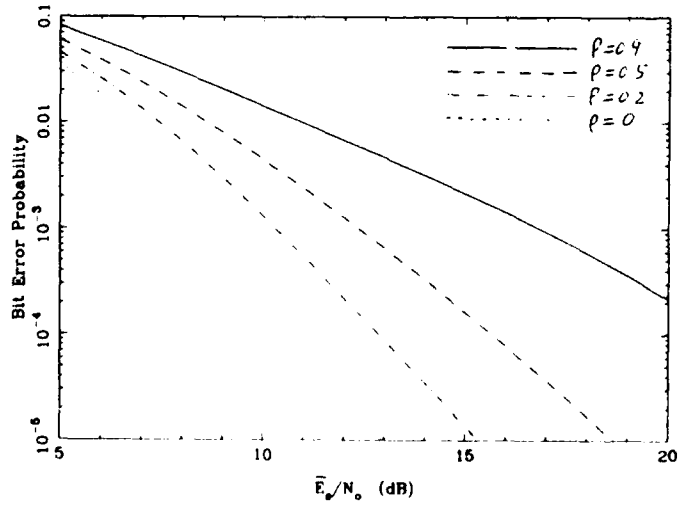


Figure 1: BER for K = 5 dB Rician and L = 4.

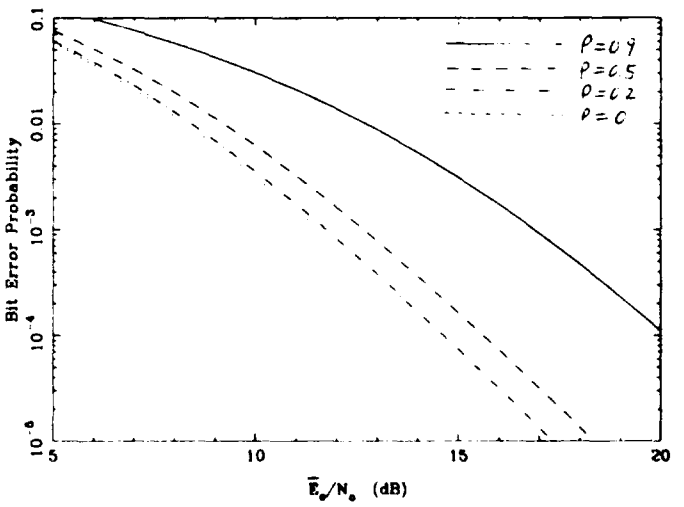


Figure 2: BER for K = 0 and L = 4.

UNDETECTED ERROR PROBABILITY OF LINEAR BLOCK CODES ON CHANNELS WITH MEMORY

Francis Swarts[1], Student Member IEEE, A.J. Han Vinck[2], Member IEEE and Hendrik C. Ferreira[1], Member IEEE.

[1]Laboratory for Cybernetics, Rand Afrikaans University, P.O. Box 524 Auckland Park 2006, Republic of South Africa.

[2]Institute for Experimental Mathematics, University of Essen, Ellernstrasse 29, 4300 Essen-12, Germany.

Abstract:- This paper addresses the problem of determining the undetected error probability, $P_u(\varepsilon)$, for linear (n,k) block codes on channels with memory. In the past, $P_u(\varepsilon)$ was investigated mainly on **memoryless channels**, such as the binary symmetric channel (BSC). We present two techniques for determining $P_u(\varepsilon)$, where both techniques employ trellis diagrams. The first technique is based upon a trellis diagram of the states of a channel model such as the Gilbert-Elliott or Fritchman channel models. The second technique involves taking the trellis diagram of the syndrome register of a code as well as the stationary and transition probabilities of any of the aforementioned channel models into account. Results indicate that in many cases $P_u(\varepsilon)$ for codes on channels with memory, far exceeds that of $P_u(\varepsilon)$ on memoryless channels for the same code. This fact therefore makes it very important to be able to calculate $P_u(\varepsilon)$ on channels with memory, seeing that $P_u(\varepsilon)$ on the BSC certainly does not represent an upperbound. We also show that the often assumed upperbound on $P_u(\varepsilon)$, $2^{-(n-k)}$, is exceeded on channels with memory. The first technique that we present is applicable to short or low rate codes, while the second can be used with high rate or long codes.

SUMMARY

Until Leung & Hellman [1] proved differently, the undetected error probability ($P_u(\varepsilon)$) for linear (n,k) block codes was assumed to be upper bounded by $2^{-(n-k)}$. In papers published after this contribution of Leung and Hellman [1], various classes of codes were investigated with respect to probability of undetected error. This was done in order to determine which codes are proper and which are improper. Proper codes are those for which $P_u(\varepsilon)$ is a monotonically increasing function in ε, over $0 \le \varepsilon \le 0.5$. Codes for which $P_u(\varepsilon)$ is not a monotonic function in ε over $0 \le \varepsilon \le 0.5$ are termed improper. From the aforementioned one can gather that for proper codes $P_u(\varepsilon)$ is always bounded by $2^{-(n-k)}$ [1]. However, in investigations published previously, it was always assumed that errors occurred independently, i.e. the channel used is the Binary Symmetric Channel (BSC).

On many real communication channels such as the switched telephone network, radio links etc. errors do not occur independently but in bursts [2]. The equation,

$$P_u(\varepsilon) = \sum_{i=1}^{n} A_i \varepsilon^i (1-\varepsilon)^{n-i}, \qquad (1)$$

with A_i the weight enumerator of the (n,k) block code, only holds for the determination of $P_u(\varepsilon)$ on channels without memory, i.e. the BSC [3].

With this paper we intend presenting techniques aimed at determining $P_u(\varepsilon)$ for linear cyclic block codes on channels modelled by the well-known Fritchman and Gilbert-Elliott channel models [2].

When determining $P_u(\varepsilon)$ for codes on channels with memory, it is of utmost importance to know the positions of errors. Therefore, calculating $P_u(\varepsilon)$ of a code on a channel with memory, compels one to take into account the underlying error mechanism present on the channel. The error mechanism is typically modelled by means of discrete Markov chains such as the Gilbert-Elliott and Fritchman channel models.

In the first technique developed, we construct a trellis diagram of the states of a channel model such as the Gilbert-Elliott or Fritchman channel models. The length of the trellis is equal to the length of a codeword, n. Assume that the transmitted codeword is v and the received word is r with the error vector being e, giving $r = v + e$. Therefore, for a linear block code, whenever e is equal to a valid codeword, r is also a valid codeword. This is exactly the process that takes place whenever an undetected error occurs. This technique determines the probability of e being any one of the nonzero codewords of a code by determining the probability of occurrence of each codeword within a code except for the all-zero's codeword.

The second technique which we present involves the construction of a trellis diagram representing the states of the syndrome register of the code. The length of the trellis is also equal to the number of bits in a codeword, n. Syndrome calculation is usually performed in order to determine whether a received word is a codeword or not. Whenever a received word is not a codeword, the syndrome associated with it is non-zero, the syndrome being zero only if the received word is a valid codeword. This very principle is the basis upon which this particular technique for the determination of $P_u(\varepsilon)$ is based. After construction of the trellis diagram of the syndrome register states, all paths leading from the all-zero state back to the all-zero state in a number of transitions equal to code word length, n, are retained. The rest of the paths terminating in non-zero states after n transitions are discarded. The paths remaining in this way can now be associated with all valid codewords of a code. This reduced trellis diagram can now be used in conjunction with any binary channel model to determine $P_u(\varepsilon)$ for the code. The advantage of this technique is that $P_u(\varepsilon)$ can be determined easily for very long codes. It furthermore removes the need of knowing the weight spectrum of the code. The limiting factor in this case is the length of the syndrome register.

The first technique takes all codewords into account making it usable with smaller and very low rate codes, this being due to the fact that considering all codewords soon becomes very complex in larger high rate codes. The second technique is usable with high rate codes, seeing that not all codewords are considered and the complexity in this case is linear and not exponential as in the first technique.

REFERENCES

[1] S.K Leung-Yan-Cheong and M.E. Hellman, "Concerning a Bound on Undetected Error Probability," *IEEE Trans. Inform. Theory*, vol IT-22, pp. 235-237, Mar. 1976.

[2] L.N. Kanal and A.R.K. Sastry, "Models for Channels with Memory and their Applications to Error Control," *IEEE Proceedings*, Vol. 66, pp. 724-744, July 1978.

[3] S. Lin and D.J. Costello, *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1983.

SIMPLIFIED RECEPTION OF CONVOLUTIONALLY ENCODED CPM SIGNALS

Ryszard Bobrowski, Witold Hołubowicz

Franco-Polish School
of New Information and Communication Technologies
ul. P. Mansfelda 4, 60-854 Poznań, Poland

In this paper, we study the simplified reception of convolutionally encoded CPM signals. Our receiver is of the Viterbi type, but the number of receiver states is smaller than that of the optimum one. We use the concept of reduced state sequence estimation (RSSE), originally introduced by Eyuboglu et al. for the reception of signals in the ISI environment.

Summary

The block diagram of the system considered is shown in Fig. 1, where G is a convolutional encoder, M a finite state sequential machine that models a CPM modulator and V is a receiver of the Viterbi type, but the number of receiver states is smaller than that of the optimum one. In the receiver we use the concept of reduced state sequence estimation (RSSE), originally introduced by Eyuboglu at al. for the reception of signals in the ISI environment [1]. The idea of Eyuboglu has been then used by Svensson [2] and Huber [3] for the reception of uncoded CPM signals. In this paper we apply this concept to the reception of convolutionally encoded CPM signals. The receiver operates on the trellis which is reduced as compared to the combined trellis of the encoder and the modulator. In our paper, following the results of our earlier research [4,5], the unsimplified trellis for coded modulation has usually fewer states than the product of the number of modulator states and the number of encoder states. The states of the unsimplified trellis are grouped into, so called, superstates. The channel in our paper is assumed to be an ideal Gaussian one.

First of all, we show that the RSSE approach is applicable to the convolutionally encoded CPM signals. Then, the asymptotic error performance of selected coded CPM schemes is estimated by means of equivalent Euclidean distance calculated from a simplified trellis. Numerical results are presented for TFM and MSK signals combined with rate-1/2 short constraint length convolutional codes. The results are also compared with computer simulation. The concept of matched convolutional encoding combined with the simplified reception allowed us to find schemes which, for the same receiver complexity and bandwidth, outperform the schemes found so far by values of up to 1.2 dB.

Finally, we perform a code-receiver opimization procedure over the set of coded schemes with optimum and suboptimum receivers that should lead to the scheme with possibly lowest error probability, regardless of whether the receiver turns out to be optimum or not. In that respect we introduce the notion of the, so called "optimum transmitter" which is our search objecttive instead of the traditionally used "optimum receiver".

For example, Fig. 2 shows two systems that may be compared. In both cases the number of receiver states is the same. System B, even though employing a suboptimum receiver, would often outperform the system A in terms of error performance. Hence, under these new constraints, the optimization moves, to a large extent, to the transmitter side. In most cases examined by us, the most efficient solutions were found not to be based on the optimum receiver but rather on the RSSE receiver with the reduction factor F=2.

References

[1] M.V. Eyuboglu, S. U. Qureshi, "Reduced-State Sequence Estimation with set partitioning and Decision Feedback, "IEEE Tran. on Commun., vol. COM-36, pp. 13-20, Jan. 1988

[2] A. Svensson, "Reduced State Sequence Detection of Partial Response CPM", submitted to IEE Proceedings, Part I

[3] J. Huber, W. L. Liu, "An Alternative Approach to Reduced Complexity CPM-Receivers", IEEE JSAC, vol. SAC-7, pp. 1437-1449, Dec. 1989.

[4] W. Hołubowicz, R. Bobrowski, "Simple Receivers for Convolutionally Encoded Continuous Frequency Modulated Signals", presented on 9-th International Conference on Digital Satellite Communication, Copenhagen, Denmark, 18-20 May, 1992

[5] F. Morales-Moreno, W. Hołubowicz, S. Pasupathy, "Optimization of convolutionally encoded TFM signals via matched encoding", accepted for publication in the IEEE Trans. on Commun.

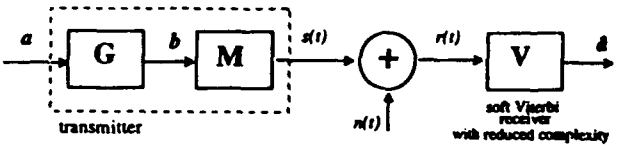


Fig.1. Block diagram of convolutionally encoded CPM communication system.

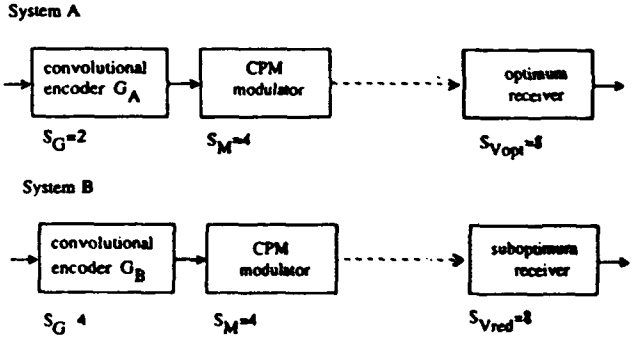


Fig 2. Comparison of two communication systems.

A Markov Analysis of Digital PLL Based MPSK Demodulators

Michael P. Fitz,* School of Electrical Engineering, Purdue University
West Lafayette, IN 47907-1285, (317)-494-0592, email: mpfitz@ecn.purdue.edu

This paper presents a statistical characterization of a uniformly sampled, first-order, decision-directed (DD), digitally implemented, phase-locked loop (DPLL) for MPSK modulations. This architecture is built in an extremely simple fashion and has near ideal coherent performance at moderate to high SNR. The phase detector (PD) presented in this paper has an ideal sawtooth form (ST-PD), but the analysis is easily modified to obtain results for the generalized Costas loops, the Mth power loop [3], or other loops for modulated signals. Figure 1 shows the analytical phase domain model for the decision-directed loop (note: $0< K \leq 1$ is the loop gain).

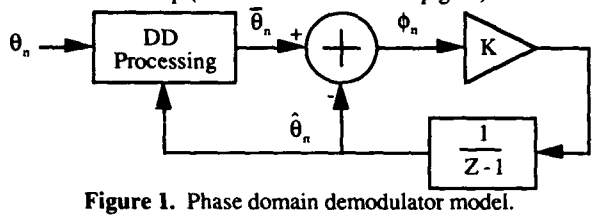


Figure 1. Phase domain demodulator model.

A discrete time Markov chain characterizes the DPLL and the associated PD. The equation describing the loop operation is

$$\begin{aligned}\hat{\theta}_{n+1} &= K\left(\bar{\theta}_n - \hat{\theta}_n\right) + \hat{\theta}_n \\ &= H\left(\hat{\theta}_n, \theta_n\right).\end{aligned} \tag{1}$$

Eq. (1) is a nonlinear equation since the phase additions and subtractions are modulo-2π operations. Assuming the input phase is a white random process (Nyquist prefiltering) then $\hat{\theta}_n$ is a first-order discrete-time Markov random process. The Chapman-Kolmogorov equation and an initial distribution function are sufficient to produce a complete statistical description of the loop operation.

A comparison of this Markov analysis with the traditional diffusion approximation is instructive. Traditionally the analysis of continuous time loops assume a narrow loop bandwidth and claimed that a diffusion approximation is valid [2, 3]. This discrete time analysis does not require a diffusion approximation and provides some advantages in analyzing synchronization systems. The advantages of this analytical technique are that no approximations are required and fast time variations of the phase error (e.g., wide bandwidth systems) can be analyzed. Table 1 summarizes the differences in the two analytical techniques.

Diffusion Approximation	Markov Chain
Continuous time model	Discrete time model
Fokker-Planck equation	Chapman-Kolmogorov equation
FP coefficients determined by the phase detector characteristics	State transition pdf determined by a transformation of random variables on the input noise
Valid for small loop bandwidths in comparison to the input noise bandwidth	Valid for all loop bandwidths
Valid for any prefilter	Input must be delta-correlated ⇨Nyquist prefiltering
Not valid for time-varying gain	Can examine time-varying loop gain
Not valid for looking at symbol-to-symbol phase error dependencies	Can examine symbol-to-symbol phase error dependencies

Table 1. Comparison of the diffusion approximation and the Markov chain model.

Steady-state performance is easily characterized with traditional Markov techniques. The chain is easily shown to be positive recurrent and asymptotically ergodic. An eigen-decomposition is used to evaluate the steady-state density function and the resulting bit error probability.

* This work partially supported National Science Foundation under Grant NCR-9115820

The acquisition performance is also easily characterized. This is accomplished using the traditional absorbing boundary/state techniques in Markov analysis. As expected, as $K \to 0$ the performance predicted by the discrete time analysis matches that predicted by the diffusion approximation [4]. Figure 2 shows the evolution of the phase error process during acquisition for a loop for an unmodulated input signal. Note that t is the time normalized by the loop bandwidth and the initial phase error was set to be $\varphi_0=180°$ which corresponds to the unstable attractor or the hangup point [1]. A major difference between the DPLL and the analog PLL acquisition performance is the effect of the hangup anomaly.

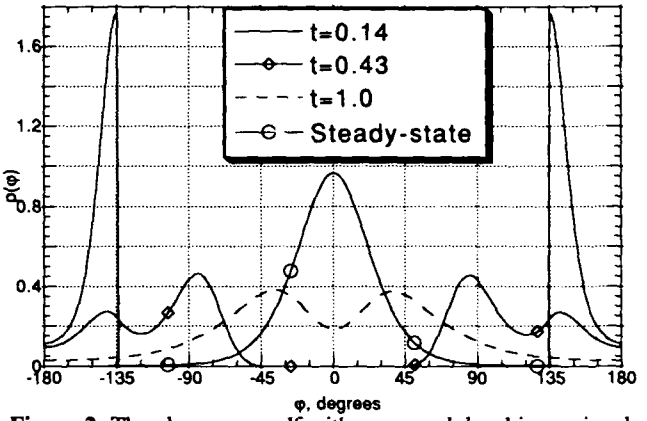


Figure 2. The phase error pdf with an unmodulated input signal during acquisition. $\varphi_0=180°$, K=0.25, ST-PD, and SNR_L=6dB.

Finally, the cycle slipping performance of the DPLL based demodulator is examined. Again, absorbing boundary techniques and some well known Markov process results [5] permit the characterization of the moments of the time to slip. Figure 3 presents the numerical results of a mean time to slip analysis for QPSK modulation.

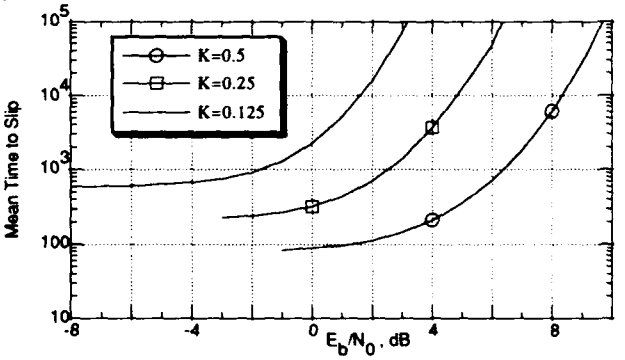


Figure 14. The mean time to cycle slip for the first-order DPLL for QPSK modulation versus loop SNR. Sawtooth PD.

References

[1] F.M. Gardner, "Hangup in Phase-Lock Loops," *IEEE Trans. Commun.*, vol. COM-25, October 1977, pp. 1210-1214.

[2] H.J. Kushner, "Diffusion Approximation to Output Processes of Nonlinear Systems with Wide-Band Inputs and Applications," *IEEE Trans. Inform. Theory*, vol. IT-26, November 1980, pp. 715-725.

[3] W.C. Lindsey and M.K. Simon, *Telecommunications Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[4] H. Meyr and L. Popken, "Phase Acquisition Statistics for Phase-Locked Loops," *IEEE Trans. Commun.*, vol. COM-28, August 1980, pp. 1365-1372.

[5] H.M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, Academic Press, Inc., Orlando, FL, 1984.

On the Applicability of the Fokker-Planck Method in Telecommunications (Summary)

L. Popken*

In the theoretical fields of telecommunications it can be observed that all too often the Fokker-PLanck (F-P) method is applied without paying (sufficient) attention to the physical foundations and to the conditions for the method's actual applicability.

A well-justified system analysis approach can be found by taking the actual equations, that describe the physical system in question, as a starting point after which one can think up approximations. Often researchers in technical literature attempt to formulate these approximations as if they were fundamental system descriptions. This, then, leads to approaches like the F-P equation being applied to processes that do not really satisfy the conditions for the applicability of the F-P method.

Continuous range Markov processes [1] contain a subclass of processes which in terms of the evolution of their pdf $p_x(x,t)$ are described by the F-P equation

$$\frac{\partial p_x}{\partial t} = -\frac{\partial}{\partial x}[K_1(x)p_x] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[K_2(x)p_x] \tag{1}$$

with two functions $K_1(x)$, $K_2(x)$ where $K_2(x) > 0$. This subclass of processes is defined by three conditions of which two are given by

$$\lim_{\Delta t\to 0}\frac{\langle \Delta x(t)\rangle}{\Delta t} = K_1(x) \tag{2}$$

$$\lim_{\Delta t\to 0}\frac{\langle (\Delta x(t))^2\rangle}{\Delta t} = K_2(x) \tag{3}$$

where x is the value of $x(t)$ at any time t, $\Delta x(t) = x(t+\Delta t) - x(t)$; the averages $\langle\cdots\rangle$ are taken with fixed $x(t)$. The third condition of

$$\lim_{\Delta t\to 0}\frac{\langle (\Delta x(t))^j\rangle}{\Delta t} = 0 \quad , j = 3,4,\ldots$$

was later replaced by the Lindeberg condition

$$\text{Prob.}\{|x(t+\Delta t) - x(t)| > \delta\} = O(\Delta t) \tag{4}$$

for any $\delta > 0$, [1].

The validity of the F-P equation is equivalent to the three conditions (2) to (4) being true for the process $x(t)$.

A major question arises in so far whether Markov processes with continuous sample paths actually exist in reality. For physical processes the Lindeberg condition (4) is at best satisfied only approximately, [1]. Therefore, the second order differential operator in the F-P equation is not a mathematical identity but an approximation only. In order to justify this approximation, a systematic expansion is required which, except for special cases, shows the F-P equation to be, in general, inconsistent because it includes terms of the order of magnitude as those terms which are neglected by omitting higher order derivatives. *Ad hoc* prescriptions for cutting off higher moments of the fluctuations seem often to be implied by (numerical) needs rather than by logic.

Markov processes with continuous sample paths do exist mathematically and can be useful in describing reality, provided that underlying conditions are proven to be adequately satisfied for the actual system in question.

In reality there is no such thing as a (continuous) Markov process. However, there may be driving processes with memory times so short that, on the time scale of interest, it is appropriate to consider the system process as well approximated by a Markov process. In this case, the question whether the sample paths are continuous or discontinuous is not relevant anymore. For deriving the pdf of the actual process the sample trajectories of the approximating Markov process are certainly *not* required to be continuous, although the physical process has almost surely continuous sample trajectories. This concept of *actual process versus Markov process* has been developed by Stratonovich in [2] which, however, has often been quite severely misinterpreted also in telecommunications literature.

For physical processes, Stratonovich has developed the pdf p_x as solution of the *kinetic* equation

$$\frac{\partial p_x}{\partial t} = \sum_{s=1}^{\infty}\frac{1}{s!}\left(-\frac{\partial}{\partial x}\right)^s[\mathbf{K}_s(x)p_x(x)] \tag{5}$$

where the individual intensity coefficients $\mathbf{K}_i(x)$ must be developed by separate expansions. In fact, for the pdf p_x a systematic two-dimensional expansion is required, i.e. the primary expansion w.r.t. x and the secondary expansion per intensity coefficient $\mathbf{K}_i(x)$ which is represented by its terms $K_{i,1}, K_{i,2}, K_{i,3}, \ldots, i = 1,2,3,\ldots$, [3].

If the physical process which drives a system, has a correlation time τ_{cor} much shorter than the system time constant τ_0, i.e. $\tau_{cor} \ll \tau_0$, and if the observation time interval $t - t_0$ is much longer than τ_0, then the intensity coefficients $\mathbf{K}_i(x)$ can be approximated by their corresponding first expansion terms, i.e. $\mathbf{K}_i(x) \approx K_{i,1}(x)$, $i = 1,2,3,\ldots$; the $\mathbf{K}_i(x), i = 2,3,\ldots$, become determined by the correlation functions of the actual wide-band driving process. This procedure is formally equivalent to the case of a mathematical, white driving process implying a Markov system process. Although the actual system process has almost surely continuous sample trajectories, it can in terms of its pdf formally be replaced, in general, by a discontinuous Markov process, [2].

It is the fundamental problem in several publications in the telecommunications area that Stratonovich's work [2] is misinterpreted such as $\tau_{cor} \ll \tau_0$ together with $(t - t_0) \gg \tau_0$ would be sufficient conditions for replacing the actual system process by a continuous Markov process to which then the F-P equation is applied, irrespective of the higher order correlation functions of the (non-Gaussian) noise.

The F-P equation (1) can provide correct results, in particular if it is applied to linear approximations; in these cases we restrict the system analysis to those features (such as low order moments of the system process) which coincide with the linear noise approximation. However, it is incorrect, as highlighted in [1], to consider the approach seriously beyond that, for instance to conclude the pdf

$$p_x(x) = \frac{\text{const}}{K_2(x)}\cdot\exp\left[-2\int_x\frac{K_1(x')}{K_2(x')}dx'\right] \tag{6}$$

which is the formal solution of the F-P equation (1).

References

[1] N.G. van Kampen, "The diffusion approximation for Markov processes," in *'Thermodynamics and Kinetics of Biological Processes* by L. Lamprecht, and Z.I. Zotin, Eds. Berlin: Walter de Gruyter 1983; (and references therein).

[2] R.L. Stratonovich, *'Topics in the Theory of Random Noise,'* Vol. I New York: Gordon and Breach, 1963.

[3] L. Popken, "On the applicability of the Fokker-Planck method in telecommunications," International Symposium on Information Theory and its Applications ISITA, Singapore, 16-20 Nov. 1992.

*European Space Research and Technology Centre, ESA/ESTEC, RF Systems Division, XRT, Keplerlaan 1, P.O. Box 299, 2200 AG Noordwijk, The Netherlands.

THE SYNCHRONIZATION GAME
PN Code Acquisition in presence of a White Noise, Average Power Constrained, Random, Symmetric Two-State Jammer

Jorge M. N. Pereira*
Communication Science Institute
EEB 500, University of Southern California
Los Angeles, CA 90089-2565

Abstract

The first, although restricted, solution of the Serial-Search PN Code Acquisition problem in presence of Slowly Time-Varying Fading Channels is herein presented, providing a determinant performance measure for Spread Spectrum Systems under these conditions. As a case study, a white noise, average power constrained, random, symmetric two-state Jammer is analyzed, and the corresponding minimax threshold is determined.

Summary

The solution of the PN Code Acquisition problem in presence of Time-Varying Fading Channels is one of the most important open problems in the area of Communications, especially in light of the present trend towards mobility.

The first, although restricted, solution of the Serial-Search PN Code Acquisition problem in presence of Slowly Time-Varying Fading Channels which can be discretely approximated is herein presented, thus providing a much needed performance measure for Spread Spectrum Systems under these taxing conditions.

Two distinct approaches led to the exact solution of the approximated (i.e., discretized) problem. One of them, original and making full use of symbolic computational capabilities now available, was found to be particularly advantageous for large uncertainty regions (i.e., large number of states in the state transition diagram). A new, quite good approximate solution, requiring even (much) less computations, was also found.

*This work was partially supported by NSF PYIA grant NCR-8552527, and by grants from *Fundação Calouste Gulbenkian*, and *Fundação Luso-Americana para o Desenvolvimento*, under sponsorship of the *Instituto Nacional de Investigação Científica*. The author is on leave from the *Centro de Análise e Processamento de Sinais*, Departamento de Engenharia Electrotécnica & Computadores, Instituto Superior Técnico, 1096 LISBOA CODEX, PORTUGAL.

The theory is then applied to a White Noise Two-State Jammer which randomly alternates betwen two equally likely jamming levels in order to satisfy an average power constraint. In this case, no approximation is involved, since the channel is intrinsecally discrete.

The results seem to indicate that at high SNRs, the On-Off Jammer, alternating periods of radio silence what periods of total jamming (i.e., reserving all the available power for the jamming ocasions), is the worst possible in what refers to acquisition. For low SNRs a more elaborate Jammer, alternating between the On-Off mode and the Continuous mode (i.e., jamming at the average power), is the worst possible. The receiver, by appropriately setting its threshold, can now play the minimax synchronization game.

References

[1] D.C. Cox, "Universal Digital Portable Radio Communications", *Proceedings of the IEEE*, Apr 87

[2] J.M. Pereira, "Study of the effects of Fading in CDMA Code Acquisition in the Personal Communication Services Environment", **Internal Report**, *Communication Sciences Institute*, University of Southern California, Jan 91

[3] A. Polydoros, *On the Synchronization Aspects of Direct-Sequence Spread Spectrum Systems*, Ph.D. Thesis, University of Southern California, Aug 82

[4] A. Polydoros, C.L. Weber, "A Unified Approach to Serial Search Code Acquisition", *IEEE Trans. on Comm.*, May 84

[5] P.M. Hopkins, "A Unified Analysis of Pseudo-Noise Synchronization by Envelope Correlation", *IEEE Trans. on Comm.*, Aug 77

PERFORMANCE ANALYSIS OF MPPM IN NOISY PHOTON COUNTING CHANNEL

Tomoaki Ohtsuki, Iwao Sasase, and Shinsaku Mori

Department of Electrical Engineering, Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223, Japan

Recently, there has been considerable interest in multi-pulse pulse position modulation (MPPM), because MPPM reduces the transmission bandwidth to about half that of pulse position modulation (PPM) [1]. In [2], the cutoff rate and the capacity of MPPM in noiseless photon counting channel are derived and MPPM is shown to outperform PPM in terms of both cutoff rate and capacity. The optical channel is often modeled by noiseless photon counting channel, but noise due to background and detector dark currents exists on the practical optical channel. However, the performance such as symbol error rate and bit error rate of MPPM in noisy photon counting channel has not been analyzed yet. In this paper, we analyze the performance of MPPM in noisy photon counting channel and propose interleaved convolutional coded MPPM system in order to reduce the average transmitter power. It is shown that the proposed system can reduce the average transmitter power compared with uncoded MPPM.

In MPPM, the laser is pulsed in p slots in one signal block consisting of m slots with duration τ, and $J=\binom{m}{p}$ pulse patterns can be formed by combining the positions of optical pulses. The optical channel is well modeled by Poisson statistics, under which the output of the channel is a doubly stochastic Poisson process with intensity $\lambda_s(t)+\lambda_n$ where $\lambda_s(t)$ is the mean rate in photons per second due to the signal impinging on the photodetector and λ_n is the noise intensity due to background and detector dark currents.

First we derive the probability of symbol error of MPPM in noisy photon counting channel. The probability of symbol error is given by

$$P(e) = \frac{1}{J}\sum_{i=1}^{J} P(e/i) \tag{1}$$

where $P(e/i)$ is the probability of symbol error when the symbol i is sent. Assuming that in the case of equal symbol counts between the correct symbol and some other symbol a wrong decision is made, we have

$$P(e/i) \le Pr[\bigcup_{j\neq i}\{N_i \le N_j\}/i] \le \bigcup_{j\neq i} Pr[N_i \le N_j/i], \tag{2}$$

where the second inequality is justified by the union bound. And we have

$$Pr[N_i \le N_j/i] = Pr[N_i' \le N_j'] \tag{3}$$

where N_i' and $\le N_j'$ are independent Poisson random variables with means $d_{ij}(\lambda_n+\lambda_s)\tau$ and $d_{ij}\lambda_n\tau$, respectively, and d_{ij} is the distance between symbol i and symbol j. The distance d_{ij} is defined as $d_{ij}=p-v$ where v is the number of overlapped pulses between symbol i and symbol j. It follows that

$$Pr[N_i \le N_j/i] = Q_1[\sqrt{2d_{ij}\lambda_n\tau}, \sqrt{2d_{ij}(\lambda_n+\lambda_s)\tau}] \tag{4}$$

where $Q_1(\alpha,\beta)$ is Marcum's Q-function. Eq. (4) can be simplified by using a Chernoff bound as

$$Pr[N_i \le N_j] = \exp[-d_{ij}(\sqrt{(\lambda_n+\lambda_s)\tau} - \sqrt{\lambda_n\tau})^2]. \tag{5}$$

Combining Eqs. (1), (2), (4) and (6), we obtain

$$P(e) \le \frac{1}{J}\sum_{i=1}^{J}\sum_{j\neq i}\exp[-d_{ij}(\sqrt{(\lambda_n+\lambda_s)\tau} - \sqrt{\lambda_n\tau})^2]. \tag{6}$$

Defining μ as the number of signal photons per information nat, we have $\lambda_s\tau = \mu R\ln(J)$ where R is the code rate. Next we show the bit error probability of MPPM in noisy photon counting channel. By using the probability of symbol error, the bit error probability P_b is approximately bounded by $P_b < [2^{L-1}/(2^L-1)]P(e)$ where L is the maximum integer satisfying $L \le \log_2\binom{m}{p}$. Figure 1 shows the bit error probability of MPPM in noisy photon counting channel as a function of signal energy μ in photons/nat. The probability of pulse occurrence is selected to be same among all schemes. It is found that increasing m and p improves the performance of MPPM. Similar trends are obtained for the noiseless case with $\lambda_n\tau$=0.0, because MPPM can form more symbols than PPM at the same probability of pulse occurrence.

In order to reduce the average transmitter power, we propose interleaved convolutional coded MPPM system in noisy photon counting channel. Each block of L input bits is fed into L parallel encoders. The encoded bits are properly interleaved and each of L bits is sent with (m,p) MPPM. On the decoding side, L parallel Viterbi decoding are employed and the symbol is hard-demodulated with deciding p slots in order of maximum counts in each frame. In this case, we model (m,p) MPPM channel as a parallel combination of binary symmetric channel (BSC) with transition probabilities q and $1-q$ given by $q \le [2^{L-1}/(2^L-1)]P(e)$. In binary using case, 2^L symbols which have the best distance characteristics are selected from $J=\binom{m}{p}$ symbols. Therefore the transition probability q is bounded by the above equation. Using the union bound on the first-event error probability, it can be shown that the bit error probability P_b for a rate $R=1/n$ convolutional code is bounded by $P_b < \sum_{h=d_{free}}^{\infty} W_h P_2(h)$ where W_h is the number of bit errors contributed by the incorrect paths which are at distance h from the correct path, and d_{free} is the minimum free distance of the code. For the BSC, the pairwise error probability $P_2(h)$ is upper bounded by $P_2(h) < \{2|q(1-q)|^{1/2}\}^h$. Figure 2 shows the bit error probability of the proposed system with $R=1/2$ and constraint length k convolutional codes where P_b is approximated by the error-event probability. It is found that the proposed system can greatly reduce the average transmitter power compared with uncoded MPPM in noisy and noiseless cases. It is also found that the system with larger constraint length k has better performance because of its higher error correction ability. For example, the proposed system with k=7 reduces signal energy μ to achieve $P_b = 10^{-9}$ over uncoded MPPM form 12.0 to 4.2 in the noisy case with $\lambda_n\tau = 1.0$ and from 8.3 to 2.3 in the noiseless case with $\lambda_n\tau = 0.0$. Therefore the proposed system is effective to reduce the average transmitter power.

References

[1] H. Sugiyama and K. Nosu, "MPPM : A method for improving the band-utilization efficiency in optical PPM," *J. Lightwave Technol.*, Vol. LT-7, no. 3, pp.465-472, Mar. 1989.

[2] T. Ohtsuki, H. Yashima, I. Sasase, and S. Mori, "Cutoff rate and capacity of MPPM in noiseless photon counting channel," *IEEE Pacific Rim Conf.*, Victoria, Canada, May. 1991.

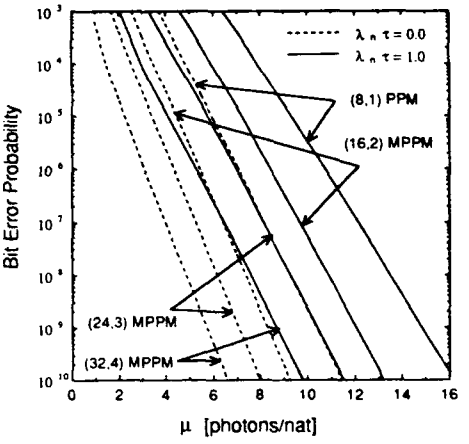


Fig.1. BER for MPPM as a function of μ photons/nat.

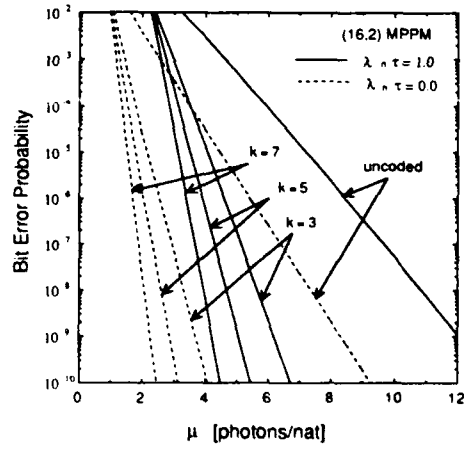


Fig.2. BER for interleaved convolutional coded MPPM as a function of constraint length k and μ photons/nat.

POWER MOMENTS OF EXPONENTIAL FUNCTIONALS OF BROWNIAN MOTION

Yeheskel E. Dallal and Shlomo Shamai (Shitz)

Department of Electrical Engineering
Technion—Israel Institute of Technology, Haifa 32000, Israel

The distribution law of the Brownian Motion exponential functional:

$$\boldsymbol{\varepsilon} = \left| \int_0^1 \exp\left[j\sqrt{2\beta}\mathbf{W}_t \right] dt \right|^2 \quad ,$$

where $\{\mathbf{W}_t, \quad t \in \mathbb{R}^+\}$ is a standard Brownian Motion, which models the laser's phase noise, plays a key role in many of the recently proposed heterodyne lightwave communication systems. The exact derivation of these statistics appears, however, a formidable mathematical task [1]. Invoking a signal flow graph formulation and combinatorial arguments, leads to a simple and computationally efficient closed form formula for the k'th power moment $\mathbb{E}\boldsymbol{\varepsilon}^k$ induced by the unknown distribution law. These results are useful in bounding the performance of a variety of lightwave communications systems operating in presence of phase noise [2],[3].

The expression for $\mathbb{E}\boldsymbol{\varepsilon}^k$ is given in terms of a $((k+1)^2 - 1/2(k+1)k)$-*fold summation*

$$\mathbb{E}\boldsymbol{\varepsilon}^k = \frac{(k!)^2}{\beta^{2k}} \sum_{m=0}^{k} \sum_{n=0}^{k-m} \frac{c(m,n)}{n!} \beta^n e^{-m^2\beta}, \quad k \geq 1 \quad ,$$

where $c(m,n)$ is a *rational coefficient given by*

$$c(m,n) = \sum_{\substack{a_1,\ldots,a_r \\ \max(m,1)\leq r\leq k \\ a_0=1,\, a_{r+1}=0 \\ \Sigma_1^r a_i = k,\, \{a_i \geq 1,\, 1\leq i\leq r\} \\ a_m + a_{m+1} \geq n+1}} \frac{2^{a_1} \prod_{i=1}^{r} \binom{a_{i-1} - 1 + a_i}{a_i}}{(a_m + a_{m+1} - n - 1)!} \cdot \frac{\partial^{a_m+a_{m+1}-n-1}}{\partial s^{a_m+a_{m+1}-n-1}} \left\{ \prod_{\substack{i=0 \\ i\neq m}}^{r} \frac{1}{(s + i^2 - m^2)^{a_i+a_{i+1}}} \right\} \Bigg|_{s=0} .$$

The computational complexity of $c(m,n)$ is of an exponential order in k rather than a factorial order characterizing previously reported methods. We address also the efficient derivation of $\mathbb{E}\boldsymbol{\varepsilon}^k$ given the set of the preceding moments $\left\{\mathbb{E}\boldsymbol{\varepsilon}^l\right\}_{l=0}^{k-1}$ motivated by the fact that power moment statistical characterization often requires the availability of a finite set of consecutive moments $\left\{\mathbb{E}\boldsymbol{\varepsilon}^k\right\}_{k=0}^{K}$ [2],[3]. It is shown that $\mathbb{E}\boldsymbol{\varepsilon}^k$ is readily given as the convolution of the preceding moments with a set of known casual functions.

References

[1] G.J. Foschini and G. Vannuci, "Characterizing filtered lightwaves corrupted by phase noise," *IEEE Trans. on Inform. Theory*, Vol. 34, No. 6, pp. 1437–1448, Nov. 1988.

[2] Y.E. Dallal and S. Shamai (Shitz), "An upper bound on the error probability of quadratic-detection in noisy phase channels," *IEEE Trans. on Commun.*, Vol. 39, No. 11, pp. 1565–1650, Nov. 1991.

[3] Y.E. Dallal and S. Shamai (Shitz), "Performance bounds for noncoherent detection under Brownian phase noise," *IEEE Trans. on Inform. Theory*, Vol. 38, No. 2, pp. 362–379, March 1992.

A New Structured Quantizer for Sources with Memory †

Rajiv Laroia, Cheng-Chieh Lee and Nariman Farvardin
Electrical Engineering Department
and Institute for Systems Research
University of Maryland
College Park, Maryland 20742

Summary

For high quantization rates, Lookabaugh and Gray have demonstrated that the advantages of optimum vector quantization over optimum scalar quantization can be separated into the *space filling* advantage, the *shape* advantage, and the *memory* advantage [1]. Eyuboğlu and Forney have proposed a lattice-based VQ in which the codebook consists of all the lattice points that lie inside a suitably chosen support region [2]. They showed, for memoryless sources, there are two significant gains — the boundary gain and the granular gain — that lattice-based vector quantizers realize over uniform scalar quantizers. For memoryless sources, the scalar-vector quantizer of Laroia and Farvardin [3] can asymptotically (in block-length) realize the optimal boundary gain. It however realizes no granular gain. The trellis coded quantizer of Marcellin and Fischer [4], on the other hand, can realize a significant portion of the total granular gain, but makes no explicit attempt to capitalize on the boundary gain. Recently, Laroia and Farvardin have combined the above two ideas and proposed a fixed-rate trellis-based scalar-vector quantizer (TB-SVQ) [5], which realizes both boundary and granular gains.

The TB-SVQ is the dual of optimally-shaped trellis-coded constellation for transmission over memoryless channels [6]. Laroia, Tretter and Farvardin have recently proposed a precoding scheme [7] that solves the problem realizing both shaping and coding gains for transmission over intersymbol interference channels. In this paper, we combine the TB-SVQ idea of [5] and the precoding idea of [7] to develop a quantization scheme for correlated sources.

We assume the source $\{X_n\}$ is the output of a linear pth-order autoregressive (AR) filter $H(z)$ driven by a stationary memoryless innovation process $\{W_n\}$ where $H(z) = 1/(1-\sum_{i=1}^{p}\rho_i z^{-i})$. We propose a quantization scheme whose block diagram is shown in Fig. 1. This quantizer — referred to as the *precoded quantizer* (PQ), motivated by the aforementioned precoding scheme, combines the precoder (to remove the source correlation) and the TB-SVQ (to achieve both boundary and granular gains).

The trellis code and SVQ share the common underlying scalar alphabet $\mathcal{Q} = \{\cdots, -3\beta, -\beta, \beta, 3\beta, \cdots\}$. The source sequence $\{X_n\}$ is first mapped to the coset trellis sequence $\{A_n\}$, which serves as a candidate quantization sequence. To check if a block of samples of $\{A_n\}$ is inside the codebook, $\{A_n\}$ is decorrelated to $\{B_n + Q_n\}$ where $\{B_n\}$ is a valid trellis sequence (congruent to $\{A_n\}$) and $\{Q_n\}$ is some noise sequence that is confined to the Voronoi region of the coset lattice of the trellis code. The coset quantizer removes $\{Q_n\}$ and produces $\{B_n\}$. For high-rate quantization, $\{A_n\}$ is a good approximation to $\{X_n\}$ and the energy of $\{Q_n\}$ can be ignored, therefore $\{B_n\}$ is a good approximation to $\{W_n\}$. A TB-SVQ designed for $\{W_n\}$ is used here for encoding $\{B_n\}$. The SVQ encoder takes a block of samples from $\{B_n\}$ and decides if the vector lies inside the codebook (defined in the innovation domain). If it does, the TB-SVQ encoding algorithm is used to encode the vector. If not, the vector $\{X_n\}$ is gradually moved closer to the codebook boundary by considering the geometric shape of the boundary induced by the distribution of $\{W_n\}$. This is repeated until the corresponding block of $\{B_n\}$ is inside the codebook. In the decoder, assuming no channel errors occur in the channel, the output of the TB-SVQ decoder is $\{B_n\}$. The precoding scheme of [7] is used here to generate the quantization sequence $\{A_n\}$.

The shape of the codebook is defined in the innovation domain. The corresponding codebook shape in the source domain is not necessarily optimal. This is due to the occurrence of the overhead noise $\{Q_n\}$. This is analogous to the increase in the transmitted power for the precoding scheme described in [7]. Reducing the energy of $\{Q_n\}$ should improve the performance of the PQ. This can be done by using higher dimensional trellis codes.

The performance of PQ for quantizing Gauss-Markov sources was obtained via simulations using 100 sequences of 32,000 samples each. The trellis decoding delay is 100 samples. Table 1 summarizes the performance (SNR in dB) of the PQ for encoding an AR(1) Gauss-Markov source ($\rho_1 = 0.9$) for dimensions 32 and 64 at rates 2,3 and 4 bits/sample. As expected, the PQ using 2D trellis code performs better than using 1D trellis code. A two-codebook [5] 64-dimensional PQ using a 2D trellis code at ra 3 bits/sample yields 0.04 – 0.06 dB performance improvement. This illu trates the advantage of applying codebooks with variable density. Compar to the predictive trellis coded quantizer (PTCQ) [4], for the same numb of trellis states (4,8,16 and 32), the 64-dimensional PQ (with 100 sampl delay) performs 0.5-1.5 dB better than PTCQ (with 1000 samples dela for rate 3 bits/sample. For rate 2 bits/sample, the performances are abo the same.

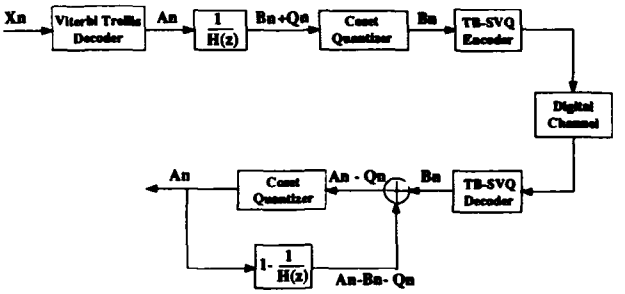


Figure 1: Block diagram of the precoded quantizer.

m	r	Trellis State									$D(R)$
			1D				2D				
		1	4	8	16	32	4	8	16	32	
32	2	16.76	16.53	16.72	16.89	16.99	17.36	17.48	17.54	17.60	19.25
	3	23.02	23.61	23.71	23.77	23.84	23.75	23.83	23.86	23.90	25.27
	4	28.68	29.49	29.56	29.61	29.65	29.53	29.60	29.64	29.67	31.29
64	2	16.88	16.65	16.81	16.91	17.07	17.63	17.72	17.78	17.84	
	3	23.26	23.86	23.95	24.01	24.08	24.04	24.10	24.14	24.18	
	4	28.95	29.85	29.93	29.99	30.03	29.79	29.85	29.89	29.91	

Table 1: Performance (SNR in dB) of the PQ for an AR(1) Gaus Markov source ($\rho_1 = 0.9$).

References

[1] T. Lookabaugh and R. Gray, "High-resolution quantization theory an the vector quantizer advantage," *IEEE Trans. Inform. Theory*, vol. I 35, pp. 1020–1033, September 1989.

[2] V. Eyuboğlu and G. D. Forney, Jr., "Lattice and trellis quantizatic with lattice- and trellis-bounded codebooks — Part I : High-rate theo for memoryless sources," *Submitted to IEEE Trans. Inform. Theor* December 1990.

[3] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantiz derived from a variable-length scalar quantizer — Part I:Memoryle sources," *IEEE Trans. Inform. Theory*, to appear.

[4] M. Marcellin and T. Fischer, "Trellis coded quantization of memor less and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. COM-3 pp. 82–93, Jan. 1990.

[5] R. Laroia and N. Farvardin, "Trellis-based scalar-vector quantization f memoryless sources," *Submitted to IEEE Trans. Inform. Theory*, Jun 1992.

[6] R. Laroia, "On optimal shaping of multidimensional constellations - an alternative approach to lattice-bounded (Voronoi) constellations *Submitted to IEEE Trans. Inform. Theory*, November 1991.

[7] R. Laroia, S. Tretter, and N. Farvardin, "A simple and effective preco ing scheme for noise whitening on intersymbol interference channels *IEEE Trans. Commun.*, to appear.

†This work was supported in part by National Science Foundation grants NSFD MIP-91-09109 and CD-88-03012.

OPTIMAL VECTOR QUANTIZED NONLINEAR ESTIMATION†

A. Gersho

Center for Information Processing Research
Dept. of Electrical & Computer Engineering
University of California,
Santa Barbara, CA 93106

Suppose we wish to estimate a random vector Y from a random vector X with an estimator $h(\cdot)$ that is constrained to take on a finite set of N values. For the mean squared error criterion, the optimal nonlinear estimator $h(x)$ is given by the cascade of the optimal unconstrained estimator $g(x) = E[Y \mid X = x]$ followed by the optimal vector quantizer for the random vector $g(X)$. The vectors X and Y may have different dimensions. We view $h(x)$ as a *generalized vector quantizer* which optimally generates a quantized approximation to Y from observation of X. The special case where $X = Y + W$ and W is independent additive noise was studied by Wolf and Ziv [1] and Ephraim and Gray [2]. Sakrison [3] considered the more general formulation of source encoding in the presence of a random disturbance.

Since $h(x)$ has finite range, its domain can be partitioned into N sets, S_i, each the pre-image of a range value, where N is the size of the range set. It is readily shown that

(a) the optimal range values $\{y_i\}$ for a given partition are given by $y_i = E[g(X) \mid X \in S_i]$, and

(b) the optimal partition regions given the range values are: $S_i = \{x : \|g(x) - y_i\| \leq \|g(x) - y_j\| \quad for\ all\ j\}$ ignoring boundary values.

In practice, design of $h(x)$ can be based on a large set of empirical data pairs (X, Y) as a statistical specification of the random vectors. In general, the domain regions S_i are neither convex nor connected sets. Thus, conventional vector quantizer design methods are inadequate. The optimal $h(x)$ must therefore be implemented as a pattern classifier (an encoder) that maps the input X to an index i followed by a decoder, a table-lookup operation with i as input.

The above formulation and resulting design methodology offers a notable improvement to a useful paradigm in vector quantization (VQ), called *nonlinear interpolative vector quantization* (NLIVQ). The basic theory of NLIVQ was introduced in [4] and has found several applications, including multiresolution image compression [5], multispectral image compression [6], nonlinear prediction of speech [7], wideband audio compression [8], and enhanced decoding of standard transform coded images [9]. In NLIVQ, a signal vector Y of dimension m is mapped by a feature extractor into a vector X of dimension k (usually $k \leq m$) which is then VQ encoded, producing an index (channel symbol) I; unlike ordinary VQ, the decoder directly reconstructs an approximation to Y (rather than to X) by a table lookup with a codebook of dimension m. In the special case where X is a subsampled version of Y, the decoder can perform optimal interpolation of Y from a digital representation of X. Until now, NLIVQ was based on an optimal decoder for a given VQ encoder. Here we see that NLIVQ can be improved by *jointly* optimizing the encoder (which digitizes X) and the decoder. This follows as an immediate application of the problem of optimal nonlinear estimation with finite range, posed in the first paragraph above. The performance achievable with NLIVQ is thereby improved and the applicability and utility of NLIVQ is correspondingly enhanced.

†This work was supported by the National Science Foundation, the UC Micro Program, Rockwell International, Hughes Aircraft, Eastman Kodak, Compression Labs, and Fujitsu Labs.

References

1. J.K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. on Inform. Theory*, vol. 16, 406-411, July 1970.
2. Y. Ephraim and R.M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inform. Theory*, vol. 34, pp. 826-834, July 1988.
3. D.J. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. Inform. Theory*, vol. 32, pp. 165-167, Jan. 1968.
4. A. Gersho, "Optimal Nonlinear Interpolative Vector Quantization," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1285-1287, September, 1990.
5. Y.S. Ho and A. Gersho, "A Variable Rate Image Coding Scheme Using Vector Quantization and Clustering Interpolation," *Conf. Record, IEEE Global Commun. Conf.*, pp. 898-902, November 1989.
6. S. Gupta and A. Gersho, "Feature Predictive Vector Quantization of Multispectral Images," *IEEE Trans. Geoscience Electronics*, vol. 30, May 1992.
7. S. Wang, E. Paksoy, and A. Gersho, "Performance of Nonlinear Prediction of Speech," *Proc. Int. Conf. Spoken Language Processing*, Kobe, Japan, November 1990, pp. 29-32.
8. W.Y. Chan and A. Gersho, "Constrained-Storage Vector Quantization in High Fidelity Audio Transform Coding," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada pp. 3597-3600, May 1991.
9. S.W. Wu and A. Gersho, "Enhancement of Transform Coding by Nonlinear Interpolation", *1991 SPIE Conf.* Visual Commun. Imag. Processing, Boston, pp. 487-498, November, 1991.

ROBUST VECTOR QUANTIZATION BY LINEAR MAPPINGS OF BLOCK-CODES

Roar Hagen and Per Hedelin

Department of Information Theory
Chalmers University of Technology
S-41296 Göteborg, Sweden

1. INTRODUCTION

Vector Quantization (VQ) is an important method for source coding and signal compression [1]. VQ is, however, not without problems, e.g. robustness and complexity. Structured VQ [2] and channel optimized VQ [3] are areas of current research, aimed to overcome some of these problems. In this paper, we introduce a new flexible concept of VQ that is robust both in terms of training databases and channel errors. The method can also be used to achieve reduced storage requirements and search complexity.

2. THE VECTOR QUANTIZER

Consider a VQ constrained by the following linear mapping

$$\mathbf{c} = \mathbf{T}\cdot\mathbf{b} + \mathbf{m} \tag{1}$$

where $\mathbf{c}$ is the reconstruction vector of dimension D generated by a code-vector $\mathbf{b}$. Let the code-vector stem from a binary systematic block-code

$$\mathbf{b} = (i_1,\ldots,i_k,c_1,\ldots,c_r)^T \tag{2}$$

where the k binary elements i_j are information bits and the the c_js are r redundant modulo-2 sums (parity bits) of the information bits. Thus, $\mathbf{b}$ is a code-vector in a $(k+r,k)$ linear block-code [4]. The 2^k different code-vectors generate an equal number of reconstruction vectors. We emphasize that only the information bits have to be transmitted in an application, the redundant bits are introduced to control the degrees of freedom in the mapping.

To achieve a suitable representation for signal-vector generation purposes, binary 0 is represented by +1 and binary 1 is represented by -1 in $\mathbf{b}$. Hence, the points $\mathbf{b}$ constitute the ordered subset of the corners of a $k+r$ dimensional cube, as specified by the block-code. The projection matrix $\mathbf{T}$ is of dimension $D\times(k+r)$ and the D-dimensional vector $\mathbf{m}$ represents the mean-value of the source. For a given source to quantize, both the projection, $\mathbf{T}$, and the block-code must be selected. For a given block-code, an optimization involves the $D\times(k+r+1)$ parameters of $\mathbf{T}$ and $\mathbf{m}$.

In order to compute an optimized mapping for an arbitrary source, we use an iterative training technique with a database of representative source vectors, $\mathbf{x}$. We adopt the squared Euclidean distance as distortion measure and restrict the discussion to zero-mean sources ($\mathbf{m}=\mathbf{0}$). The measure to be minimized is then

$$\alpha_T = E\left[\left\|\mathbf{x} - \mathbf{T}\cdot\mathbf{b}(\mathbf{x})\right\|^2\right] \tag{3}$$

where $\mathbf{b}(\mathbf{x})$ denotes the code-vector used to generate the reconstruction vector of a certain source vector $\mathbf{x}$. Minimizing this with respect to $\mathbf{T}$ gives an expression for the rows $\mathbf{t}_j$ of $\mathbf{T}$

$$E\left[\mathbf{b}(\mathbf{x})\cdot\mathbf{b}^T(\mathbf{x})\right]\cdot\mathbf{t}_j = E\left[x_j\cdot\mathbf{b}(\mathbf{x})\right] \quad j=1,\ldots,D \tag{4}$$

where x_j is component j of $\mathbf{x}$. We are now able to devise an block-iterative algorithm for computation of the mapping:

(i) Initialize the matrix $\mathbf{T}$.
(ii) Find the nearest reconstruction vector for each vector $\mathbf{x}$ of a training database (i.e. $\mathbf{b}(\mathbf{x})$). Compute the correlations in eq. (4) for each row j.
(iii) Solve eq. (4) for the D rows of $\mathbf{T}$. Evaluate the distortion α_T.
(iv) Repeat from (ii) until end of training.

3. PROPERTIES

This way of representing a Vector Quantizer has a number of desirable properties.

(i) Few parity bits are needed.

By using every available parity bit, $r = 2^k - k - 1$, there are $D\times 2^k$ free parameters in $\mathbf{T}$ and $\mathbf{m}$, i.e. we are able to generate an arbitrary set of reconstruction vectors. This gives an unconstrained VQ. For a given application, we can choose any number of parity bits between this maximum and zero. A main result of this paper is that by using only a few parity bits, or even none, one obtains a robust result close to the unconstrained case.

(ii) Robustness against channel-errors.

Eq. (1) can be expressed as

$$\mathbf{c} = \sum_{j=1}^{k} i_j\cdot\mathbf{u}_j + \sum_{j=1}^{r} c_j\cdot\mathbf{u}_{k+j} \tag{5}$$

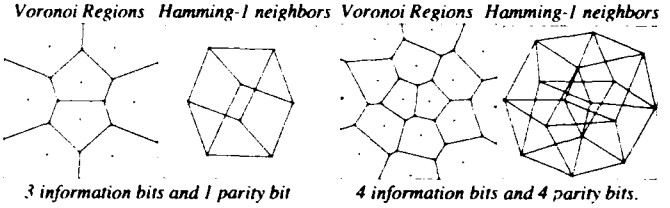
where $\mathbf{u}_j$ is column j of $\mathbf{T}$. We define the weights of the bits as the length of the corresponding vector $\mathbf{u}_j$ in (5). Robustness against channel-errors requires that neighboring reconstruction vectors $\mathbf{c}$ have small Hamming distances in the information part of the corresponding code-vectors $\mathbf{b}$. A second major result in this paper is that good neighboring properties in the VQ are obtained by assuring small weights in the redundant part of the code. The information part shall have high and fairly uniform weights. The result is a VQ with inherent robustness against channel errors.

(iii) Fast and Robust training.

The initialization of the matrix $\mathbf{T}$ is an important issue for convergence of iterations and channel-error robustness. We have devised several efficient techniques for initialization based on assigning low weights to the redundant bits. Note, moreover, that during the iterations, the proposed algorithm takes second-order effects into account in each adjustment of the VQ by solving for eq. (4). By using few redundant bits, the number of parameters to adjust is low and, hence, robustness against variability in training databases is ensured.

4. RESULTS

Figure 1 below shows reconstruction vectors with associated Voronoi-regions and a plot with vectors at Hamming-distance one connected for two 2-dimensional cases designed by the proposed method.



3 information bits and 1 parity bit *4 information bits and 4 parity bits.*

Figure 1. *Illustration of results for a 2-dimensional memoryless Gaussian source.* ***Left part:*** *Linear mapping of a (4,3) block-code.* ***Right part:*** *Linear mapping of a (8,4) block-code.*

Table 1 and 2 below illustrate some results, in terms of Signal-to-Noise ratios, obtained for a memoryless Gaussian source and a Gauss-Markov source with correlation 0.5.

Table 1. Memoryless Gauss source.

D	k	r	SNR	D	k	r	SNR
2	3	0	6.92	3	3	0	4.40
		1	6.96			3	4.48
		4	6.96			4	4.48
	4	0	9.57	4	4	0	4.39
		2	9.60			2	4.51
		11	9.68			11	4.60

Table 2. Gauss-Markov source 0.5.

D	k	r	SNR	D	k	r	SNR
2	3	0	7.56	3	3	0	5.48
		1	7.64			1	5.49
		4	7.66			4	5.49
	4	0	10.06	4	4	0	5.65
		2	10.27			2	5.69
		11	10.32			11	5.69

The results are given in dB and are obtained as the mean-value of evaluation over 3 independent data-bases, each of size 500 000 vectors.

5. REFERENCES

[1] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston 1992.
[2] T. R. Fischer, M. W. Marcellin and M. Wang, "Trellis-Coded Vector Quantization", IEEE Trans. Inf. Theory, Vol. 37, no 6, November 1991.
[3] N. Farvardin, "A Study of Vector Quantization for Noisy Channels", IEEE Trans. Inf. Theory, Vol. 36, no 4, July 1990.
[4] R. E. Blahut, *Theory and Practice of Error Control Codes*, Addison-Wesley , Reading 1983.
[5] Y. Linde, A. Buzo and R. Gray, "An Algorithm for Vector Quantizer Design," IEEE Trans. Comm., Vol. COM-28, January 1980.

AN IMPROVED TREE-STRUCTURED VECTOR QUANTIZER

David Miller and Kenneth Rose*
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Tree-structured vector quantization (TSVQ) is an important alternative to full-search vector quantization since it reduces the encoding search complexity from $O(N)$ to $O(log(N))$, where N is the decoder codebook size, while incurring some additional distortion. Typically, the methods for designing trees are greedy, growing the trees outward from the root, node-by-node, by minimizing a suitable cost function at each node. The optimization at each non-leaf node does not, in general, reflect the node's eventual role as a discriminant function, partitioning the input space in order to minimize the distortion incurred at the descendent leaves. In the splitting algorithm [1], for example, the test vectors at a non-leaf node are chosen to minimize the distortion incurred *at the given node*, instead of minimizing the distortion at the leaves. The resulting tree may be improved simply by modifying individual nodes of the tree in order to achieve better agreement with nearest neighbor classification at the leaves.

As an example, consider the TSVQ solutions of Figure 1. The sub-optimal solution of Figure 1a was achieved

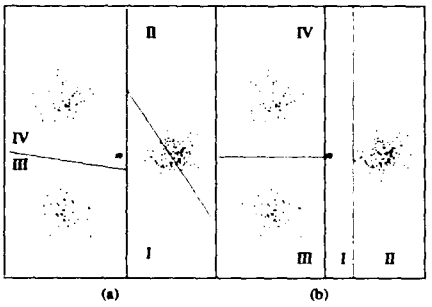


Figure 1. Binary Tree-structured VQ.

by the splitting algorithm, despite a split at the highest level which minimized the node's distortion. To improve the splitting solution, given fixed leaves, the highest level boundary should be chosen to minimize the distortion at the leaves. Finding the optimal boundary is equivalent to obtaining a minimum risk linear discriminant. Several methods from the pattern recognition field are applicable [2]. For the example of Figure 1, it suffices to choose the boundary in order to improve the agreement with nearest neighbor ownership at the leaves. This can be accomplished by choosing the highest level representatives to be the centroids of data "owned" in a nearest neighbor sense by the representatives's descendants at the leaves. Then, recalculating the leaf centroids yields the global minimum solution of Figure 1b.

*Supported by the Engineering Fondation with the cooperation of IEEE, grant RI-A-92-12

To generalize the method to trees of any depth, at each step, we fix all nodes of the tree but one and seek to minimize the risk associated with this node. Given the updated node, we can then optimize its descendant leaf codevectors as the centroids of their respective partitions. This process is repeated over all nodes of the tree. In principle, if the risk is minimized at each node, the node updates are non-increasing in the distortion of the tree and the method can be iterated until convergence. In our simulations, we applied the crude approximation of updating node representatives as centroids of the data "owned" in a nearest neighbor sense by their descendants at the leaves. This low-complexity version of our method does not guarantee a descent, though in our simulations it does improve upon the splitting solution. We tested Gaussian and Gauss-Markov sources with vector dimensions from four to eight and tree depths from five to ten. Typically, our method gained 20-30 percent of the performance gap between TSVQ via splitting and full-search VQ via the LBG method. We expect that more improvement should be possible with the use of a better linear discriminant.

A shortcoming of the approach described above is the dependence on the tree initialization. For complex data distributions, the problem of local minimum traps can become severe, and prompts us to seek a method which is insensitive to initialization, and which can avoid some local minima. Motivated by the deterministic annealing method for unstructured vector quantization and clustering [4], we have derived a related approach for the hierarchically structured clustering problem. For the structured problem, we view the hierarchical partitioning requirement as prior knowledge, and, accordingly, invoke the principal of minimum cross entropy. The resulting method has been tested on challenging problems involving normal mixtures, and has been found to obtain significant improvement over both the splitting algorithm, and, for some examples, the K-means clustering algorithm.

REFERENCES

[1] A. Buzo, A. G. Jr., R. Gray, and J. Markel, "Speech coding based on vector quantization," *IEEE Transactions on Acoustics, Speech, andSignal Processing*, vol. 28, pp. 562-574, 1980.

[2] R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*. New York, NY: Wiley-Interscience, 1974.

[3] K. Rose, E. Gurewitz, and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Transactions on Information Theory*, vol. 38, pp. 1249-1258, 1992.

Index Assignment for Progressive Transmission of Full Search Vector Quantization

Eve A. Riskin, Les E. Atlas, Ren-Yuh Wang
Dept. of Electrical Engineering, FT-10
University of Washington
Seattle, WA 98195
email: riskin@ee.washington.edu

Richard Ladner
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195

Abstract

We address progressive transmission of full search image vector quantization. We build a progressive transmission tree to define binary mergings of codewords for successively smaller sized codebooks. The tree design methods we apply are the generalized Lloyd algorithm splitting algorithm, minimum cost perfect matching, and a method of principal eigenvectors.

Vector quantization (VQ) [1] is a lossy compression technique that has been used extensively for image compression. Progressive image transmission allows an image being transmitted to be recognized early at the receiver; this saves bandwidth if the wrong image is being sent. We present three new methods for the selection of codeword indexes which allow for direct progressive transmission of images compressed with full search VQ. In all cases, we fit a tree of intermediate codewords to a full search VQ codebook and use the tree indexes as the codeword indexes.

A *full search progressive transmission tree* allows full search VQ to be sent progressively. It is a balanced tree whose terminal nodes or leaves are labeled by codewords generated by a codebook design technique and whose internal nodes are labeled by intermediate codewords derived from the leaf codewords. The tree is used to reassign the original indexes of the leaf codewords to new indexes that are compatible with progressive transmission. With each bit, the receiver displays the intermediate codeword located at the internal node being visited in the tree.

We use *region-merging* to build the progressive transmission tree and determine the intermediate codewords. A region-merging tree is formed by merging Voronoi regions of the original codebook in pairs to form larger encoding regions.

Ordered VQ codebooks provide a simple method to building the region-merging tree. In an ordered VQ codebook, the codewords with neighboring indexes are also neighbors in the input space. The region-merging tree is built by simply merging together regions with neighboring codeword indexes. We found that the generalized Lloyd splitting algorithm (GLA) gives codebooks that are reasonably well ordered.

Another method of forming a region-merging tree is minimum cost perfect matching (MCPM) from optimization theory [2]. In MCPM we have a complete graph of nodes and a cost associated with matching each different pair of nodes. The cost of the overall matching is the total sum of the costs of matching each graph node pair. To construct the region-merging tree, we choose the graph nodes to be the Voronoi regions defined by the original codebook and the cost to be the increase in distortion due to merging two Voronoi regions together. The tree is built from the bottom up by repeatedly solving the MCPM problem. Running MCPM to find the matching with minimum cost assures that the increase in distortion at the next level of the tree is minimized.

For progressive transmission, the image quality at lower rates is more important than at high rates. Unlike MCPM which is a bottom-up method, our third approach is a top-down method which seeks to minimize the distortion at lower bit rates. In this case, we start with the centroid of the codebook and successively divide the codewords in half. The problem reduces to find an optimal partition to separate the size N codebook into two size $\frac{N}{2}$ subcodebooks to maximize the decreased distortion between the centroid of the size N codebook and the two centroids of the size $\frac{N}{2}$ codebooks. Unfortunately, there are $C(N, \frac{N}{2})$ possible ways to choose the partition and solving this problem becomes impractical for moderate N. Our heuristic places a hyperplane perpendicular to the principal eigenvector of the training set to divide the codewords in half. Some codewords near the hyperplane are exchanged to maximize the decreased distortion. The tree is built from the top down by repeatedly solving this optimization problem.

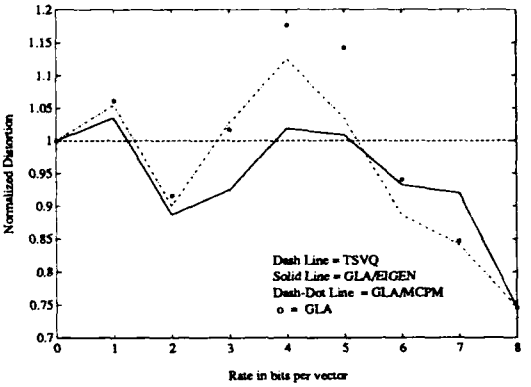


Figure 1: The normalized MSE distortion at each bit rate for intermediate codebooks

Figure 1 is the normalized MSE at each bit rate for intermediate codebooks for the test medical images. The GLA followed by one-step optimal MCPM (GLA/MCPM) slightly outperformed the GLA at most bit rates with a maximum improvement of 0.44 dB. We feel that the simplicity of the GLA and the only slight difference in performance makes the GLA more attractive than GLA/MCPM. In the same Figure, GLA/EIGEN which represents the principal eigenvector method, outperforms even TSVQ at most bit rates and of course gives lower distortion at the final bit rate.

References

[1] R. M. Gray. Vector quantization. *IEEE ASSP Magazine* 1:4-29, April 1984.

[2] E. L. Lawler. *Combinatorial Optimization: Networks and Matroids.* Holt, Rinehart and Winston, New York, 1976.

Generalised Theta Functions, for Lattice Vector Quantization

Patrick Solé,
CNRS, I3S,
250, rue A. Einstein,
06 560 Valbonne, France

Abstract: Generalized theta functions of lattices for metrics of L^p type are introduced and computed when a coding-theoretic construction of type A,B, or C of the lattice used for VQ exists. Upper bounds of the saddle point type and geometric lower bounds on certain sums of their coefficients are derived and applied to the estimation of the size of the codebook consisting of all points of the lattice within a sphere of given radius for the concerned metric.

Key words:Lattices, Vector Quantization, Theta Function, Binary Codes, Saddle Point Approximation, Voronoi Diagram

1 Problem and Motivation

When quantifying a multidimensional source with given pdf, the asymptotic equipartition principle of information theory tells us that the samples of the source will concentrate on or about the equiprobable surface. In the case of a Gaussian law, these surfaces are usual euclidean spheres. If a lattice Λ is used for quantizing the source with a sphere-shaped codebook then estimating the number of points inside the sphere is essential for determining the transmission rate.
We consider the case of equiprobable surfaces that are spheres for the L^p metric of the type

$$S_p(n,m) = \{x \in \mathbf{R}^n | ||x||_p \leq m\},$$

where $||x||_p^p = \sum_{i=1}^n |x_i|^p$. When the source is Laplacian $p = 1$, and $S_1(n,m)$ is a so-called pyramid (or hyperoctahedra). This case has practical applications in image processing [4].

2 High Rate Approximation

Gauss' counting principle says that the number of points with integer coordinates in a convex body is well approximated by its volume. This was proved to fail for $S_2(n,\sqrt{n})$ and large n in [2]. If, however, n, p are fixed and m is large this is a mere application of Riemann sums. If a lattice Λ is used with fundamental volume(=volume of its Voronoi cell) $vol(\Lambda)$ this says that

$$|S_p(n,m) \bigcap \Lambda| \sim \frac{vol(S_p(n,m))}{vol(\Lambda)}. \qquad (1)$$

Explicit formulas for $vol(S_p(n,m))$ can be found in [3].

3 Generalized Theta Function

Let

$$\theta_{p,\Lambda}(q) = \sum_{x \in \Lambda} q^{||x||_p^p}. \qquad (2)$$

When $p = 2$ (2) is just the classical theta function [1] and when $p = 1$ the Nu function of [4, 5]. When Λ is obtained from a binary code C by construction A, we have

$$\theta_{p,\Lambda}(q) = W_C(\theta_{p,2Z}(q), \theta_{p,2Z+1}(q)), \qquad (3)$$

where W_C is the weight enumerator of C.

4 Saddle Point Approximation

Let $g(s) = \theta_{p,\Lambda}(e^{-s})$, for $s > 0$.

Theorem 1

$$|\Lambda \bigcap S_p(n,m)| \leq g(s_0)e^{s_0 m^p}$$

where s_0 is the unique nonegative real solution of

$$m^p g(s) = e^{-s} g'(s).$$

If the saddle point approximation applies then this bound is Θ(RHS of (1)).

5 Voronoi Covering Bound

Partitioning the sphere $S_p(n,m)$ into Voronoi domains of the Lattice points it contains we see that RHS of (1) is always a lower bound for every m.

6 Acknowledgement

We thank M. Barlaud, D. Gardy, A.M. Odlyzko, N.J.A. Sloane for helpful discussions.

References

[1] J.H. Conway, N.J.A. Sloane "Sphere Packings, Lattices and Groups" Springer Verlag (1990).

[2] J. E. Mazo and A. M. Odlyzko, "Lattice Point in High Dimensional Spheres", Mh. Math. 110, 47-61 (1990).

[3] N.D. Elkies, A.M. Odlyzko, J.A. Rush," On the packing densities of superballs and other bodies." Inventiones Math. 105 (1991) 613-640.

[4] M. Barlaud, P. Solé, M. Antonini, P. Mathieu, T. Gaidon, "Pyramidal Lattice Vector Quantization for multiscale image coding", submitted to IEEE trans. on Image Processing.

[5] P. Solé, "Counting Lattice Points in Pyramids", submitted to the proceedings of Formal Power Series and Algebraic Combinatorics, Montréal, Canada, June 1992.

Vector Quantization Codebooks from the Nordstrom-Robinson Code and Berlekamp's Negacyclic Codes

Peter F. Swaszek
Department of Electrical Engineering
University of Rhode Island
Kingston, RI 02881

Introduction

Assume a dimension n random vector source $\mathbf{X}$. A rate r vector quantizer is a mapping of $\mathbf{X}$ onto one of $2^{nr} = N$ representation vectors. The VQ is defined by these reconstruction codevectors, $\hat{\mathbf{x}}_j$, and their associated quantization regions, $\mathcal{Q}_j$. Consider the performance measure of mean squared error per dimension. For minimum MSE each codevector should be the centroid of its corresponding region. Further, if the codevectors are the centroids, then the MSE expression reduces to the difference between the input and output variances. For the discussions below assume that $\mathbf{X}$ has independent elements whose marginal density functions are unit variance, symmetric, and have equal first absolute moment $E\{|x_i|\} = \gamma$.

Some recent work [4, 5] constructed VQ codebooks from the codewords of binary linear block codes. Rather than a full-search implementation we considered a syndrome-based mapping. Specifically, taking the dual of antipodal modulation and syndrome decoding of an (n, k) code, we considered the following rate k/n VQ implementation: hard quantize each element of $\mathbf{x}$ to one bit (with threshold zero); compute the syndrome of the resulting binary sequence and "correct" the error; use the k information bits of this codeword as the VQ index; reconstruct $\hat{\mathbf{x}}$ based upon the k information bits. With this format each quantization region was the union of 2^{n-k} orthants; the VQ reconstruction vector was the centroid of its quantization region. For the assumed source the complete symmetry of the problem resulted in all regions being identical; hence, solving for $\mathcal{Q}_1$ and $\hat{\mathbf{x}}_1$ (corresponding to the all zero codeword) was sufficient to describe performance. Letting $\mathbf{e}_l = [e_{l,1}, \ldots, e_{l,n}]$, $l = 1, 2, \ldots 2^{n-k} = M$, be the coset leaders of the code, the n elements of $\hat{\mathbf{x}}_1$ and the resulting MSE were

$$\hat{x}_{1,i} = \frac{\gamma}{M}\left(2\sum_{l=1}^{M} e_{l,i} - M\right) \quad ; \quad \text{MSE} = 1 - \frac{1}{n}\sum_{i=1}^{n}\hat{x}_{1,i}^2$$

For comparison, time-shared scalar quantization (zero or one bit quantization per element) has MSE $= 1 - r\gamma^2$.

Nonlinear Codes

An $[n, N, d]$ nonlinear code consists of N codewords of length n with minimum Hamming spacing d. Being nonlinear, the simplicity of syndrome decoding and the complete symmetry of the $\mathcal{Q}_j$ are lost. From the perspective of examining a nonlinear code as a possible VQ codebook, this forces us to describe each region $\mathcal{Q}_j$ individually, computing its probability of occurence and centroid. Let $\mathbf{c}_j$ be a typical codeword and $d_H(\cdot,\cdot)$ be Hamming distance. Implementing the VQ by hard quantizing the input and minimizing the Hamming distance to a codeword, $\min_j d_H(\mathbf{c}_j, \frac{1}{2}(1 + \text{sgn}(\mathbf{x})))$, then each $\mathcal{Q}_j$ is again the union of orthants $\mathcal{Q}_j = \bigcup_{l=1}^{M_j} \mathcal{Q}_{j,l}$ where the number of orthants, M_j for the j-th region ($\sum_{j=1}^{N} M_j = 2^n$), depends upon the detail of the encoder implementation. Paralleling the analysis above for the linear code case, the i-th coordinate of the centroid of the j-th region and the overall MSE are

$$\hat{x}_{j,i} = \frac{\gamma}{M_j}\left(2\sum_{l=1}^{M_j} \mathcal{O}_{j,l,i} - M_j\right) \quad ; \quad \text{MSE} = 1 - \frac{1}{n}\sum_{j=1}^{N}\frac{M_j}{2^n}\sum_{i=1}^{n}\hat{x}_{j,i}^2$$

where $\mathcal{O}_{j,l,i}$ depends upon the orientation of the i-th coordinate of the l-th orthant of $\mathcal{Q}_j$ with respect to the i-th element of $\mathbf{c}_j$ (unity if they match, zero if they don't).

As an example, consider the Nordstrom-Robinson [15,256,5] code [2]. An algebraic decoding algorithm [3] results in all regions having the same probability of occurence (i.e. each $M_j = 128$). The resulting performance is MSE $= 1 - 0.554\gamma^2$, slightly better than scalar quantization (MSE $= 1 - 0.533\gamma^2$). Although one could consider variations of this code (extending to dimension 16 or shortening to 14, 13, and 12) the original dimension 15 version has best MSE performance.

A Negacyclic Code Example

The first step in implementing the syndrome-based VQs consists of mapping the source vector onto a discrete sequence suitable for algebraic decoding. Our approach in the binary code case was to pair each element of the source vector with a separate codeword position, thresholding the source value (at zero) to produce a binary value. This worked since Hamming distance and Euclidean distance are directly related for binary sequences. To extend to q-ary codes ($q > 2$), two mappings which come to mind are scalar quantization of each source value to q levels (as in PAM) and partitioning of the bivariate plane into q equi-angular regions (as in PSK). Although Hamming distance does not directly reflect the Euclidean relationship of points in such constellations, the Lee metric does. Since the match is better for the PSK constellation, we pursue that mapping below.

Our example assumes an iid Gaussian source and the $q = 5$ (12,8) negacyclic code from [1, p.209] ($r = (\log_2 5^8)/24 = 0.774$ bits/dim). To map the source onto the 5^{12} 5-ary sequences, we take vectors of length 24, breaking them into 12 pairs. Each pair uniquely defines a phase angle in polar coordinates; uniform quantization of the angle specifies the 5-ary symbol. For syndrome decoding there are $5^4 = 625$ cosets which include the no error pattern, the 24 distance one errors (a single ± 1), the 288 distance two errors (a single ± 2 or two ± 1s), and 313 cosets corresponding to Lee metric equal to 3. For minimum MSE encoding we choose these remaining error vectors of the form $\pm 1, \pm 1, \pm 1$. In the centroid computation, each quantization region is the union of 625 dimension 24 cones. Completing the computation (again, noting the symmetry of the problem), the centroid of $\mathcal{Q}_1$, corresponding to the all-zero codeword, is $\hat{\mathbf{x}}_1 = [447a, 0, 507a, 0, 543a, 0, 536a, 0, 544a, 0, 547a, 0, 548a, 0, 545a, 0551a, 0, 551a, 0, 558a, 0, 552a, 0]$ where $a = (\sin\frac{\pi}{5})/125\sqrt{2}$. The overall MSE is 0.4935, again slightly better than scalar quantization (MSE = 0.5073).

References

[1] E. R. Berlekamp, *Algebraic Coding Theory*, New York: McGraw-Hill, 1968.

[2] A. W. Nordstrom & J. P. Robinson, "An optimum nonlinear code," *Information and Control*, vol. 11, pp.613–616, Nov.-Dec. 1967.

[3] J. P. Robinson, "Analysis of Nordstrom's optimum quadratic code," *Proc. Hawaii Int'l. Conf. System Sciences*, pp.157–160, 1968.

[4] P. F. Swaszek, "Vector quantization based on block and spherical codes," *Proc. Conf. Information Science & Systems*, pp.570–575, Mar. 1991.

[5] P. F. Swaszek, "Syndrome-based VQ codebooks," to appear in the *Proc. DIMACS/IEEE Workshop on Coding and Quantization*, Oct. 1992.

VECTOR QUANTIZERS TRAINED ON SMALL TRAINING SETS

David Cohn
Brain & Cognitive Sci.
Mass. Inst. of Technology
Cambridge, MA 02139

Eve A. Riskin
Electrical Engineering
Univ. of Washington
Seattle, WA 98195

Richard Ladner
Computer Sci. & Eng.
Univ. of Washington
Seattle, WA 98195

Abstract

We examine how the performance of a memoryless vector quantizer (VQ) changes as a function of its training set size. By relating the training distortion of such a codebook to its test (true) distortion, we demonstrate that one may obtain "good" codebooks at a fraction of the computational cost by training on a small random subset of the blocks in the target image.

Background

For a system with a fixed number of degrees of freedom, one may bound the difference between the error of that system on an arbitrary distribution (a test set) and its performance on a subset of that distribution (the training set). Roughly, with fixed confidence, this difference is bounded by

$$(test - train) \leq O(\frac{d}{m} \log \frac{1}{m}), \qquad (1)$$

where m is the size of the training set and d is the Vapnik-Chervonenkis (VC) dimension of the system, a measure of its number of degrees of freedom [1]. Empirically, it has been observed that for some learning systems, the expected value of this difference varies as $O(d/m)$. By bounding the difference between test and training errors, one can bound the difference between the *test error* and the "optimal" error — the test error if the system had been trained on an infinite amount of data.

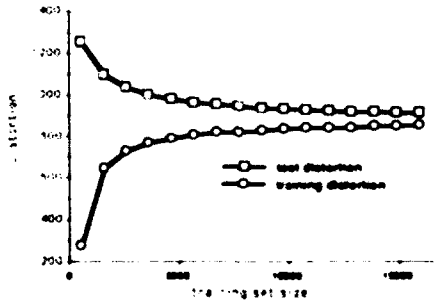


Figure 1: Test and training distortion of codebooks trained on small subsets of blocks drawn at random from a target image.

Application to Vector Quantization (VQ)

We have extended this theory to relate to the training of vector quantizer codebooks. We have also conducted empirical studies which have determined the effective VC dimension of VQ codebooks. Our results quantitatively show that a VQ codebook trained on a small random subset of vectors from a target image performs almost as well at quantizing that image as a codebook trained on the entire image, but at a fraction of the computational cost [2] (Fig. 1). Some empirical results are outlined below.

Given an image Z composed of M k-dimensional blocks, we wish to design a k-dimensional codebook with N codewords. We extract m k-dimensional blocks at random from Z (with or without replacement), and use this training set Z^m as input to the GLA codebook design algorithm [3]. The distortion that the resulting codebook imposes on Z^m is our *training distortion*.

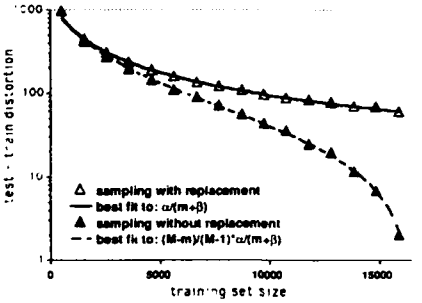


Figure 2: $(Test - train)$ distortion is described by a simple function of the training set size.

The codebook is then used to quantize image Z, and the resulting *test distortion* is measured. As predicted by theory, the test and training distortions followed a simple relationship (Fig. 2):

$$(test - train) = \frac{\alpha}{m + \beta} \text{ [replacement]}$$

$$(test - train) = \frac{M - m}{M - 1} \frac{\alpha}{m + \beta} \text{ [no replacement].}$$

Parameter α is the *learning complexity* of the image for the given codebook size, and β is an offset factor, which, for the most part, we may ignore. We found that α varied very little between natural images, but depended almost primarily on N. For normalized mean-squared distortion of 8-bit grayscale images, we found a "typical" learning complexity of

$$\alpha(N) = 0.363N^{0.79}$$

Let us say that we want our codebook's distortion to be within 3% of its asymptotic distortion. If $N = 512$ vectors, $\alpha \approx 50$. Solving $(test - train) = 0.03 \approx \alpha/m$ for m gives $m = 1672$ blocks. This indicates that if we train our 512-vector codebook on 1672 blocks drawn at random from an image (or set of images), we can expect that the codebook's performance on that entire image will be within 3% of the performance that we would get if we were to train the codebook on the entire image (or set of images), regardless of how large the image (set) is!

These results are for a particular implementation of GLA, and the exact dependence of α on N will vary with different implementations and different input domains. We have obtained slightly different results using simulated annealing to design codebooks. Still, for given training regimen, it is straightforward for users to determine the typical learning complexities of their particular problems, and then use these values to select appropriately small training set sizes.

[1] **V. Vapnik**, *Estimation of dependencies based on empirical data*, Springer-Verlag, New York, 1982.

[2] **D. A. Cohn**, "Separating formal bounds from practical performance in learning systems," Ph.D. dissertation, Univ. Washington Computer Science & Eng., 1992.

[3] **Y. Linde, A. Buzo, and R. M. Gray** "An algorithm for vector quantizer design," in *IEEE Transactions on Communications*, 28:84-95, 1980.

The Dynamics of Group Codes: Syndromes, Normal Codes, and Canonical Observers

G. David Forney, Jr.
Motorola Codex
Mansfield, MA 02048

Mitchell D. Trott
MIT
Cambridge, MA 02139

A *group code* C is a subgroup of a direct product *sequence space* G^I, where G is a group and $I \in \mathbf{Z}$ is an index set. In other words, a codeword $\mathbf{c} \in C$ is a sequence $\mathbf{c} = \{c_k \in G : k \in I\}$ of elements of G. Though G may in general be nonabelian, we refer to the group operation of G as a "sum" and its identity element as 0. The group property of C thus ensures that the componentwise sum of any two codewords is another codeword.

For the purposes of this abstract, we assume that $I = \mathbf{Z}$, so that codewords are bi-infinite sequences, and that C is time-invariant, so that shifts of codewords are codewords. These two assumptions serve primarily to simplify notation. We also require the technical condition that C be a closed set in the topology of pointwise convergence.

Let C_J denote the set of sequences of C that are zero outside the subset $J \subseteq \mathbf{Z}$. So, for example, $C_{[m,n)}$ denotes the set of sequences that are possibly nonzero only on the interval $[m, n)$. The set C_J is a normal subgroup of C. In [1] it was shown that any group code C has a well-defined state group $\Sigma_k = C/(C_{(-\infty,k)}C_{[k,\infty)})$ at each time $k \in \mathbf{Z}$, and that C has a canonical state/output realization whose state space at each time k is Σ_k. The canonical realization identifies each code sequence $\mathbf{c}$ with a unique state sequence $\sigma(\mathbf{c})$, so that one may refer unambiguously to the state $\sigma_k(\mathbf{c})$ of a sequence $\mathbf{c} \in C$ at each time k. The canonical realization is minimal.

It was further shown that a minimal feedforward encoder for C can be constructed in controller canonical form from elementary constituents of the state group Σ_k, called "granules." The controller canonical form describes code sequences as combinations of finite-length generator sequences.

In this paper we give a dual construction based on a *state observer* that recovers the state of a code sequence $\mathbf{c} \in C$ at time k from a coset decomposition of recent outputs $c_{k-m}, \ldots, c_{k-1}$. The state observer is used to construct a syndrome former and a minimal encoder in observer canonical form that describes code sequences in terms of "parity checks" that must be satisfied.

A *syndrome map* for a group code C is a function $f: G^{\mathbf{Z}} \to S$ whose kernel $f^{-1}(0)$ is C, where the *syndrome set* $S = f(G^{\mathbf{Z}})$ is a set that contains the element 0. If f is a syndrome map for C, then a sequence $\mathbf{c} \in G^{\mathbf{Z}}$ is a codeword of C if and only if $f(\mathbf{c}) = 0$.

A syndrome map $f: G^{\mathbf{Z}} \to G^{\mathbf{Z}}//C$ for a group code C may be constructed by setting $f(\mathbf{g}) = \mathbf{g}C$ for $\mathbf{g} \in G^{\mathbf{Z}}$, where the syndrome set S is the set $G^{\mathbf{Z}}//C$ of left cosets of C in $G^{\mathbf{Z}}$ and the "identity" coset C of $G^{\mathbf{Z}}//C$ is identified as the element 0 of S. This map is a homomorphism if C is a normal subgroup of $G^{\mathbf{Z}}$; then the syndrome set is the quotient group $G^{\mathbf{Z}}/C$. More generally, there exists a homomorphism $f: G^{\mathbf{Z}} \to S$ with kernel C, *i.e.*, a homomorphic syndrome map, if and only if C is a normal subgroup of $G^{\mathbf{Z}}$.

We are thus motivated to investigate group codes that are normal subgroups of their parent sequence space, which we call "normal codes." All abelian codes are normal. More generally, we show that C is normal if and only if C has abelian dynamics, *i.e.*, if and only if its state group $\Sigma_k = C/(C_{(-\infty,k)}C_{[k,\infty)})$ is abelian at each time $k \in \mathbf{Z}$.

The parallel transition subgroup $\prod_{k\in\mathbf{Z}} C_{[k,k]}$ of a normal code must include the commutator subgroup of C, so $C_{[k,k]}$ must include the commutator subgroup of $A_k = \{c_k : \mathbf{c} \in C\}$. When the output sequence space is nonabelian, therefore, there exist nontrivial parallel transition subgroups; these impose upper bounds on the minimum distance of C.

A code is *μ-observable* if, given any two code sequences $\mathbf{c}, \mathbf{c}' \in C$ that agree on a length-μ interval $[k, k+\mu)$, the sequence $\mathbf{c}''$ defined by

$$c_i'' = \begin{cases} c_i & \text{if } i < k, \\ c_i' & \text{if } i \geq k \end{cases}$$

is a code sequence in C. In other words, if two code sequences agree on an interval of width at least μ, then the past of one can be concatenated with the future of the other. The least such μ is the *observability index* of C. A 0-observable system is *memoryless*.

A *syndrome former* for a code $C \subseteq G^{\mathbf{Z}}$ is an input/state/output dynamical system with input space $G^{\mathbf{Z}}$ and output space $S^{\mathbf{Z}}$ whose output sequence is 0 if and only if the input sequence is in C. A syndrome former has *memory* m if the output s_k at any time $k \in \mathbf{Z}$ can be expressed as a function of the previous m and the current inputs, $c_{k-m}, \ldots, c_k$.

We show that the a syndrome former for a code with observability index μ must have memory $m \geq \mu$. We also show that a syndrome former must contain an underlying state-output realization of C; the syndrome former is minimal if and only if this underlying realization is minimal. As group codes have essentially unique minimal realizations, the state spaces of a minimal syndrome former are essentially unique and correspond to the state spaces Σ_k of C. A minimal syndrome former must therefore track the state sequence of C: syndrome formers are inherently state observers.

Given a μ-observable group code C, we show how to construct a minimal syndrome former for C with memory μ. The construction is based on a decomposition of the state group into *dual granules*

$$\Gamma_{[k,k+j)} = C_{\mathbf{Z}-[k,k+j)}/(C_{\mathbf{Z}-[k-1,k+j)}C_{\mathbf{Z}-[k,k+j)}).$$

At each time k the state group Σ_k has a coset decomposition into dual granules

$$\Sigma_k \leftrightarrow \bigotimes_{j=0}^{\mu-1} \prod_{i\in[k-j,k]} \Gamma_{[i,i+j)}$$

such that the value of a granule $\Gamma_{[i,i+j)}$ depends only on outputs $c_{i-1}, \ldots, c_{k-1}$. An observer constructed from this decomposition is necessarily feedforward with memory μ.

From this system one may construct a minimal syndrome former, a minimal feedforward inverter, and a minimal encoder in observer canonical form.

References

[1] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams and canonical encoders." Submitted to *IEEE Transactions on Information Theory*, February 1992.

Realizing Trellis Codes as Isometry Codes

Mitchell D. Trott
MIT
Cambridge, MA 02139

An *isometry* ϕ of n-dimensional Euclidean space R^n is a bijection $\phi: R^n \to R^n$ that preserves Euclidean distance. An *isometry code* C is a subgroup of a direct product $G^{\mathbf{Z}}$, where G is a group of isometries of R^n. In other words, a codeword $c \in C$ is a sequence $c = \{c_k \in G : k \in I\}$ of isometries of R^n. A codeword is therefore an isometry of infinite-dimensional Euclidean space $(R^n)^{\mathbf{Z}}$. Isometry codes are a type of group system, and can be analyzed using the theory developed in [2].

Many useful trellis codes may be described as the orbit $C\mathbf{x}$ of a sequence $\mathbf{x} \in (R^n)^{\mathbf{Z}}$ under an isometry code C. Certain aspects of isometry codes are studied, under different terminology, by Forney [1] and Loeliger [3]. Trellis codes generated by isometry codes are geometrically uniform; thus, when used for data transmission over an additive white Gaussian noise channel with maximum likelihood decoding, the probability of error is independent of the transmitted codeword. This property greatly simplifies performance analysis.

We develop a method for realizing trellis codes as isometry codes. First, the states of a minimal trellis for the code are assigned a group structure that is "consistent" with the trellis branches and labels. The Euclidean subset label on each trellis branch is then replaced with a coset of isometries that generates the Euclidean subset from a distinguished initial point $x \in R^n$. The resulting trellis defines an isometry code that, when applied to the initial sequence $\mathbf{x} = \{x_k = x : k \in \mathbf{Z}\}$, yields the original trellis code.

In more detail, let C be an isometry code, and let $C\mathbf{x}$ be a corresponding geometrically uniform trellis code. We assume that the signal set $S \subset R^n$ of the trellis code is partitioned into n cells (subsets) $S_0, \ldots, S_{n-1}$, and that C and $C\mathbf{x}$ have the same minimal trellis diagram. Let Σ be the state group of C, and let the branch group B be the set of all state pairs $(\sigma_1, \sigma_2) \in \Sigma \times \Sigma$ that are connected by a trellis branch. Assume without loss of generality that S_0 is the cell assigned to the branch from the identity state to the identity state, and define B_i to be the set of branches labeled with the partition cell S_i.

We show that the Euclidean and isometry labelings of the minimal trellis are related as follows: first, B_0 is a subgroup of the branch group B, and the set B_i of branches labeled with partition cell S_i is a left coset of B_0 in B. Second, for any branch $b \in B$, if $bB_i = B_j$ then the coset of isometries assigned to branch b sends cell S_i to S_j. These two conditions completely characterize the possible state groups and isometry labelings consistent with a particular Euclidean trellis.

The first of these conditions is satisfied by any trellis code described as the combination of a binary linear convolutional code and a mapping from coded bits to partition cells. The second condition, however, holds only if the mapping from bits to cells respects the symmetries of the partition.

For example, we show that a particular sixteen-state Wei code, defined over an 8-way partition of the integer lattice $\mathbf{Z}^4$, has a representation as an isometry code. The code is defined by a rate-2/3 binary linear convolutional code followed by an unusual mapping from coded bits to partition cells. We show that this mapping satisfies the conditions presented above by finding a group of 8 cosets of isometries of R^4 that generates the partition and is isomorphic to $(\mathbf{Z}_2)^3$.

As a second example, we show that Wei's nonlinear eight-state trellis code, specified in the CCITT V.32 standard, also has a representation as an isometry code. Ignoring the edge effects of the finite signal set, the V.32 code is therefore geometrically uniform. The signal set for the V.32 trellis code is a translate of the 8-way lattice partition $\mathbf{RZ}^2/4\mathbf{Z}^2$. Isometries of R^2 are denoted as follows: $t_{(a,b)}$ is translation by (a, b), r_θ is rotation by θ degrees clockwise about the origin, v_1 is reflection across the line $\{(0, y) : y \in R\}$, v_2 is reflection across the line $\{(-y, y) : y \in R\}$, and 1 is the identity isometry. The V.32 trellis code is then described as the orbit of the sequence $\mathbf{x} = \{\ldots, (0,1), (0,1), \ldots\}$ under the isometry code C generated by the time shifts of the sequences

$$\begin{aligned} g_1 &= (\ldots, 1, 1, t_{(2,2)}, t_{(2,2)}r_{180}, 1, 1, \ldots), \\ g_2 &= (\ldots, 1, 1, t_{(2,0)}, v_2, t_{(0,2)}v_1, 1, 1, \ldots). \end{aligned}$$

This description may be converted into a trellis diagram by replacing isometries by their permutation actions on the cells of the partition $\mathbf{RZ}^2/4\mathbf{Z}^2$. This converts the isometry code C, defined over an infinite isometry alphabet, into a code C' defined over a finite alphabet of permutations. The methods developed in [2] may then be applied.

The state group of the V.32 isometry code is nonabelian, and is isomorphic to the dihedral group D_4. Each of the 32 trellis branches is assigned a distinct coset of isometries, yet there are only 8 partition cells. The map from isometries to partition cells is therefore many-to-one—a property peculiar to isometry codes with nonabelian state groups. The isometry code contains the constant rotation sequences $(\ldots, r_{90}, r_{90}, r_{90}, \ldots)$, which is a sufficient condition for 90-degree rotational invariance.

As a final example, we show that the $(16, 8, 6)$ binary nonlinear Nordstrom-Robinson code may be represented as a block isometry code over a group of rotations of R^2, or equivalently as a ring code over $(\mathbf{Z}_4)^8$. The binary code is embedded in R^{16} by interpreting codewords as vertices of a 16-cube.

References

[1] G. D. Forney, Jr., "Geometrically uniform codes," *IEEE Transactions on Information Theory*, vol. IT-37, no. 5, pp. 1241–1260, 1991.

[2] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams and canonical encoders." Submitted to *IEEE Transactions on Information Theory*, February 1992.

[3] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Transactions on Information Theory*, vol. IT-37, no. 6, pp. 1675–1682, 1991.

ON GEOMETRICALLY UNIFORM SIGNAL SETS AND SIGNAL SETS MATCHED TO GROUPS

Zhe-xian Wan

Department of Information Theory
University of Lund
Box 118
S- 221 00 Lund
Sweden

Summary — Any discrete subset of an R^N,where N is any positive integer, is called a signal set. A signal set may be finite or infinite. A bijective map from R^N to itself, which preserves Euclidean distance, is called an isometry of R^N. The set of all isometries of R^N which leaves a signal set $S \subseteq R^N$ invariant forms a group with respect to the composition, called the symmetry group of S and denoted by $\Gamma(S)$. In 1991 G.D.Forney [1] introduced geometrically uniform signal sets.

Definition 1: A signal set S is said to be geometrically uniform if $\Gamma(S)$ acts transitively on S. In the same year H.-A. Loeliger [3] introduced signal sets matched to groups.

Definition 2: A signal set S is said to be matched to a group G if there is a surjective map μ from G to S such that, for all g and g' in G

$$d(\mu(g), \mu(g') = d(\mu(g^{-1}g'), \mu(e)), \qquad (1)$$

where d denotes the Euclidean distance and e denotes the identity element of G.

The purpose of this note is to show that these two concepts coincide. The case when the signal set is finite was proved by Loeliger [3].

A bijective map f from a signal set S to itself is called an isometry of S, if for all s and s' in S,

$$d(f(s), f(s')) = d(s, s'). \qquad (2)$$

Lemma 1: Any symmetry of a signal set $S \subseteq R^N$ can be extended to an isometry of R^N.

Corollary 2: Let S be a signal set in R^N and assume that Span$S = R^N$. Then any symmetry of S can be extended uniquely to an isometry of R^N.

Theorem 3: Let S be a signal set in R^N. Then S is a geometrically uniform signal set if and only if S can be matched to a group.

Corollary 4: Let S be a signal set matched to a group G. Assume that S span R^N. Then G is homomorphic to a transitive subgroup of $\Gamma(S)$.

References

[1] G.D.Forney, Geometrically uniform codes, IEEE Transactions on Information Theory, IT-37(1991), 1241-1260.

[2] W.Ledermann and S. Vajda (ed.), Handbook of Applicable Mathematics, Vol. V; Combinatorics and Geometry, Wiley, 1985.

[3] H.-A. Loeliger, Signal sets matched to groups, IEEE Transactions on Information Theory, IT-37(1991), 1675-1682.

Euclidean-Space Coding Theorems for Linear Codes and Mod-p Lattices

Hans-Andrea Loeliger
ISY / Information Theory
Linköping University
S-58183 Linköping, Sweden

All known existence proofs for capacity-achieving codes whose algebraic structure is at least a group rely on averaging arguments for linear codes over finite fields — except, seemingly, de Buda's proof [1] for lattice codes. De Buda's starting point is, instead, the Minkowski-Hlawka theorem from geometric number theory. We remove this anomaly by showing that the standard proof of that theorem has a natural interpretation as an averaging argument for linar codes in the following setup.

We consider the discrete-time Gaussian channel with p-level amplitude modulation, p prime, and hard-decision 'mod-p demodulation', i.e., the received signal is reduced mod-p and quantized to the nearest integer. (The approach can be extended to soft-decision, however.) We thus have created a channel with mod-p additive noise, for which linear codes over GF(p) are the natural choice. Moreover, existence results for linear codes imply corresponding results for the associated mod-p lattices, as is demonstrated by some examples. In particular, the Minkowski-Hlawka theorem is shown to follow from a Gilbert-Varshamov-type argument for linear codes.

Having thus seen that the first step of de Buda's proof can be derived with averaging arguments for linear codes, it is natural to ask whether an alternative, more direct, existence proof for good lattice codes can be based on such arguments, which we show is indeed possible.

It is shown that spherically shaped cosets of linear codes over GF(p) achieve the capacity of almost every channel with p inputs, each associated with a certain cost, under a constraint on the total cost of each codeword. Application of this result to the Gaussian channel with p-level amplitude modulation implies, in the limit for $p \to \infty$, the coding theorem for the corresponding mod-p lattice codes. (Such a limit $p \to \infty$ is also part of the standard proof of the Minkowski-Hlawka theorem and thus implicitly contained in de Buda's proof.)

An unsatisfactory point of our proof — as well as of de Buda's, even in its corrected version [2] — is that the upper bound on error probability holds only for the average over all codewords of the code; if the code is 'cleaned' be deleting weak codewords, then its algebraic structure is destroyed. We thus conclude by emphasizing that there is no proof known that reasonably shaped lattice codes *without weak codewords* can achieve the capacity of the Gaussian channel at any finite SNR.

References

[1] R. de Buda, 'Some optimal codes have structure', *IEEE J. Select. Areas Comm.*, vol. 7, pp. 893–899, Aug. 1989.

[2] T. Linder, Ch. Schlegel, and K. Zeger, 'Comments on "Some Optimal Codes have Structure"', submitted to *IEEE Trans. Inform. Theory*.

Analysis of Block Codes Designed over the Real-Field

PETER MASSEY and PETER MATHYS
Dept. of Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309

Abstract

Error-correcting channel codes designed over the real-field are advantageous over finite-field codes in certain cases. The Real-field codes are sensitive to small deviations from the ideal continuous values as well as to large errors which are desired to be removed. The former caused by things like finite precision, receiver noise, or channel noise. A special channel model is used which includes both background and impulsive noises. Real-field codes can "correct" up to $(N-K-1)$ impulsive errors per word if the ratio of noise powers (σ_m^2/σ_n^2) is large enough.

Summary

The Background and Impulsive Noise (BIN) channel has additive white Gaussian background noise which affects every vector component while another independent Gaussian noise is switched-on with a probabilistic switch. The BIN channel has a larger capacity for a continuous Gaussian input than for a finite-alphabet input, thus motivating the use of Real-codes on this channel. Large alphabet inputs approach the capacity of the continuous input channel. Block channel coding is accomplished by a linear transformation on the information symbols, $\underline{c} = G\underline{u}$, where Euclidian distance is preserved by requiring $G^TG = I$.

The optimal minimum mean-squared-error estimator (MMSEE) is the best decoding algorithm when the decoded MSE is to be minimized. It has the form:

$$\hat{\underline{u}} = E[\underline{u}|\underline{y} = \underline{\beta}] = \frac{\sum_{\underline{s}_j} \hat{\underline{u}}^{(j)}\, p_{\underline{y}|\underline{s}=\underline{s}_j}(\underline{\beta})\, Pr[\underline{s} = \underline{s}_j]}{\sum_{\underline{s}_j} p_{\underline{y}|\underline{s}=\underline{s}_j}(\underline{\beta})\, Pr[\underline{s} = \underline{s}_j]} \tag{1}$$

$$\hat{\underline{u}}^{(j)} = E[\underline{u}|\underline{y} = \underline{\beta}, \underline{s} = \underline{s}_j] \tag{2}$$

$\underline{u}$ is the input vector; $\underline{y}$ is the output vector;
$\underline{s}$ is the impulsive error location pattern (ie. $\underline{s}$= (0,1,0,1,0))

Unfortunately, it requires an exhaustive search over all 2^N possible impulsive-error location patterns, $\underline{s}_j$. A lower bound shows that the best decoding MSE per component cannot be much smaller that the white Gaussian background noise variance. The ratio (σ_m^2/σ_n^2) of impulsive-error variance to background noise variance is critical to the performance of the decoding. Larger ratios of (σ_m^2/σ_n^2) permit a decoding MSE which is closer to the background noise level.

Using a parity-check matrix, H^T, which is orthogonal to the generator matrix G, eliminates the need for knowledge of the source, but at the expense of increased MSE. The optimal $\text{MMSEE}_{\underline{z}}$ for the syndrome subspace $\underline{z}$, is difficult to derive analytically. The estimator is developed by using a form similar to the MMSEE, but which uses an indirect estimate of the source word.

$$\hat{\hat{\underline{u}}} = \frac{\sum_{\underline{s}_j} \hat{\underline{u}}^{(j)}\, p_{\underline{z}|\underline{s}=\underline{s}_j}(\underline{\zeta})\, Pr[\underline{s} = \underline{s}_j]}{\sum_{\underline{s}_j} p_{\underline{z}|\underline{s}=\underline{s}_j}(\underline{\zeta})\, Pr[\underline{s} = \underline{s}_j]} \tag{3}$$

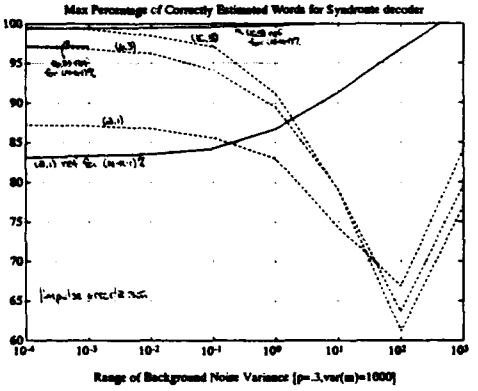
Again, a lower bound can be derived which shows that no decoding which uses this syndrome method can have a MSE less than the background noise variance. It also predicts the MSE for each possible error-pattern. The MSE increases with the number of estimated errors. The MSE performance of generator-parity check matrix pairs can be quantified. This decoding algorithm also requires an exhaustive search of the 2^N possible impulsive-error location patterns.

Some applications require that the important performance criterion be the probability that impulsive errors be correctly detected. We define an impulsive error as any channel noise with amplitude greater than some multiple of the background noise standard deviation, $\theta\sigma_n$. The optimal estimators for the error-location patterns $\underline{s}_j$ are the MAP decision rules in the codeword space and the syndrome subspace which are given by:

$$max_{\underline{s}_j}[p_{\underline{s}|\underline{y}=\underline{\beta}}(\underline{s}_j)] = max_{\underline{s}_j}[p_{\underline{y}|\underline{s}=\underline{s}_j}(\underline{\beta})Pr[\underline{s} = \underline{s}_j]] \tag{4}$$

$$max_{\underline{s}_j}[p_{\underline{s}|\underline{z}=\underline{\zeta}}(\underline{s}_j)] = max_{\underline{s}_j}[p_{\underline{z}|\underline{s}=\underline{s}_j}(\underline{\zeta})Pr[\underline{s} = \underline{s}_j]] \tag{5}$$

Several interesting trends can be observed in the performance of the real-codes when constant energy codes are compared at different rates. Larger ratios (σ_m^2/σ_n^2) of impulse-variance to background-variance improves the error location estimation. The syndrome $\text{MAP}_{\underline{z}}$ estimator can "correct" up to $(N-K-1)$ errors in a word if the ratio (σ_m^2/σ_n^2) is large enough. Short codelengths can "correct" some of the $(N-K)$ patterns. The idea of "correct" means to exactly estimate impulsive errors when no background noise is present, or with a high probability of being within some $\theta\sigma_n$ of the impulsive amplitudes when background noise is present. The optimal MAP estimator can "correct" more than the $\text{MAP}_{\underline{z}}$, up to N impulsive errors in a word, but this amount decreases as the input power increases until an infinite input power equates it to the performance of the $\text{MAP}_{\underline{z}}$ estimator. It should be noted that long codelengths can make the percentage of words with $(N-K-1)$ or fewer impulsive errors as close to 100% as desired.



In many applications only very small decoding errors are considered correct, while any larger errors are all considered just as bad. Instead of using the MMSEE decoders, an alternative is to use the Largest Density (LD or $\text{LD}_{\underline{z}}$) decoder which first selects the error-location pattern $\underline{s}_j$ which corresponds to the largest conditional a posteriori density (MAP or $\text{MAP}_{\underline{z}}$). Then it uses the corresponding conditional estimate $\hat{\underline{u}}^{(j)}$ or $\hat{\hat{\underline{u}}}^{(j)}$ to estimate the source word. The LD decoders will have a higher MSE, but they will have a higher percentage of decoded words with very small errors. If the error-location pattern estimate is correct, then the error is essentially the background noise. If the pattern estimate is wrong, then the error is very large. Thus the decoding error can be made to be near the background variance for words with $(N-K-1)$ or fewer impulsive errors if the ratio (σ_m^2/σ_n^2) is large enough.

On Minimality Conditions for Linear Systems and Convolutional Codes

Hans-Andrea Loeliger
ISY / Information Theory
Linköping University
S-58183 Linköping, Sweden

Thomas Mittelholzer
Signal and Inform. Proc. Lab.
ETH-Zentrum
CH-8092 Zürich, Switzerland

The recent generalizations of convolutional codes to rings [1], [2] and groups [3], [4] have redirected some attention to the old topic of characterizing minimal and catastrophic encoders and how these notions are related to the minimality concept for linear systems. We address these questions in a universal-algebra framework that treats codes over groups, rings, and fields in a unified way.

Any standard description of a convolutional encoder leads to a state-transition diagram (or 'transition graph' [4]) whose branches B form a subspace of $S \times Y \times S$, where S is the state space and Y is the code symbol (encoder output) by which the branch is labeled. (For convolutional codes over groups, the branches form a sub*group* of the direct product $S \times Y \times S$, where S and Y are groups.)

A branch $(s, 0, s') \in B$ is called *left-neutral (right-neutral)* if $s = 0$ ($s' = 0$); it is *two-sided neutral* if it is part of a zero loop. It will be assumed that the state space (state group) S satisfies the so called descending-chain condition, which is a generalization of finite dimensionality to modules and groups. (This condition is always satisfied for finite S).

Theorem: Each of the following conditions is equivalent to the minimality of the transition graph:

1. No state other than the zero state is the ending or starting state of a semi-infinite path all of whose labels are zero.
2. The set of left-neutral branches, the set of right-neutral branches, and the set of two-sided neutral branches all consist only of the zero-branch.
3. There exists a nonnegative integer L such that any length-L path segment is uniquely determined by its label sequence.

Most other published minimality conditions for convolutional encoders can easily be derived from these conditions.

If the branches are labeled with input-output pairs rather than with output symbols only, then the above conditions characterize the minimality with respect to the transfer function, which is the traditional viewpoint in system theory [5], and the standard minimality condition for realizations of linear transfer functions — minimal $\Leftrightarrow$ controllable and observable — is an easy consequence of the theorem.

References

[1] R. B. Filho et al., 'Systematic linear codes over a ring for encoded phase modulation', Int. Symp. on Inform. and Coding Theory, Campinas-SP-Brasil, 1987.

[2] J. L. Massey, T. Mittelholzer, 'Convolutional codes over rings', *Proc. 4th Swedish-Soviet Int. Workshop on Inform. Th.*, Gotland, Sweden, pp. 14-18, Aug. 27-Sept. 1, 1989.

[3] G. D. Forney, Jr., and M. D. Trott, 'The dynamics of linear codes over groups: state spaces, trellis diagrams and canonical encoders', submitted to *IEEE Trans. Inform. Theory*.

[4] H.-A. Loeliger and T. Mittelholzer, 'Convolutional codes over groups', submitted to *IEEE Trans. Inform. Theory*.

[5] T. Kailath, *Linear Systems*, Prentice-Hall, 1980.

MULTILEVEL CODES FOR UNEQUAL ERROR PROTECTION

A. R. Calderbank
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue, 2C-363
Murray Hill, NJ 07974

N. Seshadri
Information Principles Research Laboratory
AT&T Bell Laboratories
600 Mountain Avenue, 2C-477
Murray Hill, NJ 07974

In many speech and image coding schemes, some of the coded bits are extremely sensitive to channel errors while some others exhibit very little sensitivity. In order to make the best use of channel redundancy, unequal error protection (UEP) codes are needed. In a bandlimited environment, such coding and the modulation should be integrated. In this work, we propose two combined UEP coding and modulation schemes.

The first method multiplexes different coded signal constellations, with each coded constellation providing a different level of error protection. The novelty here is that a codeword specifies the multiplexing rule and the choice of the codeword from a fixed codebook is used to convey additional important information. The decoder determines the multiplexing rule before decoding the rest of the data.

The second method is based on partitioning a signal constellation into disjoint subsets, where the most important data sequence is encoded, using most of the available redundancy, to specify a sequence of subsets. The partitioning and code construction is done to maximize the minimum Euclidean distance between two different valid subset sequences. This leads to novel ways of partitioning the signal constellations into subsets. Finally, the less important data selects a sequence of signal points to be transmitted from the subsets. A side benefit of the proposed set partitioning procedure is a reduction in the number of nearest neighbors, sometimes even over the uncoded signal constellation.

Many of the codes we have designed provided virtually error free transmission (greater than 6 dB coding gain) for some fraction(for example, 25%) of the data while providing a coding gain of 1 to 2 dB for the remaining data with respect to uncoded transmission. The two methods can also be combined to realize new coded signal constellations for unequal error protection.

QPSK MODULATION CODES FOR UNEQUAL ERROR PROTECTION[1]

Robert H. Morelos-Zaragoza and S. Lin

Department of Electrical Engineering
University of Hawaii at Manoa
Honolulu, Hawaii 96822 USA

SUMMARY

Unequal error protection (UEP) codes [1] find applications in broadcast channels, as well as in some digital communication systems, where messages have different degrees of importance. In this paper, we propose to use binary linear UEP (LUEP) codes, in combination with a QPSK signal set and Gray mapping, to obtain new efficient block QPSK modulation codes with *unequal squared Euclidean distances*. We present several examples of QPSK block modulation codes that have the same minimum squared Euclidean distance (MSED) as the best QPSK block modulation codes of the same length and rate. In the proposed new constructions of QPSK block modulation codes, even-length binary LUEP codes are used. It is shown that good LUEP QPSK block modulation codes are obtained by combining shorter - simpler to encode and decode - binary linear codes using the well known |u|u+v|-construction or the so-called construction X. Both constructions have the advantage of yielding optimal or near optimal binary LUEP codes of short to moderate lengths, using very simple constituent codes, and may be used as component codes in the proposed constructions of QPSK modulation codes. In addition, LUEP codes lend themselves quite naturally *multi-stage decodings* [4], using the decodings of component codes. In this paper, we present a new suboptimal two-stage soft-decision decoding of binary LUEP codes and apply it to the proposed constructions of LUEP QPSK block modulation codes.

Constructions via Gray mapping

In a QPSK signal constellation with *Gray mapping* between labels and signal points, the squared Euclidean distance between signal points is *twice* the Hamming distance between their corresponding labels. We say that this QPSK signal constellation forms a *second-order Hamming space*. Our proposed new construction consists of a Gray mapping between two-bit blocks and signal points in a QPSK signal set, together with (2n,k) binary LUEP codes, with *separation vector* $\mathbf{s} = (s_1, s_2)$, to obtain (n,k) LUEP QPSK modulation codes which have *squared Euclidean separation* $\mathbf{S} = (2s_1, 2s_2)$. Some of the resulting LUEP QPSK block modulation codes have the same MSED as that of *optimal* QPSK block modulation codes of the same rate and length [2-3]. These LUEP QPSK modulation codes offer, in addition, a larger MSED between code sequences associated with most important message bits,as shown in Table 1, where * indicates LUEP QPSK modulation codes based on the |u|u+v|-construction. G_1 and G_2 in Table 1 are asymptotic coding gains corresponding the components of the squared Euclidean separation, for the most and least significant message parts, respectively. R denotes the code rate in bits per dimension. It should be noted that *all* the optimal QPSK modulation codes found by Sayegh [2-3], of lengths 5 to 10, can be obtained based on the |u|u+v|-construction and Gray mapped QPSK signal sets. All these codes are in fact LUEP QPSK modulation codes, and this appears to be the first time that this has been pointed out.

Table 1: Some LUEP QPSK block modulation codes

2n	k	k_1	k_2	s_1	s_2	R	G_1	G_2	
10	5	1	4	5	4	1/2	3.98	3.01	*
10	7	1	6	4	2	7/10	3.65	0.64	
12	6	1	5	6	4	1/2	4.77	3.01	*
12	6	2	4	5	4	1/2	3.98	3.01	
14	7	1	6	7	4	1/2	5.44	3.01	*
14	7	4	3	5	4	1/2	4.07	3.01	

Main References

[1] B. Masnick and J. Wolf, "On Linear Unequal Error Protection Codes," *IEEE Transactions on Info. Theory*, Vol. IT-13, No. 4, pp. 600-607, July 1967.

[2] S.L. Sayegh, "A Class of Optimum Block Codes in Signal Space," *IEEE Transactions on Communications*, Vol. COM-34, No. 10, pp. 1043-1045, October 1986.

[3] S.L. Sayegh, Private communication (tables of codes from reference [2]), 1992.

[4] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error Correcting Codes," *IEEE Transactions on Info. Theory*, Vol. IT-23, No. 3, pp. 371-377, May 1977.

[1] This work was supported in part by NSF under Grant NCR-88813480 and by NASA under Grant NAG 5-931.

Jump-Diffusion Processes for Unknown Model Order Estimation Problems *

Michael I. Miller
Department of Electrical Engineering
Washington University
St. Louis, MO. 63130

Yali Amit
Department of Statistics
University of Chicago
Chicago, IL. 60637

Ulf Grenander
Division of Applied Mathematics
Brown University
Providence, RI. 02912

A new class of random sampling algorithms is presented for the solution of estimation problems over hypothesis spaces $\mathcal{E}$ which are countable unions of Euclidean spaces of varying dimension: $\mathcal{E} = \cup_{k=1}^{\infty} R^{n_k}$ with model k of dimension n_k. The estimation problem is to choose parameters in $\mathcal{E}$ given some data. The existence of a distribution μ on the parameter space $\mathcal{E}$ is assumed relating the parameters to the data, with μ taken as a convex combination of μ_k-s each a distribution on subspace $\Re^{n_k}$. The Bayesian conditional mean estimates of the parameters are generated by constructing a Markov process sampling μ.

The Markov process $X(t)$ is said to satisfy *jump-diffusion* dynamics through $\mathcal{E}$ in the sense that (i) on random exponential times the process jumps from one of the countably infinite set of spaces in $\Re^{n_k}, k = 1, 2, \ldots$ to another, and (ii) between jumps it satisfies stochastic differential equations over the respective spaces. We have proven [1, 2] that as long as the diffusions have drifts which make the μ_k measures invariant within each subspace, and that the distribution μ on $\mathcal{E}$ is invariant for the jump process, then μ is the invariant measure of the process. This coupled with the assumptions that it is possible to get from one space to another with a finite number of jumps allows proof of Harris reccurence and uniqueness of the invariant measure. From this it follows that ergodic averages generated from the process converge to their expectations, and that the transition distribution of the process converges in variational norm to the invariant measure.

These results are summarized via the following two Theorems taken from [1]. We assume that each of the distributions μ_k on $\Re^{n_k}$ have densities with respect to n_k dimensional Lebesgue measure of the Gibb's form $\frac{e^{-E_k(x)}}{Z}$

Theorem 1 *Let the jump diffusion process $X(t)$ have the properties that*

(a) the diffusion $X(t)$ within any subspace $\Re^{n_k}$ satisfies the stochastic differential equation

$$dX(t) = -\frac{1}{2}\nabla E_k(X(t))dt + dW_{n_k}(t) \qquad (1)$$

with $X(t), \nabla(\cdot)$ and $W_{n_k} \in \Re^{n_k}$ the state, gradient and standard vector Brownian motion, respectively, with the gradient $\nabla(\cdot)$ satisfying Lipschitz continuity, and

(b) the jump intensities $q(x, dy), q(x)$ defined in the standard way

$$q(x, dy) = \lim_{\epsilon \to 0} \frac{\Pr\{X(t+\epsilon) \in dy | X(t) = x\} - 1_{dy}(x)}{\epsilon} \qquad (2)$$

and $q(x) = \int_{\mathcal{E}\setminus x} q(x, dy)$ both bounded continuous functions satisfying

$$q(x)\mu(dx) = \int_{\mathcal{E}} q(y, dx)\mu(dy) \qquad (3)$$

Then μ is an invariant measure of $X(t)$.

Theorem 2 *Let $X(t)$ be the Markov process satisfying Theorem 1, along with the assumption that the Euclidean spaces are connected under the jumps, i.e. $\forall k, k'$, $\exists j(k, k') < \infty$ sequence of jumps carrying the process from $\Re^{n_k}$ to $\Re^{n_{k'}}$.*

Then μ is the unique invariant measure of the jump-diffusion process $X(t)$, and the associated chain $X(i\Delta), \Delta > 0$ converges in total variation norm $\|\cdot\|$ to μ the invariant measure:

$$\text{for all } x \in \mathcal{E}, \quad \lim_{i \to \infty} \|\Pr\{X(i\Delta) \in \cdot | X(0) = x\} - \mu(\cdot)\| = 0 \qquad (4)$$

We emphasize that the aforementioned results have been generalized to unions of manifolds such as the n_k-dimensional Torus [2]. The motivation for introducing jump-diffusions arises in object recognition [1, 3, 4] in which the different continuous and discrete components of the discovery of both the shape and number of objects in a scene are accommodated. Given a fixed number of objects, call it k, the problem is to reshape via group transformations such as scale, rotation and translation- the k objects to fit the acquired data. For this the model k consisting of n_k parameters is fixed, with the hypothesis generation a continuous diffusion through scale-rotation-translation parameter space $\Re^{n_k}$ following Langevin's stochastic differential equation. The second part of the sampling process, the jump process hypothesizes new objects and removes objects, with a jump corresponding to a transition from one continuum (model-order) to another.

References

[1] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society*, in review February 1992.

[2] Y. Amit, U. Grenander, and M.I. Miller. Ergodic properties of jump-diffusion processes. *Annals of Applied Probability*, submitted December 1992.

[3] M.I. Miller, D. Maffitt, J. Shrauner, B. Roysam, and U. Grenander. Automated segmentation of biological shapes in electron microscopic autoradiography. *Proceedings of the Twenty-Fifth Annual Conference on Information Sciences and Systems*, pages 637-642, 1991.

[4] A. Srivastava, M.I. Miller, and U. Grenander. Jump-diffusion processes for object tracking and direction finding. In *Proceedings of the 29th Annual Allerton Conference on Communication, Control and Computing*, pages 563-570, Urbana, Champaign, 1991. University of Illinois.

*Supported by NSF PYIA ECE-8552518, ARO DAAL-03-86-K-0110, ARO P-29349-MA-SDI, ONR 59922

The Optimal Error Exponent for Markov Order Estimation

Lorenzo Finesso, Chuang-Chun Liu, and Prakash Narayan

1. Introduction

A wide variety of approaches [1],[3-5] have been developed over the years to estimate the order of a Markov chain. However, as stated in Merhav *et al* [3], only recently has attention been focused on estimators with optimality properties beyond consistency. Our results are in the spirit of [3] and provide, under more general conditions, a complete characterization of the error exponents and consistency properties of a class of order estimators.

Let $\{X_n, n \geq 1\}$ be a stochastic process with values in $\mathcal{X} := \{1, \ldots, r\}$ and let P be the probability measure on $\mathcal{X}^\infty$ induced by $\{X_n\}$. The measure P is Markov of *order* k iff: $P(x_n|x_1^{n-1}) = P(x_n|x_{n-k}^{n-1})$ for $n > k$, where k is the smallest constant for which the equality above holds. Let $\mathcal{P}_k$ be the set of all stationary ergodic Markov measures on $\mathcal{X}^\infty$ of order k. We observe the process $\{X_n\}$ of *unknown measure* $P \in \bigcup_{k=1}^{k_0} \mathcal{P}_k$ where k_0 is a known constant and wish to estimate its order. We focus on estimators which satisfy a generalized Neyman-Pearson criterion of optimality. Specifically, the optimal order estimator minimizes the probability of underestimation among all estimators whose probability of overestimation lies below a prespecified level. Our main result identifies the best exponent of decay of the probability of underestimation. We further construct an estimator which achieves the best exponent.

2. Preliminaries

Given a sequence x_1^n in $\mathcal{X}^n, n > k_0$, we define its k_0-th order Markov type as the empirical distribution on $\mathcal{X}^{k_0} \times \mathcal{X}$ given by $Q := \{q_{sa}; s \in \mathcal{X}^{k_0}, a \in \mathcal{X}\}$ where:

$$q_{sa} := \frac{1}{n-k_0} \sum_{t=1}^{n-k_0} 1(X_t^{t+k_0-1} = s, X_{t+k_0} = a)$$

Let $q_s := \sum_a q_{sa}$. We define the conditional entropy of Q to be:

$$H(Q) := -\sum_{s,a} q_{sa} \log \frac{q_{sa}}{q_s}$$

with the convention that $q_{sa}/q_s = 0$ if $q_s = 0$. For $P \in \bigcup_{k=1}^{k_0} \mathcal{P}_k$ we define the conditional divergence of Q and P as:

$$D(Q\|P) := \sum_{s,a} q_{sa} \log \frac{q_{sa}/q_s}{P(a|s)}$$

Note that if $P \in \mathcal{P}_k$ for some $k \leq k_0, P(a|s)$ will depend only on the latest k components of $s \in \mathcal{X}^{k_0}$.

Let $\mathcal{Q}$ be the set of all sequences in $\mathcal{X}^n$ with (common) k_0-th order Markov type Q. The following bounds have been proved in Gutman [2]:

Lemma 1

$$|\mathcal{Q}| \geq n^{-r^{k_0}} (n+1)^{-(r^{2k_0}+1)} \exp\{(n-k_0)H(Q)\}, \text{and}$$

$$|\mathcal{Q}| \leq r^{k_0} \exp\{(n-k_0)H(Q)\}.$$

Moreover, for $P \in \mathcal{P}_k, 1 \leq k \leq k_0$, the following large deviation estimates hold:

$$P(Q) \geq (n+1)^{-r^{k_0+1}} [\min_{s \in \mathcal{X}^{k_0}} P(s)] \exp\{-(n-k_0)D(Q\|P)\},$$

$$P(Q) \leq r^{k_0} \exp\{-(n-k_0)D(Q\|P)\}. \quad \square$$

3. Main Results

Theorem 1 below specifies the rates of decay to zero of the probabilities of underestimation and overestimation for the following estimator.

Given $x_1^n \in \mathcal{X}^n, n > k_0, \hat{k}_n(x_1^n) = k$ iff:

$$(i) \quad D(Q\|P') > \epsilon_n \quad \forall P' \in \mathcal{P}_\ell, \quad 1 \leq \ell \leq k-1$$

$$(ii) \quad D(Q\|P) \leq \epsilon_n \quad \text{for some } P \in \mathcal{P}_k$$

where $\epsilon_n := (r^{k_0+1} + \delta)\frac{\log n}{n-k_0}$ and $\delta > 0$ is a constant that will be specified later. If neither condition above is satisfied, set $\hat{k}_n(X_1^n) = k_0$.

Theorem 1

Fix $\delta > 0, \gamma > 0$. Fix $P \in \mathcal{P}_k$ for some $1 \leq k \leq k_0$.

$$(i) \quad P(\hat{k}_n(X_1^n) > k) \leq r^{k_0} n^{-\delta} \qquad n \geq N(\delta, \gamma, k_0)$$

$$(ii) \quad P(\hat{k}_n(X_1^n) < k) \leq \exp\{-(n-k_0)[\min_{k' \leq k-1} D(\mathcal{P}_{k'}\|P) - \gamma]\} \qquad n \geq N(\delta, \gamma, k_0)$$

where $D(\mathcal{P}_{k'}\|P) := \inf_{P' \in \mathcal{P}_{k'}} D(P'\|P)$. $\square$

Remarks

1. Any choice of $\delta > 1$ yields strong consistency for our estimator, i.e., $\hat{k}_n(X_1^n) \to k$ P-a.s. Clearly the overestimation probability can be reduced by choosing a larger value of δ but only at the expense of a larger sample size $N(\delta, \gamma, k_0)$ in (ii).

2. Observe that $D(\mathcal{P}_{k'}\|P)$ in (ii) is strictly positive since the closure of $\mathcal{P}_{k'}$ does not intersect $\mathcal{P}_k$ for $k' < k$.

Theorem 2 below establishes that the rate of decay in Theorem 1 (ii) cannot be bettered.

Theorem 2

Let $0 < \alpha < 1$ be given. Let $\tilde{k}_n(X_1^n)$ be any estimator such that for each $P \in \mathcal{P}_k, 1 \leq k \leq k_0 : P(\tilde{k}_n(X_1^n) > k) < \alpha$ for $n \geq N(\alpha, k_0, P)$. Then for any $\gamma > 0$ it holds that:

$$P(\tilde{k}_n(X_1^n) < k) \geq \exp\{-(n-k_o)[\min_{k' \leq k-1} D(\mathcal{P}_{k'}\|P) - \gamma]\}$$

for $n \geq N(\alpha, \gamma, k_0, P)$. $\square$

References

[1] I. Basawa and B. Prakasa Rao, *Statistical Inference for Stochastic Processes*, Academic Press, 1980.

[2] M. Gutman, Asymptotically Optimal Classification for Multiple Tests with Empirically Observed Statistics," *IEEE Trans. on Inform. Theory*, *IT-35*, pp. 401-408, 1989.

[3] N. Merhav, M. Gutman, and J. Ziv, "On the estimation of the order of a Markov chain and universal data compression," *IEEE Trans. on Inform. Theory*, *IT-35*, pp. 1014-1019, Sept. 1989.

[4] J. Rissanen, "Complexity of strings in class of Markov sources," *IEEE Trans. Inform. Theory*, *IT-32*, pp. 526-532, 1986.

[5] H. Tong, "Determination of the order of a Markov chain by Akaike's information criterion," *J. Appl. Prob.*, vol. 12, pp. 488-497, 1975.

This research was supported by the Institute for Systems Research at the University of Maryland, College Park, under NSF Grant OIR-85-00108.

L. Finesso is with LADSEB-CNR, Padova, Italy. He is at present on leave of absence at the Institute for Systems Research, University of Maryland, College Park, MD 20742, USA.

C.-C. Liu is with IBM Corporation, Poughkeepsie, NY, 12062, USA.

P. Narayan is with the Electrical Engineering Department and the Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

On the Convergence of the EM Algorithm

A.O. Hero
Dept. of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109-2122

ABSTRACT

The EM algorithm is a popular iterative method for finding the maximum likelihood estimate when the likelihood function is either non-analytical or its functional form is too difficult to maximize directly. In this paper we analyze the convergence properties of the EM algorithm. By representing the E step in a Taylor series with remainder we obtain a derivation of region of convergence and asymptotic convergence rates for a specified complete data space. These results can help one tailor the choice of complete data space so as to achieve an optimal tradeoff between ease of implementation and rapid convergence of the EM algorithm.

I. Main Results

Let θ denote a point in parameter space $\Theta \subset \mathbf{R}^p$ which parameterizes the density $f(\mathbf{y};\theta)$ of the set of observations $\mathbf{Y}$. Now define a hypothetical data set $\mathbf{X}$ with density $g(\mathbf{x};\theta)$ which is related to the actual data $\mathbf{Y}$ in the sense that the conditional distribution $dP(\mathbf{y}|\mathbf{x};\theta)$ is functionally independent of θ. Equivalently $\mathbf{Y}$ can be interpreted as the output of a θ-independent communications channel C with input $\mathbf{X}$. $\mathbf{X}$ is called the *complete data* and $\mathbf{Y}$ is called the *incomplete data*. The EM algorithm has been widely applied to iteratively approximate the maximum likelihood estimate $\hat{\theta} = \mathbf{argmax}_\theta \ln f(\mathbf{Y};\theta)$ [1, 7, 4, 5, 2, 6]. For an initial point θ^0 the EM algorithm produces a sequence of points $\{\theta^i\}_{i=1}^{\infty}$ via a recursion whose form is equivalent to [3]:
EM Algorithm:

$$\theta^{i+1} = \mathbf{argmax}_\theta Q(\theta;\theta^i), \qquad i = 1, 2, \ldots \tag{1}$$

where $Q(\theta;\overline{\theta})$ is the difference between the incomplete data likelihood function $L(\theta) = \ln f(\mathbf{y};\theta)$ and the Kullback-Liebler information divergence:

$$Q(\theta;\overline{\theta}) \stackrel{\text{def}}{=} L(\theta) - D(\theta\|\overline{\theta}),$$

where

$$D(\theta\|\overline{\theta}) \stackrel{\text{def}}{=} \int \log \frac{g(\mathbf{x}|\mathbf{y};\overline{\theta})}{g(\mathbf{x}|\mathbf{y};\theta)} g(\mathbf{x}|\mathbf{y};\overline{\theta}) d\mathbf{x}. \tag{2}$$

For any non-negative definite symmetric matrix A define the spectral radius $\rho(A)$ as the maximum eigenvalue of A. For $\theta \in \Theta$ define the Hessian matrices:

$$\begin{aligned} \mathbf{Q} &\stackrel{\text{def}}{=} -\nabla^{20} Q(\hat{\theta};\hat{\theta}) \\ \mathbf{D} &\stackrel{\text{def}}{=} -\nabla^{20} D(\hat{\theta};\hat{\theta}) \\ \mathbf{L} &\stackrel{\text{def}}{=} -\nabla^{2} L(\hat{\theta}). \end{aligned}$$

Define $\mathcal{R}_+ \subset \Theta$ as the largest open ball with center $\hat{\theta}$ such that for each $\overline{\theta} \in \mathcal{R}_+$:

$$-\int_0^1 (1-t)\nabla^{20} D(t\theta + (1-t)\hat{\theta}\|t\overline{\theta} + (1-t)\hat{\theta}) dt > 0, \qquad \forall \theta \tag{3}$$

In the sequel $\mathcal{R}_+$ will be identified as a set of initial points θ^0 for which the EM algorithm is guaranteed to converge which is identified with a *region of convergence* in the following theorem.

Theorem 1 *Assume: i) the MLE, $\hat{\theta} = \mathbf{argmax}_\theta L(\theta)$, occurs in the interior of the parameter space $\Theta \subset \mathbf{R}^p$; ii) $L(\theta)$ and $D(\theta\|\overline{\theta})$ are twice continuously differentiable in θ and $\overline{\theta}$. For θ^0 an initial point let $\{\theta^i\}_{i=1}^{\infty}$ denote the sequence of values produced by the EM algorithm. Then:*

1. *if $\theta^0 \in \mathcal{R}_+$ the EM sequence converges to $\hat{\theta}$,*
2. *and the asymptotic convergence rate is linear with root convergence factor $\rho(\mathbf{I} - \mathbf{Q}^{-1}\mathbf{L}) = \rho(\mathbf{QD})$.*

Theorem 1 is proven via a simple application of Taylor's Theorem with remainder. While Theorem 1 requires stronger assumptions (differentiability of L and D) than the convergence results stated in [8], our proof is more elementary and we come up with a region of convergence $\mathcal{R}_+$ for $\{\theta_k\}_{k\geq 0}$. One can apply the results of Theorem 1 to compare different choices of complete data in terms of radius of convergence and and speed of convergence.

REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Society, Ser. B*, vol. 39, pp. 1–38, 1977.

[2] M. Feder, A. Oppenheim, and E. Weinstein, "Maximum likelihood noise cancelation using the EM algorithm," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. 37, no. 2, pp. 204–216, Feb. 1989.

[3] A. O. Hero and J. A. Fessler, "Convergence properties of the EM algorithm," Technical Report in prep., Comm. and Sig. Proc. Lab. (CSPL), Dept. EECS, University of Michigan, Ann Arbor.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomography*, vol. 8, no. 2, pp. 306–316, April 1984.

[5] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: applications to estimating point-process intensities and Toeplitz constrained covariances," *IEEE Proceedings*, vol. 75, no. 7, pp. 892–907, July 1987.

[6] M. Segal, E. Weinstein, and B. Musicus, "Estimate-maximize algorithms for multichannel time delay and signal estimation," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. 39, no. 1, pp. 1–16, Jan. 1991.

[7] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-1, No. 2, pp. 113–122, Oct. 1982.

[8] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Annals of Statistics*, vol. 11, pp. 95–103, 1983.

Necessary and Sufficient Conditions of Channel Identifiability Based on Second-Order Cyclostationary Statistics

Lang Tong
Dept of Electrical & Computer Eng.
West Virginia University
WV 26506-6101

Guanghan Xu
Dept of Electrical & Computer Eng.
The University of Texas at Austin
Austin, TX 78712

Thomas Kailath
Information Systems Laboratory
Stanford University
Stanford, CA 94305

1 Introduction

Consider the following communication model

$$x(t) = \sum_k s_k h(t - kT) + n(t), \tag{1}$$

where $h(\cdot)$ is the channel impulse response; $\{s_k\}$ and T are the information symbol sequence and symbol interval, respectively. *The "blind" identification problem addressed in this paper is the identifiability of a possibly nonminimum phase channel $h(\cdot)$ given only the observation process $x(\cdot)$.*

The following assumptions are imposed to the above model:
(1) $\{s_k\}$ is an i.i.d. sequence.
(2) The symbol interval T is an integer.
(3) The channel has a finite impulse response. (4) The noise process is uncorrelated with $\{s_k\}$ with known second-order statistics.

2 A Necessary and Sufficient Condition in Frequency Domain

Under the assumed condition, the observation process $x(\cdot)$ is a cyclostationary process. Different from the stationary case, the second-order statistics contain the phase information of the channel. The identification of $H(z)$ is approached by identifying its zeros from those of $\{\Gamma^{(k)}(z)\}$, where $\{\Gamma^{(k)}(z)\}$ is obtained from the observation spectra. The relation between $\Gamma^{(k)}(z)$ and $H(z)$ is given by

$$\Gamma^{(k)}(z) = H(z)H^*(e^{j k \frac{2\pi}{T}} \frac{1}{z^*}), k = 1, 2, \cdots. \tag{2}$$

The problem of channel identification is then equivalent to identifying $H(z)$ by $\Gamma^{(k)}(z)$.

The following theorem provides a necessary and sufficient condition for the channel identifiability.

Theorem 1 *$H(z)$ is uniquely determined (identifiable) by $\{\Gamma^{(k)}(z)\}$ up to a constant if and only if $H(z)$ does not have uniformly $\frac{2\pi}{T}$ - spaced zeros. More over, if the channel is identifiable,*

$$\mathcal{Z}(H(z)) = \bigcap_k \mathcal{Z}(\Gamma^{(k)}(z)), \tag{3}$$

where $\mathcal{Z}(H(z))$ stands for the set of zeros of $H(z)$.

3 A Necessary and Sufficient Condition in Time Domain

A time-domain necessary and sufficient condition of channel identifiability is obtained by using a vector representation of the baseband model

$$\mathbf{x}(n) = [x(nT)x(nT+1)\ldots x(nT+T-1)]^t. \tag{4}$$

We then have, from the channel model,

$$\mathbf{x}(n) = \sum_k s_k \mathbf{h}_{n-k} + \mathbf{n}_n, \tag{5}$$

where

$$\mathbf{h}_k = [h(kT)h(kT+1)\ldots h(kT+T-1)]^t, \tag{6}$$
$$\mathbf{n}_k = [n(kT)n(kT+1)\ldots n(kT+T-1)]^t. \tag{7}$$

With the above formulation, we have the following theorem.

Theorem 2 *The channel impulse response can be determined uniquely up to a constant if and only if there exists an integer d such that matrix $\mathbf{H}^{(d)}$ has a full column rank, where*

$$\mathbf{H}^{(d)} = \underbrace{\begin{pmatrix} \mathbf{h}_0 & \mathbf{h}_1 & \ldots & \mathbf{h}_L & 0 & \ldots & 0 \\ 0 & \mathbf{h}_0 & \mathbf{h}_1 & \ldots & \mathbf{h}_L & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \mathbf{h}_0 & \mathbf{h}_1 & \ldots & \mathbf{h}_L \end{pmatrix}}_{d}. \tag{8}$$

The proof of the above theorem establishes the connection between the rank condition and the condition involving the location of the zeros. The sufficient part of this theorem is equivalent to the one [1].

References

[1] L. Tong, G. Xu, and T. Kailath. "Blind identification and equalization based on second-Order statistics: A time domain approach". *Submitted to IEEE Trans. Information Theory.*

ACKNOWLEDGEMENT

This research was supported in part by the Joint Service Program at Stanford University (U.S. Navy, and U.S. Air Force) under Contract DAAL03-91-C-0011 and by SDIO/IST, managed by the Army Research Office under Contract DAAL03-90-G-0108.

ENTROPY AND THE CONSISTENT ESTIMATION OF JOINT DISTRIBUTIONS

Katalin Marton
Mathematics Institute
Hungarian Academy of Sciences

Paul C. Shields
University of Toledo and
Eötvös Loránd University

The kth-order joint distribution for an ergodic finite-alphabet process can be estimated from a sample path of length n by sliding a window of length k along the sample path and counting frequencies of k-blocks. If k is fixed the procedure is consistent in that the resulting empirical k-block distribution will almost surely converge to the true distribution of k blocks as $n \to \infty$, a fact guaranteed by the ergodic theorem. The consistency of such estimates is important when using training sequences, that is, finite sample paths, to design engineering systems. The empirical k-block distribution for a training sequence is used as the basis for design, after which the system is run on other, independently drawn sample paths. There are some situations, such as data compression, where it is good to make the block length as long as possible. Thus it would be desirable to have consistency results for the case when the block length function $k = k(n)$ grows as rapidly as possible, as a function of sample path length n. This is the problem addressed in this paper.

A sequence $\{k(n)\}$ will be said to be *admissible* for a given ergodic process μ if the variational distance between the true distribution and the empirical distribution of $k(n)$-blocks converges almost surely to 0 as $n \to \infty$. Every ergodic process has an admissible sequence such that $\lim_n k(n) = \infty$, by the ergodic theorem, and, for any sequence $k(n) \to \infty$ there is an ergodic measure for which $\{k(n)\}$ is not admissible.

Entropy plays a role in this problem, because if $k(n) \geq (1 + \epsilon)(\log n)/H$, then the empirical k-block distribution cannot be close to the true distribution, for, by the entropy theorem, most of the probability is concentrated on a set of k-blocks of cardinality $2^{k(H+\epsilon/2)}$. Thus the interesting question is whether there are any processes for which consistent estimation is possible if $k(n) \sim (1 - \epsilon(\log n)/H$. It is shown in this paper that the answer is yes for the class of Markov processes as well as for somewhat larger classes, such as the class of finite state processes, and in a slightly weaker form for the class of processes for which past and future become asymptotically independent in the weak Bernoulli sense. The proofs depend on an extension of the Sanov-Hoeffding large deviations bound, together with an inequality due to Pinsker.

For the class of functions of Markov chains, this work sharpens prior results obtained for more general classes of processes by Ornstein and Weiss and by Ornstein and Shields which used the $\bar{d}$-distance rather than the variational distance.

Extensions and applications of these results will also be discussed.

Acknowledgements. The authors were partially supported by the Hungarian National Foundation for Scientific Research Grant OTKA 1906 and by NSF grant DMS-9024240.

A New Bound on the Estimation of the Probability Density Function Using Spectral Analysis

Marcelo S. Alencar *†
Universidade Federal da Paraiba
Departamento de Engenharia Eletrica
Av. Aprigio Veloso, 882
58.100 Campina Grande PB Brasil

Abstract

A new upper bound is introduced on the estimation of the probability density function through spectral analysis. The upper bound is shown to decrease steadly as the modulating index is increased, for the Gaussian case.

Summary

Estimation of the probability density function (pdf) of stochastic process is commonly based on a time series approach [1]. This is done by measurement of the time spent by the signal between two specified levels or through a pulse counting process, for discrete signals. This usually leads to biased and inconsistent estimates, and to mean square errors that depend on the pdf itself [2]. It is a common practice to assume the stationarity and ergodicity of the random process into analysis [3].

The main purpose of this paper is to present a new bound on the estimation of the probability density function of random signals, using Woodward's theorem, correlation techniques and spectral analysis [4] [5]. The proposed method is based on the spectral analysis of the random process.

Woodward's theorem asserts that the spectrum of a high-index frequency modulated waveform can be approximated by the probability distribution of its instantaneous frequency deviation [5]. A new proof of the theorem was developed previously, which gave the following result for the power spectrum density (PSD) of the modulated signal [6].

$$\hat{S}_S(w) = \frac{A^2}{2D}[p_M(\frac{w + w_c}{D}) + p_M(\frac{w - w_c}{D})]. \quad (1)$$

where the constant parameters A, w_c and D represent respectively the carrier amplitude, frequency (rd/s) and frequency deviation index. The signal whose pdf ($p_M(m)$)one intends to analyse is represented by $m(t)$, here considered a zero mean random stationary process, limited in frequency to w_M. The phase of the carrier ϕ is random, uniformly distributed in the range $(0, 2\pi)$ and statistically independent of $m(t)$.

The difference between the estimate of the PSD function and the actual PSD is the estimation error E_S. An upper bound for this error is evaluated below and is shown to decrease with the an increase in the modulation index β. The approximation error is given by

$$E_S(\beta) = S_S(w) - \hat{S}_S(w) \quad (2)$$

where $S_S(\omega)$ represents the actual spectrum and $\hat{S}_S(\omega)$ stands for the approximation.

Considering the limiting case ($\tau = \pi/\beta w_M = 1/\beta f_M$), an upper bound on the normalised error can be determined. Substituting the expressions for $S_S(w)$ and $\hat{S}_S(w)$ into equation 2, evaluating the expectancies at $w_c = 0$ and using the following inequality [7]

$$\begin{aligned} \tau^2 E[(v'(t))^2] &\geq E[(v(t+\tau) - v(t))^2] \\ &\geq \frac{4\tau^2}{2\pi^2} E[(v'(t))^2] \end{aligned} \quad (3)$$

leads to

$$E_S(\beta) \leq [\frac{\pi - 2}{\beta w_M}][1 - 2Q(\frac{\pi}{\beta w_M})] \quad (4)$$

where $Q(x)$ is the Q-function, defined by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy. \quad (5)$$

The above expression is a very tight bound and shows the error dependency on the modulating index β. The efficiency of the estimation used is assured because the variance goes to zero. The estimation always gets better as the modulating index is increased, which implies a decrease in frequency or an increase in the power of the signal [8]. This also implies the consistency of the method. All the relevant information is available for the estimation, giving sufficiency to the estimator.

A digital computer implementation of the method was performed through contract No. C.NE.085.16 with EMBRATEL, and has been tested in practice with good results [9].

References

[1] G. Mirsky. *Radioelectronic Measurements.* Mir Publishers, Moscow, 1978.

[2] M. B. Priestley. *Spectral Analysis and Time Series.* Academic Press, London, 1981.

[3] M. Schwarts and L. Shaw. *Signal Processing: Discrete Spectral Analysis, Detection, and Estimation.* McGraw-Hill, Tokyo, 1975.

[4] N. M. Blachman and G. A. McAlpine. "The Spectrum of a High-Index FM Waveform: Woodward's Theorem Revisited". *IEEE Transactions on Communications Technology*, 17(2), April 1969.

[5] P. M. Woodward. "The Spectrum of Random Frequency Modulation". Memo. 666, Telecommunications Research Establishment, Great Malvern, England, December 1952.

[6] Marcelo S. Alencar and Benedito G. Aguiar Neto. "Estimação da Função Densidade de Probabilidade Através da Análise Espectral". In *Anais do Simpósio Brasileiro de Telecomunicações*, pages 10.3.1–10.3.5, São Paulo, Brasil, 1991.

[7] A. Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, Tokyo, 1981.

[8] Marcelo S. Alencar. "Measurement of the Probability Density Function of Communication Signals". In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference - IMTC'89*, pages 513–516, Washington, D. C., April 1989.

[9] Marcelo S. Alencar. "Estimação da Densidade Espectral de Potência do Sistema FDM-FM da EMBRATEL". In *Anais do Simpósio Brasileiro de Telecomunicações*, pages 297–300, Campina Grande, Brasil, Setembro 1988.

*This work was partially supported by CNPq and EMBRATEL.

†The author is currently with the Department of Electrical and Computer Engineering, University of Waterloo, Canada.

A Cramer-Rao Type Lower Bound for Estimators Satisfying a Bias Constraint*

Alfred Hero

In this paper we give a Cramer-Rao (CR) type lower bound on estimator covariance which applies to any estimator whose bias gradient lies within a user specified ellipsoidal region of parameter space. In addition to providing a useful lower bound which is insensitive to small unknown estimator biases, the rate of change of the new bound provides a quantitative bias "sensitivity index" for the conventional bias-dependent CR bound. We give an analytical form for this sensitivity index which indicates that small estimator biases can make the new bound significantly less than the unbiased version of the CR bound when there exist important but difficult-to-estimate nuisance parameters. This implies that the application of the CR bound is unreliable for this situation due to severe bias sensitivity. As a practical illustration of these results, we consider the problem of estimating elements of the 2×2 covariance matrix associated with a pair of independent, identically distributed, zero-mean Gaussian random sequences.

Reference

A. O. Hero, "A Cramer-Rao type lower bound for essentially unbiased parameter estimation," MIT Lincoln Laboratory, Lexington, Mass., Technical Rep. 890, (3 January 1992). DTIC AD-A246666.

* This work was supported by M.I.T. Lincoln Laboratory under Air Force Contract F19628-90-C-0002. Dr. Hero is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122.

NON-LINEAR, NON-BINARY CYCLIC GROUP CODES

G. Solomon
Los Angeles, CA

Abstract

New cyclic group codes of length 2^m -1 over $(m-j)$-bit symbols are introduced. These codes may be systematically encoded and decoded algebraically. The code rates are very close to RS codes and are much better than BCH codes (a former alternative). The $(m-j)$-binary tuples are identified with a sub-group of the binary m-tuples which represent the field GF(2^m). Encoding is systematic and involves a two stage procedure, the usual linear feedback register (using the division or check polynomial), and a small table look up. For low rates, a second shift register encoding operation may be invoked. Decoding uses the Reed-Solomon error correcting procedures for the m-tuple alphabet, i.e., the field elements GF(2^m).

SUMMARY

Group codes of lengths up to 2^m over binary $(m-1)$ tuples are first introduced and are shown to be cyclic and then systematically encodable. These $(m-1)$-tuples are identified with an additive subgroup of the field GF(2^m). These codes are not linear. That is, a codeword does not admit multiplication by a GF(2^m) field element to yield another codeword.

Consider the field GF(2^m) along with a primitive element β which generates the $n = (2^m - 1)$ roots of unity. In addition, β is chosen with the following properties: 1) m odd: Tr $\beta^i = 0$ for $1 \le i \le m-1$, where Tr denotes the linear field operator trace. Tr $\beta = \beta + \beta^2 + \beta^4 + \cdots + \beta^{2^{m-1}}$. So Tr $\beta \in GF(2)$, Tr β^2 = Tr β , Tr cx^2 = Tr $\sqrt{c}x$, for $c, x \in$ GF(2^m). 2) m even: Tr $\beta^i = 0$ for $0 \le i \le m-1$ except for a single odd integer p , $p < m$, and Tr $\beta^p = 1$.

The following are polynomials for β which satisfy the conditions 1) and 2) above for $3 \le m \le 12$.

m	*Polynomial for* β	*explanation*
3	3 1 0	$(x^3 + x + 1)$
4	4 1 0	
5	5 3 0	
6	6 1 0	(Tr $\beta^5 = 1$)
7	7 3 0	
8	8 4 3 2 0	(Tr $\beta^5 = 1$)
9	9 5 0	
10	10 3 0	(Tr $\beta^7 = 1$)
11	11 9 0	
12	12 6 4 1 0	(Tr $\beta^{11} = 1$)

Codes of length greater than 4096 are rarely invoked in present day block coding techniques. Do these properties extend beyond $m = 12$?

An element $c \in$GF(2^m) may be represented by $c = \sum_{i=0}^{m-1} c_i\beta^i$ One may identify Tr c by its binary representation (c_i); $0 \le i \le m-1$ and single out c_0 for m odd, and c_p for m even. Thus the binary value Tr c is determined by only the trace one position (0 or p) in its binary m-bit representation. Choose an $(n, k; d)$ Reed-Solomon code over GF(2^m) so that the codewords are values of sets of polynomials $P(x)$ with coefficients in GF(2^m) of fixed highest degree $(n-d)$ or $(n-d-1)$. A codeword $\mathbf{a} = (a_j)$ is represented by the values of a polynomial $P_{\mathbf{a}}(x)$ so that $a_j = P_{\mathbf{a}}(\beta^j)$, $0 \le j \le n-1$.

Restrict $P_{\mathbf{a}}(x)$ for all codewords $\mathbf{a}$ to a $(m-1)$ order subgroup of GF(2^m) by stipulating that Tr $P(x) = 0$ for $x \in$GF(2^m). ($P(x)$ as writtten here is generic for all $P_{\mathbf{a}}(x)$). The codes thus generated are cyclic group codes over $(m-1)$-bit symbols and are systematically encodable for codes meeting the conditions in the main theorem.

Examples:

1. Take the RS code **A** of dimension 5 over GF(2^m), $\mathbf{a} \in \mathbf{A}$, $\mathbf{a} = (a_i)$, $a_i = P_{\mathbf{a}}(\beta^i)$. The polynomials $P_{\mathbf{a}}(x)$ are of degree 4 with Tr $P_{\mathbf{a}}(x) = 0$, for all $x \in$ GF(2^m). For a general $P(x)$, dropping the subscript $\mathbf{a}$, $P(x) = A + Bx + Cx^2 + Dx^3 + Ex^4$; $A, B, C, D, E \in$ GF(2^m). The conditions that Tr $P_{\mathbf{a}}(x) = 0$ gives Tr $A = 0$, $B^4 + C^2 + E = 0$, $D = 0$.

   This code has binary dimension $(m-1) + 2m$.

   For $m = 3$, we get binary dimension 8 or dimension 4 over 2-tuples. i.e., a (7, 4; 3) code over binary doubles. This is a reduction from the (7, 5; 3) RS code over binary triples!

   There exist no integer dimension over $(m-1)$-tuples for $m > 3$ since $(m-1) + 2^m$ is not a multiple of $(m-1)$.

2. Take a RS code of dimension 11 over GF(16) but choose as your Mattson-Solomon (M-S) set the polynomials $P(x)$ of degree 11, setting the constant term equal to zero.

   $P(x) = \sum_{i=1}^{11} c_i x^i$; Tr $P(x) = 0$ leads to

   $c_1^8 + c_2^4 + c_4^2 + c_8 = 0$
   $c_3^4 + c_6^2 + c_9 = 0$
   $c_5 + c_5^4 + c_{10}^2 + c_{10}^8 = 0$
   $c_{11}^2 + c_7 = 0$

   The number of binary dimensions is 12 + 8 + 6 + 4 = 30 which is dimension 10 over binary triples. Thus the (15, 11; 5) RS code over GF(16) is transformed into the non-systematic (15, 10; 5) code over trace zero elements of GF(16).

3. Similarly the RS (15, 7; 9) code over GF(16) using polynomials of degree 6 from 0 to 6, under analogous techniques, gives the relations Tr $c_0 = 0$; $c_1^4 + c_2^2 + c_4 = 0$; $c_6 + c_3^2 = 0$; $c_5 = 0$.

   Binary dimension count is 3 + 8 + 4 = 15. This yields a (15, 5; 9) code over triples.

Compare this to

- (15, 5; 11) RS over 4-tuples,
- (15, 5; 7) BCH over GF(8) and GF(2).
- (15, 4; 10) BCH over GF(4). (doubles).

These non-systematic codes are cyclic. The extension to $(m-j)$-bit symbols and the systematic construction of these codes can be found in "Nonlinear, Nonbinary Cyclic Codes" by G. Solomon NASA Code 310-10-63-53-00 TDA Progress Report 42-108 Jet Propulsion Laboratory October-December 1991.

*This work was performed by the author while acting as a consultant to the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration

Classification of Cosets of the Reed-Muller Code $R(m-3,m)$

Xiang-dong Hou
Department of Mathematics and Statistics
Wright State University
Dayton, Ohio 45435

The covering radius of the $(m-3)$rd order Reed-Muller code $R(m-3,m)$ of length 2^m has been known. This talk aims at a complete classification and further properties of the cosets of $R(m-3,m)$.

The general affine group $GA(m,2)$ is an automorphism group of $R(m-3,m)$, and is the full automorphism group when $m \geq 4$. Hence $GA(m,2)$ acts on the set $\mathcal{C}$ of all cosets of $R(m-3,m)$. The cosets of even weight in $\mathcal{C}$ correspond to $m \times m$ symmetric matrices over $GF(2)$, and their $GL(m,2)$-orbits correspond to the congruence classes of $m \times m$ matrices over $GF(2)$. The same thing happens with respect to the cosets of odd weight in $\mathcal{C}$. Using the well-known classification of symmetric matrices over $GF(2)$ under congruence, we get the calssification of $\mathcal{C}$ under the action of $GL(m,2)$. The classification of $\mathcal{C}$ under the action of $GA(m,2)$ follows immediately. Representatives and sizes of the $GA(m,2)$-orbits in $\mathcal{C}$ are given. The minimal weights of the cosets in $\mathcal{C}$ are determined.

We also identify all the orphan cosets of $R(m-3,m)$. It turns out that all the orphans of $R(m-3,m)$ are 0-covered, i.e., for any orphan C of $R(m-3,m)$ and any coordinate position, there is a coset leader of C whose coordinate at the given position is 0. This implies that $R(m-3,m)$ is normal.

Finally, we turn to the weight distributions of the cosets in $\mathcal{C}$. For any $C \in GF(2)^{2^m}$, we derive a general recursive formula for computing the weight distribution of C. The recursion starts at the minimal weight μ of C. When $w \geq \mu$ is not far away from μ, the formula gives the number of vectors of weight w in C rather easily in terms of certain functions $|A_\zeta(C)|$ of C. $|A_\zeta(C)|$ is difficult to compute for general C. However, when C is one of those representatives of the $GA(m,2)$-orbits in $\mathcal{C}$, we are able to give explicit formulas for $|A_\zeta(C)|$.

IDEMPOTENTS AND MINIMUM WEIGHTS OF PRIME POWER LENGTH CYCLIC CODES OVER ARBITRARY FIELDS

Vanessa Job, Marymount University, Arlington, VA 22207

Let p be a prime and let q be a prime power relatively prime to p. Let z be the greatest integer such that $p^z|(q^t - 1)$ where t is the order of q modulo p. Assume that p and q have been chosen so that $z = 1$. (Note that is very unusual to have $z > 1$.) We give a characterization of idempotents of length p^{m+1} cyclic codes over $GF(q)$ in terms of the idempotents of length p cyclic codes over $GF(q)$. We define two classes of length p^{m+1} cyclic codes, the repeated p codes and the expanded p^m codes, which are derived from length p and p^m cyclic codes, respectively, and give the idempotents of these codes in terms of the idempotents of the codes from which they were derived. We also give the weight enumerators of codes in these classes as a function of the weight enumerators of the codes from which they were derived. Finally, we show that every length p^{m+1} code can be uniquely expressed as a sum of a repeated p code C_1 and an expanded p^m code C_2 and show that the sum must have minimum weight less than or equal to $\min(p^m d_1, d_2)$ where d_1 is the minimum weight of of the length p code from which C_1 was derived and d_2 is the minimum weight of the length p^m code from which C_2 was derived. Using these results, we give an algorithm for constructing idempotents for prime power length duadic, triadic, and polyadic codes, generalizations of quadratic residue codes to nonprime lengths n and dimensions other $(n-1)/2$ and $(n+1)/2$. We show that the minimum weight of prime power length polyadic codes is unlikely to be greatest possible, distinguishing them from polyadic codes of square free length, which frequently have greatest possible or greatest possible known minimum weight for codes of their length and dimension.

Constructing Reed-Muller Codes from Reed-Solomon Codes over $GF(q)$

Frank R. Kschischang

1 Summary

The $[q,k]$ extended Reed-Solomon codes over $GF(q)$ are nested, that is,

$$[q,q] \supset [q,q-1] \supset \cdots \supset [q,1] \supset [q,0] = \{\mathbf{0}\}.$$

This means that the $[q,k-1]$ code is a subgroup of the $[q,k]$ code and hence the $[q,k]$ code may be *partitioned* into the q cosets of the $[q,k-1]$ code. This code, in turn, may be partitioned into the q cosets of the $[q,k-2]$ code, etc., thus forming the set partition chain $[q,q]/[q,q-1]/\cdots/[q,1]/[q,0]$. Since these codes are all maximum distance separable or MDS (the minimum Hamming distance of the $[q,k]$ code is $q-k+1$), the intrasubset distances form the sequence $\{1,2,3,\ldots,q,\infty\}$, where ∞ is used to denote the intrasubset distance of a set with one element (a singleton).

Let g_1 be a nonzero codeword in the $[q,1]$ code. Then g_1 is a generator for the code. Let g_2 be a nonzero codeword in the $[q,2]$ code that is not in the $[q,1]$ code, i.e., an element of the relative complement of $[q,1]$ in $[q,2]$. Then $\{g_1,g_2\}$ generates the $[q,2]$ code. Proceeding in this way we can obtain a *universal generator matrix*

$$G = \begin{bmatrix} g_q \\ g_{q-1} \\ \vdots \\ g_1 \end{bmatrix},$$

the last k rows of which generate the $[q,k]$ code.

The basic *multilevel* (sometimes called generalized concatenated or hierarchical) code construction technique (see, e.g., [1, 2]) is based on precisely the type of set partitioning described above. The construction combines q *component* codes of block length n to obtain (in this case) a code of length nq. Although not necessary for the construction, we consider here only linear codes. Denoting the parameters of the component codes by $[n,k_1,d_1]$, $[n,k_2,d_2]$, $\ldots$, $[n,k_q,d_q]$, the multilevel construction combines these codes to obtain an

$$[nq, k_1+k_2+\cdots+k_q, d = \min(d_1, 2d_2, 3d_3, \ldots, qd_q)].$$

code. Wu and Costello used this construction method in [3] to obtain new codes over $GF(q)$. In "code formula" form, we have

$$C = [n,k_1,d_1] \otimes g_q + [n,k_2,d_2] \otimes g_{q-1} + \cdots + [n,k_q,d_q] \otimes g_1$$

where "$\otimes$" denotes a Kronecker product.

For example, when $q=2$, we take $g_1 = 11$ and $g_2 = 01$. Applying the multilevel construction, we combine two codes $V = [n,k_1,d_1]$ and $U = [n,k_2,d_2]$ to obtain

$$C = [n,k_1,d_1] \otimes 01 + [n,k_2,d_2] \otimes 11.$$

If V has generator matrix G_V and U has generator matrix G_U, then C has generator matrix

$$G_C = \begin{bmatrix} 0 & G_V \\ G_U & G_U \end{bmatrix}.$$

By the basic properties of the multilevel construction, C has minimum distance $d = \min(d_1, 2d_2)$. Clearly this is the well-known $(U|U+V)$ construction [4].

Similarly, the first three rows of Pascal's triangle, reduced modulo 3, form a universal generator matrix for a family of nested ternary MDS codes of block length 3. Applying the basic multilevel construction method results in a ternary "$(U|2U+V|U+V+W)$" code construction method [5], in which $C = U \otimes 121 + V \otimes 110 + W \otimes 100$.

This construction is easily extended to $GF(p)$ for any prime p. Massey *et al.* have shown in [6] that the first p rows of Pascal's triangle reduced modulo p form a universal generator matrix for a family of nested MDS codes of block length p over $GF(p)$. Thus, the basic multilevel construction method may be applied to this case.

For arbitrary fields $GF(q)$, we construct q-ary "Reed-Muller" codes as follows. Beginning with the bi-infinite partition chain

$$\cdots/[1,1]/[1,1]/[1,1]/[1,0]/[1,0]/[1,0]/\cdots$$

we apply the multilevel construction technique taking as component codes every sequence of q consecutive codes in the partition chain. From this we obtain the bi-infinite partition chain

$$\cdots/[q,q]/[q,q]/[q,q-1]/\cdots/[q,1]/[q,0]/[q,0]/\cdots.$$

We then apply the multilevel construction technique to this chain of codes, resulting in a chain of codes of block length q^2. This process may be repeated indefinitely, resulting in a family of block codes having block lengths that are integer powers of q.

In our work we derive formulas for the dimension and minimum distance of these codes, investigate their duality properties, and show that these codes are natural analogs of the binary Reed-Muller codes.

[0]The author is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, CANADA M5S 1A4, tel: (416) 978-0461, fax: (416) 978-7423, e-mail: frank@comm.toronto.edu

References

[1] V. A. Zinoviev, "Generalized cascade codes," *Probl. Peredach. Inform.*, vol. 12, pp. 2-9, 1976.

[2] E. Biglieri and A. Spalvieri, "Generalized concatenation: A tutorial," in *Coded Modulation and Bandwidth-Efficient Transmission* (E. Biglieri and M. Luise, eds.), pp. 27-39, Elsevier Science Publishers, B.V., Amsterdam, The Netherlands, 1992.

[3] J. Wu and D. J. Costello, Jr., "New multilevel codes over $GF(q)$," *IEEE Trans. on Inform. Theory*, vol. 38, pp. 933-939, May 1992.

[4] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. New York: North-Holland, 1977.

[5] F. R. Kschischang and S. Pasupathy, "Some ternary and quaternary codes and associated sphere packings," *IEEE Trans. on Inform. Theory*, vol. 38, pp. 227-246, March 1992.

[6] J. L. Massey, D. J. Costello, Jr., and J. Justesen, "Polynomial weights and code constructions," *IEEE Trans. on Inform. Theory*, vol. IT-19, pp. 101-110, Jan. 1973.

ON THE APPARENT DUALITY OF THE KERDOCK AND PREPARATA CODES

ROGER HAMMONS AND P. VIJAY KUMAR

ABSTRACT. The Kerdock and Preparata codes are something of an enigma in coding theory since they are both Hamming distance invariant and have weight enumerators that are dual under the MacWilliams transform just as if they were dual linear codes. In this paper, we explain, by constructing in a natural way a Preparata-like code $\mathcal{P}_L$ from the Kerdock code $\mathcal{K}$, why the existence of a distance-invariant code with weight distribution that is the MacWilliams transform of that of the Kerdock code is only to be expected. The construction involves quaternary codes over the ring $\mathbf{Z}_4$ of integers modulo 4. We exhibit a quaternary code $\mathcal{Q}$ and its quaternary dual $\mathcal{Q}^\perp$ which, under the Gray map, give rise to the Kerdock code $\mathcal{K}$ and Preparata-like code $\mathcal{P}_L$, respectively. The code $\mathcal{P}_L$ is identical in weight and distance distribution to the Preparata code. The linearity of $\mathcal{Q}$ and $\mathcal{Q}^\perp$ ensures that $\mathcal{K}$ and $\mathcal{P}_L$ are distance invariant, while their duality as quaternary codes guarantees that $\mathcal{K}$ and $\mathcal{P}_L$ have dual weight distributions.

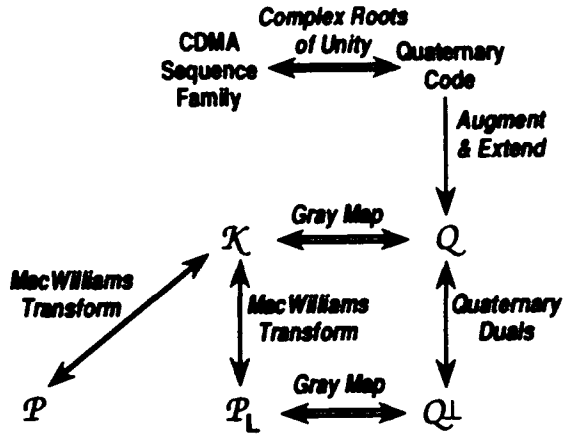


FIGURE 1. QUATERNARY CONNECTIONS

SUMMARY

Recently, a family of nearly optimal four-phase sequences of period $N = 2^r - 1$, r odd, with alphabet $\{1, j, -1, -j\}$, $j = \sqrt{-1}$, was discovered first by Solé [1] and later independently by Boztaş, Hammons, and Kumar [2]. After replacing each complex fourth root-of-unity j^a by its exponent $a \in \{0, 1, 2, 3\}$, this family may be viewed as a linear quaternary code over the ring $\mathbf{Z}_4$ of integers modulo 4. Since the family has low correlation values, it also possesses large *minimum Euclidean* distance and thus the potential for excellent error-correcting capability.

An analysis [2] of the correlation properties of the four-phase sequences led us to consider the 2-adic (*i.e.*, base 2) expansions of the quaternary codewords. Interestingly, these bore a striking resemblance to the original expression [3] for the nonlinear binary Kerdock code. A second connection with the Kerdock code arose during attempts to construct good binary codes from the four-phase sequence family using the Gray map. This was a logical step to pursue as the Gray map translates a quaternary code with large minimum Euclidean distance into a binary code of twice the length having large minimum Hamming distance. The codes that resulted were nonlinear and had the same parameters as shortened versions of the Kerdock code.

In exploring these connections, it was discovered that the original quaternary code could be enlarged in a natural way, as shown in Figure 1, to a linear quaternary code $\mathcal{Q}$ whose image under the Gray map is *precisely* the Kerdock code. It was only natural to consider whether the interesting link between the Kerdock code and a linear quaternary code could also be used to explain the apparent duality of the Kerdock and Preparata codes.

The new perspective does indeed provide an explanation, although not the one that might first be suspected. We show that the binary images $\mathcal{G}(C)$ and $\mathcal{G}(C^\perp)$ under the Gray map of a linear quaternary code C and its $\mathbf{Z}_4$-dual are always Hamming distance invariant. Furthermore, these binary codes have the property that their weight distributions are always dual under the MacWilliams transform. As a consequence, the Kerdock code possesses a *natural* "quaternary-dual" code $\mathcal{P}_L = \mathcal{G}(\mathcal{Q}^\perp)$, identical in size, weight, and distance to the extended Preparata code. Although the Preparata and Preparata-like ($\mathcal{P}_L$) codes have similar finite field transform descriptions, they are in general not the same.

Interestingly, at length 16, the Preparata and Preparata-like codes do coincide. In fact, the Kerdock code, the extended Preparata code, and the Preparata-like code all coincide with the Nordstrom-Robinson code $\mathcal{N}_{16}$. Thus, the Nordstrom-Robinson code can be generalized in one way to get the extended Preparata codes, in another way to get the Kerdock codes, and in yet another way to get the Preparata-like codes!

From the standpoint of decoding, it is not necessary to distinguish between the binary codes and their quaternary parents. An important advantage in working in the $\mathbf{Z}_4$-domain, where the codes are linear, is that it is meaningful to speak of syndromes. Moreover, the codes $\mathcal{Q}$ and $\mathcal{Q}^\perp$ are $\mathbf{Z}_4$-analogs of the binary first-order Reed-Muller code $RM(1, r)$ and its dual $RM(r-2, r)$. This connection makes decoding of the Kerdock and Preparata codes, at least conceptually, easier.

REFERENCES

[1] P. Solé, *A Quarternary Cyclic Code and a Family of Quadriphase Sequences with Low Correlation Properties*, Lecture Notes in Computer Science **388** (1989), 193–201.

[2] S. Boztaş, R. Hammons, and P. V. Kumar, *4-phase Sequences with Near Optimum Correlation Properties*, IEEE Transactions on Information Theory **38** no. 3 (May 1992), 1101–1113.

[3] A. M. Kerdock, *A Class of Low-Rate Nonlinear Binary Codes*, Information and Control **20** (1972), 182–187.

R. Hammons is with Hughes Aircraft Company, 8433 Fallbrook Avenue, Canoga Park, CA 91304-0445. P. V. Kumar is with the Communication Sciences Institute, EE - Systems, University of Southern California, Los Angeles, CA 90089-2565.

This work was supported in part by the National Science Foundation under Grant NCR-9016077 and by Hughes Aircraft Company under its Ph.D. fellowship program.

Normal and Abnormal Codes

TUVI ETZION* GADI GREENBERG† IIRO S. HONKALA‡

Lot of research in the area of covering radius is on the normality of codes. The main reason is that by using the amalgamated direct sum [1] construction one can generate from normal codes sparse covering codes with larger covering radius. An $(n,d)R$ code C is a code of length n, covering radius at most R, and minimum distance at least d. An interesting question in this context is to determine which codes are normal and which codes are abnormal. One important factor is the ratio between the covering radius of the code and its minimum distance. van Wee [5] proved that all $(n,2R)R$ codes and all $(n,2R+1)R$ codes are normal. Hou [3] proved that all linear quasi-perfect codes are normal. These results are strengthen with the following theorem.

Theorem 1. If C is an $(n,2R-1)R$ code, where R does not divide n, then C is normal and all its coordinates are acceptable.

All the abnormal codes which are known [2],[4],[5] have minimum distance 1. Three constructions (A, B, and C) for generation of abnormal codes are given. The constructions that we use are modifications of the constructions of Frankl [4] and van Wee [5]. The constructions differ in the structure of the codes which they use.

By applications of Construction A we show that for most lengths there exists abnormal $(n, R)R$ codes, $R <= 6$.

Theorem 2. By applying Construction B on an $(n,d)R$ code we obtain an abnormal $(n,d-1)R+1$ code.

Theorem 3. For each $R >= 1$, there exists an n_0 such that for each $n >= n_0$ there exists an $(n, M, R-1)R$ code.

*Computer Science Department, Technion — Israel Institute of Technology, Haifa 32000, Israel.

†Mathematics Department, Technion — Israel Institute of Technology, Haifa 32000, Israel.

‡Department of Mathematics, University of Turku, 20500 Turku, Finland.

By Theorem 2 and the results of Vleduts and Skorobogatov [6] we have

Theorem 4. For each t there exists an m_0 such that for all $m >= m_0$ there exist abnormal $(2^m - 1, 2t)2t$ and $(2^m, 2t+1)2t+1$ codes.

By applying Construction C on the extended Hamming code, the punctured Preparata code, and the Preparata code, we obtain that there exists an n_0 such that for each $n >= n_0$, there exist abnormal $(2^n, 3)2$, $(2^{2n} - 1, 4)3$, and $(2^{2n}, 5)4$ codes, respectively. One consequence is that it would be difficult to extend Theorem 1.

References

[1] G. D. COHEN, A. C. LOBSTEIN, AND N. J. A. SLOANE, *Further results on the covering radius of codes, IEEE Trans. on Inform. Theory*, IT-32 (1986) 680–694.

[2] I. S. HONKALA AND H. O. HAMALAINEN, *Bounds for abnormal binary codes with covering radius 1, IEEE Trans. on Inform. Theory*, IT-37, (1991) 372–375.

[3] X. HOU, *Binary linear quasi-perfect codes are normal, IEEE Trans. on Inform. Theory*, IT-37, (1991) 378–379.

[4] K. E. KILBY AND N. J. A. SLOANE, *On the covering radius problem dor code II. Codes of low dimension; normal and abnormal codes, SIAM J. Algebraic Discrete Methods*, 8 (1987) 619–627.

[5] G. J. M. VAN WEE, *More binary covering codes are normal, IEEE Trans. on Inform. Theory*, IT-36, (1990) 1466–1470.

[6] S. G. VLEDUTS AND A. N. SKOROBOGATOV, *Covering radius for long BCH codes, Problemy Peredachi Informatsii*, 25, (1989) 38–45.

TCH: A NEW FAMILY OF CYCLIC CODES LENGTH 2^m

(†) F. A. B. CERCAS * (‡) M. TOMLINSON (†) A. A. ALBUQUERQUE

(†) Instituto Superior Técnico, DEEC, Av. Rovisco Pais, 1096 Lisboa Codex, PORTUGAL

(‡) University of Plymouth, Satellite Centre, Plymouth PL4 8AA, ENGLAND

SUMMARY

A new class of block codes length 2^m, named TCH (Tomlinson, Cercas and Hughes), has been found based on finite field theory. Sophisticated computer techniques have been used to refine and extend these codes to other binary number lengths not directly achievable, as the number of code lengths lying in the range of most practical applications is extremely scarce.

TCH codes have the advantage of an easy implementation of the receiver and are suitable for a wide range of applications in communications particularly those taking place in adverse environments like fading, Doppler effects, reflections and all types of interference. For this reason we looked for codes which could be at least cyclic and with code length $n = 2^m$, m being a positive integer. This allows the implementation of a maximum-likelihood decoder with a bank of correlators using transform techniques, such as the Fast Fourier Transform (FFT), while keeping the total number of correlators as low as possible.

TCH codes are nonlinear cyclic codes of length $n = 2^m$ as the linear addition modulo 2 of two codewords does not necessarily produce another valid codeword. They are cyclic in the sense that every cyclic shift of any codeword is always a valid codeword. A TCH code is then a block code closed under cyclic shifting but with the all-zero codeword excluded. TCH codes can be defined in terms of h code polynomials, $P_i(x)_{i=1\ to\ h}$, where $P_i(x) \neq P_j(x^r)\ mod\ n$, $i \neq j$, for all time shifts r. TCH codes are also non-systematic and cannot be defined in terms of a set of parity check equations, except in special cases. The number of information bits k of a TCH(n,k,t) code, able to correct t errors, or simply TCH(n,k), is given by :

$$k = m + log_2 h + 1 \qquad (1)$$

where the term 1 accounts for including the inverses of all codewords, which are also valid codewords.

TCH codes can be generated in the following way : once we want cyclic codes length $n = 2^m$, we must find polynomials $P(x)$ length n with coefficients a_i, $i = 0, 1, \ldots, n-1$ from GF(2). Finite field theory tells us that polynomials of degree n with coefficients from a Galois field GF(q), where q is related to a prime number p by $q = p^k$, k a positive integer, can be the field elements of GF(q). Restricting the coefficients to GF(2), as required, and for $k = 1$, we can easily construct basic TCH polynomials for all prime numbers p verifying the following equation :

$$p = n + 1 = 2^m + 1 \qquad (2)$$

Basic TCH polynomials have the form :

$$P_1(x) = \sum_{i=0}^{(\frac{n}{2}-1)-1} a_i x^{K_i} \qquad (3)$$

where the number of terms is $n/2$ so the number of ones equals the number of zeros in the polynomial. The exponent values K_i are those which verify the equation

$$\alpha^{K_i} = 1 + \alpha^{2i+1} \qquad i = 0, 1, \ldots, \frac{p-1}{2} - 1 \qquad (4)$$

for any given primitive root α of GF(q). Each of the existing α generates a different $P_1(x)$, containing $2n$ codewords, which is the basis, $h = 1$, for a new TCH code.

A good method to expand the codeword set is to use the first polynomial in a shift and add procedure which consists of cyclically shifting $P_1(x)$ and adding the shifted polynomial with the original one, in order to get a second polynomial $P_2(x)$. If $P_2(x)$ has a good auto-correlation function and a good cross-correlation function with $P_1(x)$, for all time shifts, then $P_2(x)$ is included in the codeword set, $h = 2$, therefore doubling the total number of codewords. The procedure is continued in the same way in order to increase the number of information bits k for a given value of minimum distance d_{min} required. The good results obtained with this method rely on the structure of TCH codes itself. The j^{th} cross-correlation coefficient between $P_1(x)$ and $P_1'(x) = P_1(x)[1 + x^r]$, where r is a time shift, is given by :

$$C_j = n - 2W[P_1(x)(1 + x^j + x^{r+j})] \qquad (5)$$

where $W[.]$ is the Hamming weight. C_j is virtually zero for $j = -r$. For other time shifts the coefficients are evaluated and tested.

Although the ratio k/n of TCH codes found so far is relatively small, it is shown[1, 2] that low-rate coding using TCH codes can have significant advantages. The fact that TCH codes have length 2^m and not $2^m - 1$ like BCH codes, dramatically reduces the total number of transforms through a decoding process with just two steps : in the first step the modulus of h transforms is evaluated, choosing the code polynomial which best matches, and in the second step the computation of its 2^m phases is performed so to decide what was the most likely codeword sent. The speed gain S_g defined as the ratio between the total number of operations needed to perform maximum likelihood decoding, and the number of operations needed by a TCH decoder is given by :

$$S_g = \frac{h2^m}{h + 2^m} = \frac{h}{1 + \frac{h}{n}} = \frac{n2^k}{2n^2 + 2^k} \qquad (6)$$

For the TCH(512,16,111) codes found so far the decoding process can be speeded up by more than 100 times ($S_g = 102.4$), and just a bit less ($S_g = 83.3$) for the TCH(256,16,54) codes[2].

The use of TCH polynomial codewords as PN sequences can also be advantageous as there is very little spectral overlay between them[2], or in other words, low cross-correlation. It can be shown that the average signal-to-interference ratio for the mentioned TCH(256,16) code is 24.3 dB, which is identical to the best code of this approximate length, the (255,16) Kasami code.

References

[1] M. Tomlinson, F. A. B. Cercas, C. D. Hughes. **"Aspects of Coding for Power Efficient Satellite VSAT Systems"**, ESA Journal 1991, Vol. 15, pp. 165-185.

[2] F. A. B. Cercas, M. Tomlinson, A. A. Albuquerque, **"New Developments and Applications Using TCH Codes"**, VII Symposium National de la Union Científica Internacional de Rádio, 23-25 September 1992, Málaga, Spain.

*Supported by a grant from JNICT.

Digital Signature Schemes Based on Error-Correcting Codes

Mohssen Alabbadi and Stephen B. Wicker
School of Electrical Engineering, Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract

We examine the security of several digital signature schemes based on algebraic block codes. It is shown that Xinmei's digital signature scheme can be totally broken by a known plaintext attack with complexity $O(k^3)$, where k is the dimension of the code used in the scheme. Harn and Wang have proposed a modified version of Xinme. - scheme that prevents selective forgeries. Their scheme is also shown to be vulnerable to a known plaintext attack. We then present a new signature scheme that we believe to be resistant to the previously described attacks.

1 Xinmei's Digital Signature Scheme

Xinmei's digital signature scheme [1] attempts to base its security on the intractability of the general decoding problem and the difficulty of factoring large matrices. Each user, say user A, chooses an (n,k) binary Goppa code C_A that has the ability to correct t_A errors. A $k \times n$ binary generator matrix G_A and an $(n-k) \times n$ binary parity check matrix H_A are selected for C_A. User A then finds the $n \times k$ binary matrix G_A^* such that $G_A G_A^* = I_k$, where I_k is the $k \times k$ identity matrix. User A selects a nonsingular binary $n \times n$ matrix P_A and a nonsingular binary $k \times k$ matrix S_A. User A completes the set-up of the system by constructing the matrices $J_A = P_A^{-1} G_A^* S_A^{-1}$, $W_A = G_A^* S_A^{-1}$, and $T_A = P_A^{-1} H_A^T$.

The public key consists of J_A, W_A, T_A, H_A, t_A, and t', where t' is an integer such that $t' < t_A$. The private key consists of the two matrices $S_A G_A$ and P_A.

User A obtains the n-bit signature $\underline{c}_j$ of the k-bit message $\underline{m}_j$ by computing $\underline{c}_j = (\underline{e}_j \oplus \underline{m}_j S_A G_A) P_A$, where $\underline{e}_j$ is an n-bit error vector with Hamming weight $w_H(\underline{e}_j) = t'$ chosen at random by user A.

The receiver validates the possibily noise corrupted signature $\underline{c}'_j$ through the use of the Berlekamp-Massey algorithm and the public key [1].

2 Cryptanalysis and a Modification of Xinmei's Scheme

In [2] the authors showed that the linearity of the code and knowledge of the error vectors could be exploited in a chosen-plaintext attack that results in a complete break of Xinmei's scheme. The attack transforms the cryptanalytic problem into a pair of systems of linear equations, one containing n equations in n variables, and the other containing k equations in k variables. The complexity of the attack is thus $O(n^3)$.

It was also observed by Harn and Wang in [3] that the combination of valid signatures of some messages yields a valid signature for another message. Harn and Wang [3] proposed a modification of Xinmei's scheme that appears to secure it against selective forgery. Their scheme requires that user A publish the same public keys as in Xinmei's scheme, with the further restriction that P_A b- a permutation matrix. In addition, they introduced a one-way hashing function h that is made public. The hashing function accepts an l-bit vector and produces a k-bit vector, where $l \geq k$, thus implementing a form of compression.

The n-bit signature $\underline{c}_j$ of the l-bit message $\underline{m}_j$ is obtained by computing $\underline{c}_j = h(\underline{m}_j) S_A G_A P_A$. When the signature $\underline{c}_j$ is transmitted, it becomes susceptible to errors induced by additive channel noise $\underline{e}_j$. The received signature is thus denoted by $\underline{c}'_j$ where $\underline{c}'_j = \underline{e}_j \oplus h(\underline{m}_j) S_A G_A P_A$. The signature is verified in a manner similar to that in Xinmei's original scheme.

In [4] it is shown that Harn and Wang's scheme is susceptible to a known-plaintext attack. Since the error vectors are revealed during the verification process, we can obtain the expression $\underline{c}'_j \oplus \underline{e}_j = h(\underline{m}_j) S_A G_A P_A$. Let $\underline{m}_1, \underline{m}_2, \ldots, \underline{m}_k$ be k distinct messages, $h(\underline{m}_1), h(\underline{m}_2), \ldots, h(\underline{m}_k)$ their respective images under the function h, and $\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k$ the corresponding signatures. A linear system of equations is then created: $[\underline{c}'_j \oplus \underline{e}_j] = [h(\underline{m}_j)] S_A G_A P_A$.

If the vectors $[h(\underline{m}_j)]$ are linearly independent, $S_A G_A P_A$ can be obtained in $O(k^3)$ operations.

3 A New Digital Signature Scheme

A system is proposed that uses a series of intentional error vectors that are in the same coset as the maximum likelihood error pattern, but have higher weight. These error vectors thus cannot be obtained through standard decoding techniques, making the proposed system immune to the above attacks.

A function $f(\underline{x}, \underline{y})$ is made available to all users. f is a nonlinear invertible function where $\underline{x}$ is a binary k-tuple, $\underline{y}$ is a binary n-tuple, and the output value is a binary k-tuple.

Each user, say User A, selects an (n,k) binary irreducible Goppa code C_A that has the ability to correct some t_A errors. User A then selects a generator matrix G_A and a parity check matrix H_A, and finds an $n \times k$ binary matrix G_A^* such that $G_A G_A^* = I_k$. A nonsingular binary $n \times n$ matrix P_A is generated and the matrices $G'_A = P_A^{-1} G_A^*$ and $H'_A = P_A^{-1} H_A^T$ computed.

Finally User A selects an $n \times l$ binary matrix W_A of rank n, where $n < l$, and determines W_A^* such that $W_A W_A^* = I_n$.

The public key consis's of G'_A, H'_A, W_A^*, t_A, and t'_A, where t'_A is an integer such that $t' < t_A$. The private key consists of the matrices G_A, P_A, G_A^*, and W_A.

A k-bit message $\underline{m}_j$ is signed in the following manner. A random binary error vector $\underline{z}_j$ of length n and weight t_A is selected. A random l-bit vector $\underline{e}_j$ of arbitrary weight is also selected. The l-bit signature $\underline{s}_j$ is then computed using the following expression.

$$\underline{s}_j = \{(\underline{z}_j \oplus [f(\underline{m}_j, \underline{z}_j) \oplus \underline{z}_j G_A^*] G_A) P_A \oplus \underline{e}_j W_A^*\} W_A \oplus \underline{e}_j. \qquad (1)$$

The signature is validated by first computing $\underline{v}_j = \underline{s}_j W_A^*$. The Berlekamp-Massey algorithm is then applied to $\underline{v}_j$ to obtain an estimate of $\underline{z}_j$. The remainder of the public key is used to obtain $f(\underline{m}_j, \underline{z}_j)$, which is then compared to the value computed by the receiving user.

References

[1] W. Xinmei. Digital signature scheme based on error-correcting codes. *Electronics Letters*, 26(13):898–899, 21st June 1990.

[2] M. Alabbadi and S. Wicker. Comments on the security of Xinmei's digital signature scheme. *Electronics Letters*, 28(9):890–891, 23rd April 1992.

[3] L. Harn and D. -C. Wang. Cryptanalysis and modification of digital signature scheme based on error-correcting codes. *Electronics Letters*, 28(2):157–159, 16th January 1992.

[4] M. Alabbadi and S. Wicker. Cryptanalysis of the Harn and Wang modification of the Xinmei digital signature scheme. *Electronics Letters*, 28(18):1756–1757, 27th August 1992.

The Trellis Complexity of Equivalent Binary [17, 9] Quadratic Residue Codes Is Five

Yan-Yih Wang *Chung-Chin Lu*
Department of Electrical Engineering
National Tsing Hua University
Hsinchu, Taiwan 300, R.O.C.

Abstract: It is known that equivalent linear block codes may have different minimal trellis structures. The minimum complexity among all minimal trellis structures of equivalent codes is defined as the trellis complexity of the class of equivalent codes. Sharper lower bounds for trellis complexity are derived when more information about the infrastructure of codes is supplied. These bounds serve as a starting specification for a search algorithm to find optimal permutations under which the permuted codes achieve the trellis complexity. A simple application to the class of equivalent binary [17,9] quadratic residue codes finds the trellis complexity is five.

Let $\mathcal{C}$ be an $[n, k, d]$ linear block code over $GF(q)$. Let $\mathcal{D}$ be its dual code with minimum distance $d^\perp$. Let S_n be the set of all permutations on the n coordinates of codewords. Let $\sigma(\mathcal{C})$ be the equivalent code of $\mathcal{C}$ under a permutation σ in S_n. Let $\sigma(\mathcal{C})_{p,i}$ ($\sigma(\mathcal{C})_{f,i}$) be the past (future) subcode of $\sigma(\mathcal{C})$ which consists of codewords whose future (past) coordinates to position i are all zero. Let $k_{p,i}(\sigma)$ ($k_{f,i}(\sigma)$) be the dimension of the past (future) subcode $\sigma(\mathcal{C})_{p,i}$ ($\sigma(\mathcal{C})_{f,i}$). The dimension $k_{s,i}(\sigma)$ of the state space at position i in a minimal trellis of $\sigma(\mathcal{C})$ is [1]

$$k_{s,i}(\sigma) = k - k_{p,i}(\sigma) - k_{f,i}(\sigma).$$

Let $s(\sigma(\mathcal{C}))$ be the maximum value of $k_{s,i}$ over $0 \le i \le n$. The trellis complexity s of the class of equivalent codes of $\mathcal{C}$ is defined as

$$s = \min_{\sigma \in S_n} s(\sigma(\mathcal{C})).$$

Let

$$K_{p,i} = \max_{\sigma \in S_n} k_{p,i}(\sigma), K_{f,i} = \max_{\sigma \in S_n} k_{f,i}(\sigma), K_{s,i} = k - K_{p,i} - K_{f,i}.$$

Note that $K_{p,i} = K_{f,n-i}$. Since $k_{p,i}(\sigma) \le K_{p,i}$ and $k_{f,i}(\sigma) \le K_{f,i}$, we have

$$k_{s,i}(\sigma) \ge K_{s,i}.$$

In general, $K_{p,i}$ and $K_{f,i}$ are intrinsic attributes of the class of equivalent codes of code $\mathcal{C}$. $K_{f,i}$ may be estimated by $N(\alpha, \beta)$ [2] which is the minimum possible block length for a linear block code to have minimum distance α and dimension β as follows:

1. If $i \le N(d^\perp, j) - 1$, then $K_{f,i} \le k - i + j - 1$.
2. If $i \ge n - N(d, j) + 1$, then $K_{f,i} \le j - 1$.

More precisely, for binary codes and early and late positions i, $K_{f,i}$ can be evaluated as follows:

$$K_{f,i} = \begin{cases} k - i, & \text{if } 0 \le i \le d^\perp - 1, \\ k - i + 1, & \text{if } d^\perp \le i < d^\perp + \left\lceil \frac{d^\perp}{2} \right\rceil - 1, \\ 1, & \text{if } n - \left(d + \left\lceil \frac{d}{2} \right\rceil - 1\right) \le i \le n - d, \\ 0, & \text{if } n - d + 1 \le i \le n. \end{cases}$$

Two monotone sequences $0 = \tilde{k}_{p,0} \le \tilde{k}_{p,1} \le \ldots \le \tilde{k}_{p,n} = k$ and $k = \tilde{k}_{f,0} \ge \tilde{k}_{f,1} \ge \ldots \ge \tilde{k}_{f,n} = 0$ together are called a specification of past and future dimensions if they satisfy

1. $0 \le \tilde{k}_{p,i} - \tilde{k}_{p,i-1}(\tilde{k}_{f,i-1} - \tilde{k}_{f,i}) \le 1, \forall 1 \le i \le n$;
2. $k - \tilde{k}_{p,i} - \tilde{k}_{f,i} \ge 0$ for all $0 \le i \le n$.

The minimal trellis structure of an equivalent code $\sigma(\mathcal{C})$ is said to be dominated by a specification as in above if all its past dimensions $k_{p,i}(\sigma)$ and future dimensions $k_{f,i}(\sigma)$ are upper bounded by $\tilde{k}_{p,i}$ and $\tilde{k}_{f,i}$ respectively at each position i. Necessary and sufficient conditions for the existence of a permutation σ under which the minimal trellis structure of the permuted code is close to and dominated by a specification are developed. And a constructive algorithm is then built to search for optimal permutations under which permuted codes can achieve the trellis complexity.

Let $\mathcal{C}$ be the binary [17,9] quadratic residue code generated by $g(x) = 1 + x^3 + x^4 + x^5 + x^8$.. Let $\mathcal{D}$ be its dual code, The minimum distance of $\mathcal{C}$ and $\mathcal{D}$ are $d = 5$ and $d^\perp = 6$. By applying the above results, we can list $K_{f,i}, K_{p,i}$, and $K_{s,i}$ in the following table:

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$K_{p,i}$	0	0	0	0	0	1	1	1	2	2	3	4	4	5	6	7	8	9
$K_{f,i}$	9	8	7	6	5	4	4	3	2	2	1	1	1	0	0	0	0	0
$K_{s,i}$	0	1	2	3	4	4	4	5	5	5	5	4	4	4	3	2	1	0

Hence, the trellis complexity of the class of equivalent binary [17,9] QR codes is not smaller than 5. To find optimal permutations and then to determine the exact trellis complexity, we start our search algorithm with the following specification of future and past dimensions $\tilde{k}_{p,i}$, $\tilde{k}_{f,i}$, a very slight variation from the above table, listed in the next table:

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\tilde{k}_{p,i}$	0	0	0	0	0	0	1	1	2	2	3	3	4	5	6	7	8	9
$\tilde{k}_{f,i}$	9	8	7	6	5	4	3	3	2	2	1	1	0	0	0	0	0	0
$\tilde{k}_{s,i}$	0	1	2	3	4	5	5	5	5	5	5	5	5	4	3	2	1	0

With the above specification, we have constructed four optimal permutations:

$$\begin{aligned} \sigma &= (1,4,5,6,9,7,10,2,14,17,3,15,8,11,12,13,16) \\ &\text{or } (1,4,5,6,9,10,7,2,14,17,3,15,8,11,12,13,16) \\ &\text{or } (1,4,5,6,9,7,10,2,14,17,15,3,8,11,12,13,16) \\ &\text{or } (1,4,5,6,9,10,7,2,14,17,15,3,8,11,12,13,16). \end{aligned}$$

With anyone of the above permutations, we have

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$k_{p,i}(\sigma)$	0	0	0	0	0	1	1	1	2	2	3	3	4	5	6	7	8	9
$k_{f,i}(\sigma)$	9	8	7	6	5	4	3	3	2	2	1	1	1	0	0	0	0	0
$k_{s,i}(\sigma)$	0	1	2	3	4	4	5	5	5	5	5	5	4	4	3	2	1	0

Thus, the trellis complexity of the class of equivalent binary [17, 9] QR codes is 5.

[1] G. D. Forney, JR., "Coset codes - Part II: Binary lattices and related codes," *IEEE Trans. on Inform. Theory*, vol. IT-34, no. 5, pp. 1152-1187, Sept. 1988. Part II.

[2] H. C. A. van Tilborg, "The smallest length of binary 7-dimensional linear codes with prescribed minimum distance," *Discrete Mathematics*, **33** (1981), pp. 197-207.

CHANNEL EQUALIZATION FOR BLOCK TRANSMISSION SYSTEMS

Ghassan Kawas Kaleh

Ecole Nationale Supérieure des Télécommunications, 46, rue Barrault, 75 634 Paris 13, France

In Block Transmission System the information symbols are arranged in the form of blocks alternating with blocks of L known symbols, L being the length of channel memory*. The latters help to identify the channel and to process each information block independently from the others; their interferences on the sampled output of the matched filter are calculated and then subtracted . We thus obtain the observation vector $Y = R\,D + U$, where R is an $M \times M$ Toeplitz Hermitian symmetric matrix which represents channel distorsion, $D = (d_M, d_{M-1}, \ldots, d_1)^t$ is the information symbol vector and U is a Gaussian zero-mean noise vector whose covariance matrix is $2N_0$ R. The known receiver for this system is the **Nonlinear Data-Directed Estimator** (NDDE)*. Its complexity is $O(M^3)$; it has to solve, using the Levinson Algorithm, $M/2$ Toeplitz systems of decreasing order.

In this paper we extend the equalization techniques to Block Transmission Systems and deduce decision-feedback equalizers that have better performance and less complexity than the NDDE. The fact that the observation and its noise are snapshots from stationary time-series suggests the use of a nonstationary innovations representation based on Cholesky factorization of the matrix R as $R = H^{*t}\,\Sigma^2\,H$, where H is an upper triangular matrix with 1's along the main diagonal and Σ is diagonal with positive real entries $\Sigma_{ii} = \sigma_i$; i=0, 1, ...M-1. The factors H and Σ are obtained from the Schur algorithm. We use this representation to deduce the following processors.

1. Noise Whitener ($\Sigma^{-1}H^{*t-1}$). It transforms the observation Y into $Z = \Sigma\,H\,D + V$. The covariance matrix of the noise V is $2N_0$ I. The vector Z is obtained from Y without inverting H^{*t}, thanks to its triangular structure.

2. Maximum Likelihood Block Detector. It may use the Viterbi algorithm. The main difference with the conventional maximum likelihood sequence detector is that the channel seen by the detector is time-varying.

3. Zero-Forcing Block Linear Equalizer (ZF-BLE) It gives $X = (\Sigma\,H)^{-1}Z = R^{-1}\,Y = D + W$, also without inverting H. Suboptimum symbol-by-symbol decisions are obtained from X via a threshold detector. The signal-to-noise ratio in the decision variable on symbol d_i, is $SNR_i = E_s/(N_0\sigma_0^2\,[R^{-1}]_{M-i,M-i})$, where E_s is the symbol energy.

4. Zero-Forcing Block Decision-Feedback Equalizer (ZF-BDFE). We obtain from the noise whitener $\Sigma^{-1}\,Z = H\,D + \Sigma^{-1}\,V$. The transformation H is causal and triangular. Thus, starting with a decision on d_1, the decision on symbol d_i can be obtained with the help of decisions on previous symbols, as made in the conventional DFE. We have $SNR_i = \sigma_{M-i}^2\,E_s/(N_0\sigma_0^2)$. We show that this performance is better than that of NDDE, ZF-BLE, and the conventional ZF-DFE. The complexity is $O(LM)$.

5. Minimum-Mean-Squared-Error Block Linear Equalizer (MMSE-BLE). The performance degradation of the ZF-BLE can be reduced by inserting between R^{-1} and the threshold detector a Wiener estimator Ψ to obtain $X' = \Psi\,X = D + W'$, where the power of every components w'_i of W is minimized. The cascade of R^{-1} and Ψ is a transformation $R'^{-1} = \Psi\,R^{-1}$ whose input is Y and output is X', where $R' = [\,R + (2N_0/E|d_k|^2)\,I\,]$. The covariance of the error W' is $2N_0\,R'^{-1}$. The SNR_i is $[E|d_i|^2/\,E|w'_i|^2\,] - 1$.

6. Minimum-Mean-Squared-Error Block Decision-Feedback Equalizer (MMSE-BDFE). The observation can be written as $Y = R'D + U'$, where the error U' has a covariance matrix $2N_0R'$. The matrix R' can be Cholesky-factored as $R' = H'^{*t}\,\Sigma'^2\,H'$. As above, a noise whitener ($\Sigma'^{-1}H'^{*t-1}$) can be used and its output is processed using decision feedback strategy as in ZF-BDFE. The decision vector is $S = D + \Delta$, where the error Δ is a mixture of noise and residual intersymbol interference. Its covariance is $2N_0\,\Sigma'^{-2}$. The SNR_i is $[E|d_i|^2/\,E|\delta_i|^2] - 1$, where δ_i is component of Δ. This performance is better than the that of NDDE, ZF-BDFE, and the conventional MMSE-DFE. The complexity is $O(LM)$.

Conclusions: Whereas conventional equalizers use transversal filters, the derived ones involve the use of matrix transformations, as expected. These transformations can be implemented exactly, while what is deduced from the theory of conventional equalization is approximated by simple implementable filters. Moreover, assuming the channel impulse response (or its estimate) is available, the equalizer coefficients (the matrix entries) are easily calculated using the Levinson or Schur algorithms. We use the latter because it implies complexity reduction and allows us to use a decision-feedback strategy. We have evaluated the performances of the deduced equalizers and compared them with that of known ones. The ZF and MMSE block decision feedback equalizers are particularly attractive because of their better performance and lower complexity as compared with the known NDDE.

---

* F. Hsu, "Data Directed Estimation Techniques for Single-Tone HF Modems," IEEE Military Commun. Conf., Boston, MA., Oct. 1985.

UPPER BOUNDING THE PERFORMANCE OF ISI CHANNELS*

Sreenivasa A. Raghavan
ComStream Corporation
10180 Barnes Canyon Road
San Diego, CA 92121

Jack K. Wolf
Center for Magnetic Recording Research
University of California, San Diego
La Jolla, CA 92093

Laurence B. Milstein
Dept. of Electrical & Computer Engineering
University of California, San Diego
La Jolla, CA 92093

Summary

We consider a communication channel that is corrupted by both finite ISI and additive white Gaussian noise. The impulse response of the channel is $h(t)$ and we assume BPSK modulation. We assume that $h(t)$ is time-limited to nT, where $1/T$ is the rate at which data is transmitted on the channel. Hence the output of the channel is given by

$$r(t) = \sum_k (2a_k - 1)h(t - kT) + n(t), \quad (1)$$

where $n(t)$ is additive white Gaussian noise with two sided power spectral density $N_0/2$ and a_k is equally likely to be 0 or 1. The receiver filter is $g(t)$, the whitened matched filter corresponding to $h(t)$. The noise samples at the output of the sampler are uncorrelated Gaussian random variables with zero mean and variance equal to $N_0/2$. If we denote the sampled outputs of $g(t)$ by $\{y_k\}$, then

$$y_k = v_k + n_k,$$

where

$$v_k = \sum_{i=0}^{n-1} f_i(2a_{k-i} - 1)$$

and

$$E[n_k n_m] = \frac{N_0}{2}\delta_{km}. \quad (2)$$

In (2), the symbol $E[\cdot]$ stands for expectation and

$$\delta_{km} = \begin{cases} 1 & \text{if } k = m \\ 0 & \text{else} \end{cases}$$

The set of constants $\{f_k\}$ depends on the pulse autocorrelation function of the impulse response.

Another way to represent the same ISI channel is as a trellis code. It is equivalent to a rate $1/n$ binary input, linear convolutional code, followed by a nonlinear mapping. Each input to the channel a_k results in an n-bit codeword at the output of the encoder given by

$$\mathbf{C}^k = (a_k, a_{k-1}, a_{k-2}, \cdots, a_{k-n+1}). \quad (3)$$

The nonlinear mapping in our case (i.e., for the ISI channel) is given by

$$M(\mathbf{C}^k) = \sum_{i=0}^{n-1} f_i(2a_{k-i} - 1). \quad (4)$$

Therefore, using this notation,

$$y_k = M(\mathbf{C}^k) + n_k. \quad (5)$$

Consider the following definitions:

Definition 1: The polynominal, $f(D)$, that characterizes the ISI channel is given by

$$f(D) = \sum_{i=0}^{n-1} f_i D^i. \quad (6)$$

Definition 2: Let $\mathbf{E}^k$ be a binary n-tuple given by $(e_0^k, e_1^k, \cdots, e_{n-1}^k)$. Then the squared Euclidean error weight of $\mathbf{C}^k$ with respect to $\mathbf{E}^k$ is given by

$$d^2(\mathbf{C};\mathbf{E}) = \|M(\mathbf{C}) - M(\mathbf{C} \oplus \mathbf{E})\|^2, \quad (7)$$

where the symbol $\oplus$ stands for bit-by-bit modulo-2 (logical XOR) operation of two vectors of length n. Since there are at most 2^n channel signals, there can be at most $2^{n-1}(2^n - 1) + 1$ possible values for the Euclidean weight.

Definition 3: Let A be a set of binary n-tuple vectors $\mathbf{C}_i$ (or, equivalently, a set of channel signals). The weight profile of the set A with respect to a given error vector $\mathbf{E}^k$, denoted $F(A, \mathbf{E}^k, W)$, is given by

$$F(A, \mathbf{E}^k, W) = \sum_\alpha m_\alpha W^\alpha, \quad (8)$$

where m_α is the number of elements in the set A that have a squared Euclidean error weight α with respect to $\mathbf{E}^k$.

It is now straightforward to prove the following lemma:

Lemma: Let A be the set of all binary n-tuples. Then the subset A_c of A defined by

$$A_c = \left\{ \mathbf{C} \mid \mathbf{C} = (\underbrace{x\,x\,\ldots\ldots x}_{n-1 \text{ arbitrary binary numbers}}\ 0) \right\}$$

forms a sub-group of A, under the operation $\oplus$ defined in (7).

The set A_c has cardinality 2^{n-1}. The only coset of A_c, denoted $\hat{A}_c$, may be formed from A_c as

$$\hat{A}_c = \{\mathbf{C}' \mid \mathbf{C}' = \oplus(\underbrace{1\,1\,1\,\ldots\ldots 1}_{\text{length-}n \text{ vector of all ones}}), \text{ where } \mathbf{C} \in A_c\}.$$

This follows from the fact that the all-one vector is not in A_c. Hence adding this vector to all elements of A_c forms the only coset of A_c. $\hat{A}_c$ also has cardinality 2^{n-1}.

Theorem: For any error pattern $\mathbf{E}$ of length n and for any partial response channel $f(D)$ with $(n-1)$ interfering symbols,

$$F(A_c, \mathbf{E}, W) = F(\hat{A}_c, \mathbf{E}, W)$$

The proof is presented in [1].

For the ISI channel with $n = 2$, the subset A_c corresponds to all possible outputs from the all zero state and $\hat{A}_c$ is the only coset of A_c. Hence this channel satisfies the conditions imposed in [2] for a class of trellis codes, namely a) the trellis is based upon a binary, linear convolutional code of rate $(n-1)/n$ with a nonlinear mapping from the encoder output to channel input symbols, and b) $F(A_c, \mathbf{E}, W) = F(\hat{A}_c, \mathbf{E}, W)$. Thus, we can apply a modified generating function of the ISI channel with one interfering symbol that involves only 2 states in the state diagram. This modified generating function can then be used to compute the probabilities of both event errors and bit errors. In the computation of the generating function, we may assume that the initial state is the all zero state, without loss of any generality. The edge labels are, however $LF(A_c, \mathbf{E}, W)I^r$, where r is the number of data bit errors. If we denote the resulting generating function $T(W, L, I)$, then the probability of event error P_e and the probability of bit error P_b are upper-bounded by

$$P_e \le Q\left(\sqrt{\frac{d_f^2 E_s}{2N_0}}\right) \exp\left(\frac{d_f^2 E_s}{4N_0}\right) T\left(W = \exp\left(\frac{-E_s}{4N_0}\right), L = \frac{1}{2}, I = 1\right),$$

and

$$P_b \le Q\left(\sqrt{\frac{d_f^2 E_s}{2N_0}}\right) \exp\left(\frac{d_f^2 E_s}{4N_0}\right) \frac{\partial T}{\partial I}\left(W = \exp\left(\frac{-E_s}{4N_0}\right), L = \frac{1}{2}, I = 1\right),$$

respectively, where d_f^2 is the normalized minimum squared free Euclidean distance of the ISI channel and E_s is the average channel symbol energy.

References

[1] S. A. Raghavan, J. K. Wolf, and L. B. Milstein, "On the Performance Evaluation of ISI Channel". Accepted in *IEEE Trans. on Inform. Theory*.

[2] E. Zehavi and J. K. Wolf, On the Performance Emulation of Trellis Codes," *IEEE Trans. on Inform. Theory* IT-33, 196-202, March 1987.

*This work was partially supported by the National Science Foundation under grant NCR 9105639, and the Center for Magnetic Recording Research at the University of California, San Diego

Performance of M-Algorithm Receivers With Imperfect Channel Estimates

F. Gozzo
IBM - Federal Systems Company
Owego, NY 13827-1298

J.B. Anderson
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

I. Introduction

In any practical system, the channel estimate can be inaccurate for one of many reasons including finite-length training sequences, quantization, time-varying channels, and truncation. Thus, the performance of any receiver will typically degrade under mismatched channel conditions. Although this degradation in performance was extensively studied for MLSE receivers by Divsalar [1], the *optimality* of the mismatched MLSE receiver was not addressed. Unfortunately, deriving the optimal receiver under arbitrary mismatched channel conditions is clearly an intractable problem. This paper presents test results in an effort to better understand the performance of MLSE receivers in arbitrary channel mismatch conditions.

II. The M-Algorithm

The M-algorithm, which has become increasingly popular in communications applications [2]- [7], gives the designer the ability to easily trade-off complexity and performance. It performs a breadth-first search of an ISI tree (or trellis), but only keeps the best M paths, up to a decision-depth, D. As each new signal is received, the algorithm extends the M paths, sorts them according to their cost, and retains only the best M paths. We will denote the M-algorithm receiver as MA(M,D), where M is the number of paths to keep and D is the decision depth. Assuming the decision depth is long enough, the single parameter, M, enables one to investigate a continuum of practical MLSE-based receivers — from reduced-search to full-complexity MLSE.

III. Test Results

The results from [8] shown in Figure 1 are for a channel whose transfer function is $H(z) = \sqrt{1-a^2} + az^{-1}$. A series of tests were run to determine what value of M would yield MLSE performance. We began with a complexity of $M = 1$ and proceeded to increase M until marginal performance changes were found under a broad range of mismatch conditions. Bit error rates are shown as a function of a_{train}, the value of the parameter a during training. The highest value of M shown represents the saturating value of M, i.e., increasing M further did not significantly alter the curve. Based on the test results, several observations can be made:

1. *M-Algorithm quickly converges to MLSE Performance.* Relatively few paths were required to achieve near-MLSE performance, in spite of the fact that merged paths were ignored by our implementation.

2. *Increasing M-Algorithm Complexity May Degrade Performance In Mismatch Conditions.* Analogous to results for other receivers discussed in [8], we feel confident that:

Under channel mismatch conditions, a (well-trained) full-complexity MLSE receiver will be optimally matched to the wrong channel, and may therefore achieve a deeper level of mismatch than a reduced-complexity MLSE scheme.

3. *Eigenvalue Spread Is Useful In The Analysis of Channel Mismatch.* In Figure 1, we have superimposed the eigenvalue spread of Channel C. As can be seen, the point at which the MLSE receiver (i.e., MA(5,10)) is no longer optimum coincides with the point at which the eigenvalue spread approaches infinity. In general, this *complexity-inversion* phenomenon — the situation when increasing a receiver's complexity could actually degrade its performance — was found to occur whenever the eigenvalue spread during training and decoding differed significantly.

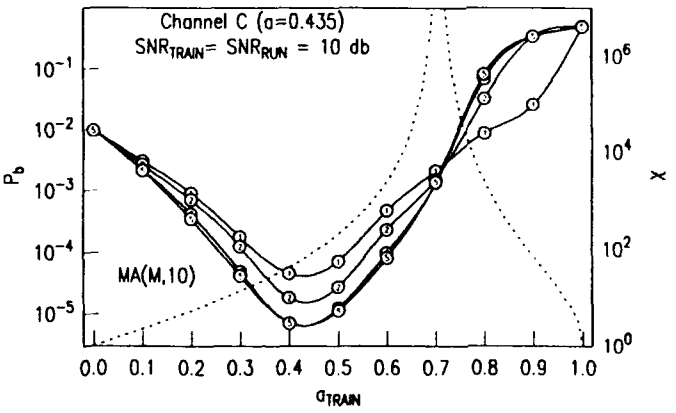


Figure 1. Performance of M-Algorithm In Channel Mismatch. Solid curves represent P_b for selected M. Dashed curve represents the eigenvalue spread, χ.

References

[1] D. Divsalar, "Performance of Mismatched Receivers on Bandlimited Channels," Ph.D. Dissertation, UCLA, 1978.

[2] J.B. Anderson and S. Mohan, "Sequential Coding Algorithms: A Survey and Cost Analysis," *IEEE Trans. Commun.*, Vol. COM-32, No. 2, pp. 169-176, Feb. 1984.

[3] J.B. Anderson and S. Mohan, *Source and Channel Coding: An Algorithmic Approach*, Boston: Kluwer, 1991.

[4] J.B. Anderson, "Limited Search Trellis Decoding of Convolutional Codes," *IEEE Trans. Inform. Theory*, Vol. 35, No. 5, pp. 944-955, Sept. 1989.

[5] D. Bauer, H. Speer, and C.M. Zhao, "Performance of Sequential Detection Schemes For Trellis-Encoded Data Signals Distorted By Intersymbol Interference," *Proc. IEE, Sixth Int. Conf on Dig. Proc. of Sig. in Comm.*, pp. 141-146, Sept. 1991.

[6] F-Q Wang and D.J. Costello, "A Hybrid M-Algorithm/Sequential Detector For Convolutional and Trellis Codes," *NASA Tech Report*, CR-186863, June 29, 1990.

[7] G. Zimmerman and W. Rupprecht, "Adaptive Receiver Structures With Sequential Detection Algorithms For Digital Mobile Radio Systems," *Proc. IEE, Sixth Int. Conf on Dig. Proc. of Sig. in Comm.*, Sept. 1991, pp. 141-146.

[8] F. Gozzo, "Robust Sequence Estimation in the Presence of Channel Mismatch," Ph.D. Dissertation, Rensselaer Poly. Inst., May. 1992

Presented at the 1993 IEEE International Symposium on Information Theory
This work was supported by the IBM Resident Study Program

New Results in Signal Design for the AWGN Channel

M. Steiner
Naval Research Laboratory

Summary

There has been a fair amount of work done in the area of signal design. Unfortunately, there are few results on the optimality of signal sets (throughout the paper an optimal signal set is one that maximizes the average probability of detection). The optimal selection of M signal vectors embedded in even the most fundamental type of noise, white Gaussian noise, is not known in general. One of the most famous of communication conjectures, dating back to 1948, states that the optimal signal vectors are vertices of an n dimensional regular simplex for which $M = n + 1$. When the signal vectors are constrained only by an average power limitation, this conjecture has been referred to as the strong simplex conjecture (SSC). To avoid confusion, we refer to the conjecture of simplex optimality when the signal vectors lie on the surface of a sphere as the weak simplex conjecture (WSC). The validity of the SSC implies the validity of the WSC, although the converse statement is not true.

Under the assumption that the signal vectors are equal energy, Balakrishnan proved in his seminal work [1] that the regular simplex is 1) optimal (in terms of maximizing the average probability of detection) as the signal to noise ratio (SNR) λ approaches infinity, 2) optimal as λ approaches zero, and 3) locally optimal at all λ. He also proved that if there *does* exist a signal set which is optimal at all λ it is necessarily the regular simplex signal set. Dunbridge[2][3] in 1967 extended Balakrishnan's work where only an average power constraint is imposed on the signal set. It was proven by both Balakrishnan [4] and Weber[5] that the regular simplex maximizes the minimum distance under a peak power constraint.

A number of new results are presented. The strong simplex conjecture is disproven. A signal set is shown which performs better than the regular simplex at low signal to noise ratios for all $M \geq 7$. This leads to a theorem which states that in general there is no signal set which is optimal at all signal to noise ratios. Furthermore it is found that the optimal solution at low signal to noise ratios is not an equal energy solution for all $M \geq 7$. The regular simplex is shown to be the unique 'ope which maximizes the minimum distance between signals. This extends the result by Balakrishnan who proved that the regular simplex maximizes the minimum distance under a peak power constraint. This result leads to the corollary that a signal set which maximizes the minimum distance is not necessarily optimum. This is an interesting result since much signal design work has been based on maximizing the minimum distance due to the inherent simplicity of the criteria. A simple proof that the regular simplex maximizes the minimum distance under a peak power constraint is also shown. The global optimality of the regular simplex under the performance measure of the union bound on the probability of detection is addressed. The union bound is often used to assess the performance of many signal sets at medium to high SNR when computation of the probability of detection is intractable. It is proven that the regular simplex uniquely maximizes the union bound at all signal to noise ratios.

References

[1] A. V. Balakrishnan, "A contribution to the sphere-packing problem of communication theory," *Journal of Mathematical Analysis and Applications*, vol. 3, pp. 485–506, 1961.

[2] B. Dunbridge, "Asymmetric signal design for the coherent Gaussian channel," *IEEE Trans. on Inform. Theory*, vol. IT-13, pp. 422–431, July 1967.

[3] B. Dunbridge, *Optimal signal design for the coherent Gaussian channel.* PhD thesis, University of Southern California, Los Angeles, December 1965.

[4] A. V. Balakrishnan, *Advances in communication systems.* New York: McGraw-Hill, 1965.

[5] C. L. Weber, *Elements of Detection and Signal Design.* Springer Verlag, 1987.

PRACTICAL USE OF IMPORTANCE SAMPLING IN DIGITAL COMMUNICATION SYSTEM SIMULATIONS

Kung Yao and Dongrin Kim
Electrical Engineering Department
University of California, Los Angeles, CA 90024-1594

Importance sampling scheme using a fixed mean translation (MT) without conditioning on the intersymbol interference (ISI) in a digital communication system is proposed. The reduction in simulation samples is significant. MT shift found by adaptive algorithm agrees with the one from numerical method based on large deviation theory for nonlinear system.

Introduction

In a digital communication system the bit error probability (P_e) is an important performance measure, but often it can not be derived analytically in closed form due to the complexity of the system. For this reason the Monte Carlo (MC) simulation method has been commonly used. Since the number of MC samples is of the order of $1/P_e$, for a small P_e, this number can be quite large. Important sampling(IS) is one variance reduction method to evaluate P_e with the same degree of of accuracy but using a smaller number of samples than the conventional MC method. We will consider a simplified satellite digital communication system model with uplink and downlink noise. The nonlinear element of the system contains the third order term for modeling the saturating amplifier.

Mean Translation Importance Sampling

In the MT IS scheme [1] the new random vector n^* is obtained by $n^* = n + c$, where c is a constant vector. Therefore, it is critical to find a proper value c in order for the IS technique to effectively decrease the variance of the sampled data. For the linear system most efficient MT shift $c_j(opt)$ can be found easily if the conditioned ISI pattern is used.

In order to find a c_j for nonlinear system we use the result of large deviation theory and an adaptive scheme. From the large deviation theory [2] we can use a numerical method which provides c_j whose behavior is more desirable than the $c_j(opt)$ of the linearized version of the system. Adaptive scheme [3] includes the process of finding c_j in the simulation program estimating the system performance. Both methods can be used to obtain identical result of $c_j(opt)$ when applied to the linear system.

Application

From analysis we can show that for a linear system the maximum MT shift which gives variance reductions as compared to the MC scheme is about twice as large as c_{opt}. This illustrates the adverse effect of a large MT shift, while a small amount of shift is innocuous. We consider unconditioned stream ISI simulation which generates both the ISI sequence vectors and noise vectors. Rather than using $c_j(opt)$ for each ISI sequence generated, a fixed shift c_{fix} is used for the MT IS. When this method is applied to our nonlinear model, the c_{fix} will be taken as the optimum c_j for the least signal output which can be evaluated.

For systems with memory of length M, unconditioned ISI is significantly more effective than conditioned ISI simulation as this scheme does not require 2^{M-1} simulations with different c's. An unconditioned ISI stream simulation is applied to a nonlinear system with M=12. Results are shown in the following figure for three different downlink to uplink noise ratios. The efficiency (i.e., the ratio of number of MC simulations to the number of unconditioned stream IS for the same variance) versus P_e is shown. As can be seen, the efficiency increases for small P_e.

Conclusion

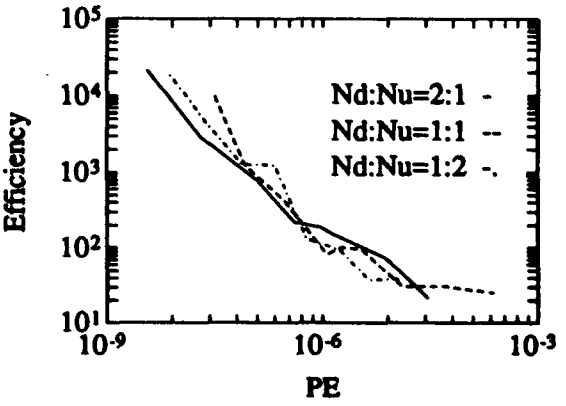
We presented a practical method to estimate P_e for a nonlinear system with memory by biasing the Gaussian noise and not conditioning on the ISI. In order to find the proper MT shift amount, numerical method was used. When the adaptive method was used assuming no knowledge of the ideal shifting value, the numerically obtained minumum rate point and the mean value of the error causing noise vector were almost identical.

References

[1] D. Lu and K. Yao,"Improved Importance Sampling Tech. for efficient Simulation of Digital Comm. Systems," *IEEE J. Sel. Ar. Comm.* Jan. 1988, pp.67-75.

[2] J.S. Sadowsky and J.A. Bucklew,"On Large Deviations Theory and Asymp. Efficient Monte Carlo Estimation," *IEEE Trans. on Infor. Theory*, No. 3, May 1990.

[3] J.S. Stadler and S. Roy,"Adaptive Importance Sampling," *To appear.*



A New Class of Optimum Importance Sampling Strategies Derived from Statistical Distance Measures

Geoffrey C. Orsak[1]
Dept. of Electrical and Comp. Eng.
George Mason University
Fairfax, VA 22030-4444

Behnaam Aazhang[2]
Dept. of Electrical and Comp. Eng.
Rice University
Houston, TX 77251-1892

Summary

Performance analysis of discrete time systems often requires the evaluation of the expected value of a cost function of the system output or equivalently the expected value of a functional of the system input vector. In either case, analytical expressions for the performance in many practical situations are typically unavailable. As an alternative to approximations or bounds, Monte Carlo simulations are often employed as a convergent method for obtaining arbitrarily accurate estimates of the performance. Unfortunately, in many important applications, the required computations can often be prohibitive. Therefore, much recent research has focused on the development of new and efficient forms of the method of Monte Carlo known as Importance Sampling simulations.

The fundamental problem in Importance Sampling is to determine the appropriate statistics for the simulation. It is well known that the minimum variance statistics can not be implemented since they require the explicit knowledge of the expectation one is attempting to estimate. Therefore, researchers have worked to determine "good" suboptimal statistics from probability measure constraint classes which exclude the degenerate optimal solution. These so-called "biasing strategies" are typically obtained by minimizing the variance of the Importance Sampling estimator with respect to the biasing statistics over a class of probability measures which exclude the optimal distribution.

Recently[3], it was shown by the authors that minimizing the Importance Sampling variance is equivalent to minimizing an Ali-Silvey distance [1] of equivalently an f-divergence [2] between the admissible biasing densities and the well known optimal biasing density. This result has led to a new approach in the design of Importance Sampling strategies where one merely determines the biasing density from an arbitrary constraint class with minimum "distance" to the global optimal distribution to minimize the simulation variance.

Extending this previous research, we derive in this work the minimum variance biasing distribution from a constraint class whose controlling parameter is fundamental in the performance analysis of Importance Sampling. In addition, we will show that for the special case of estimating the probability of rare events, the constrained optimal biasing distribution from this class is independent of the unknown parameter and as such, leads to solutions which are both amenable to implementation and yet still optimal with respect to a relevant criterion.

To motivate the proposed constraint class, we note that it has been analytically established for the very important special case of estimating the probability of a rare event that effective biasing strategies *must* generate the rare event with greater frequency than the statistics of the original Monte Carlo simulation. In fact, the probability of the rare event with respect to the Importance Sampling statistics is the controlling parameter in an achievable lower bound on the variance of the Importance Sampling simulation. Therefore, generalizing these arguments, we consider the class of all biasing distributions with probability mass over an arbitrary Borel subset bounded by some an arbitrary constant.

We eliminate the mathematical formalities and developments and simply state the major results. The optimum biasing distribution from any constraint class is that distribution which minimizes a *specific* f-divergence to the global optimal distribution. Via a theorem developed in this work, we derive the unique biasing distribution from the constraint class described above which minimizes *every* f-divergence or Ali-Silvey distance to the optimal statistics. As a special case, we show that this distribution minimizes the Importance Sampling variance among all distributions from this constraint class. Moreover, it is shown that the savings over standard Monte Carlo simulations obtained through the use of this distribution can be made arbitrarily close to the optimal savings by varying the selection of the parameters of the constraint set.

It is shown that for the case of estimating the probability of rare events, the constrained optimal biasing distribution can be made completely independent of the parameter of interest, thus readily admitting to a practical implementation. We show further that the computational savings obtained through the use of this biasing distribution can be made arbitrarily large by adjusting the parameters of the constraint class.

We conclude this work with an asymptotic analysis of the performance gains of these biasing distributions. Necessary and sufficient conditions are given for the asymptotic gains achieved through these distributions to become unbounded as the probability of the rare event diminishes. Furthermore, a methodology for selecting sequences of Borel bounding sets which satisfy these conditions and yet renders simulations which are amenable to implementation is presented.

[1]Supported in part by the National Science Foundation under Grant NCR-9109858 and in part by Rome Laboratories under contract F30602-92-C-0053.

[2]Supported in part by the IBM Contract 89-832302 and the Texas Advanced Technology Grant 003604-018.

References

[1] S. M. Ali and D. Silvey, "A General Class of Coefficients of Divergence of One Distribution from Another," *J. Royal Stat Soc.*, vol. 28, pp. 131–142, 1966.

[2] I. Csiszar, "Information-Type Measures of Difference of Probability Distributions and Indirect Observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 299–318, 1967.

[3] G. C. Orsak and B. Aazhang, "Constrained Solutions in Importance Sampling via Robust Statistics," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 2, pp. 307–316, March 1991.

IMPORTANCE SAMPLING USING GEOMETRY*

A. Dabak *D.H. Johnson*

Computer & Information Technology Institute
Electrical & Computer Engineering Department
Rice University
Houston, Texas 77251-1892

ABSTRACT

The problem of finding an importance sampling biasing density for estimating the performance of an optimal binary detection system is addressed geometrically. This geometric approach allows us to find an importance sampling biasing density for any pair of mutually absolutely continuous nominal densities unlike other methods which are problem specific. We prove that the biasing density lying *geometrically halfway* between the nominals gives an asymptotically infinite importance sampling gain as system performance improves and the probability of error tends to zero.

1. BACKGROUND

Consider the standard binary hypothesis testing problem of determining which of two hypotheses (H_0 and H_1) is true based upon a set of observations $\underline{\omega} = \{\omega_1, ..., \omega_n\} \in \Omega$, the observation space. P_0 and P_1 denote the respective probability measures on Ω that correspond to these hypotheses. The *optimal* detector under a variety of performance criteria is the likelihood ratio test. We focus here in two criteria: minimum probability of error and Neyman-Pearson.

Calculation of performance probabilities is usually so complicated that numerical estimation is required. The number of Monte Carlo simulations to be performed to obtain a reliable numerical estimate of the performance can be very large since it is inversely related to the performance. A technique known as *importance sampling* has been employed to greatly reduce the number of simulations required to produce accurate estimates [3]. Here, observations are generated according to an alternate model specified by the *biasing density*; the utility of importance sampling hinges on finding a biasing density that can greatly reduce the number of required simulations to achieve accurate performance estimates. A measure used to quantify this reduction is the *importance sampling gain* $\Gamma = \frac{N}{M}$, the ratio of the number of trials N required for Monte Carlo simulations and the number of trials M required for the importance sampling technique such that the estimates' variances are equal. In general, for arbitrary nominal densities, no straightforward method is known to produce a biasing density that provably yields gain; usually problem specific, *ad hoc* methods are used. In this paper, we employ the natural geometry underlying detection problems to find the importance sampling biasing density.

2. GEOMETRIC IMPORTANCE SAMPLING

We assume that P_0 and P_1 are mutually absolutely continuous with respect to each other. Let p_0 and p_1 denote the probability densities of P_0 and P_1 with respect to some other absolutely continuous measure P. Employing the tools of differential geometry, we have analyzed elsewhere [1, 2] the non-Riemannian geometry of the space of all probability measures on Ω mutually absolutely continuous with respect to P_0 and P_1. In this geometry, the *natural* path, the so called *geodesic* connecting p_0 and p_1, is the exponential mixture density.

$$p_u(\underline{\omega}) = \frac{p_0^{(1-u)}(\underline{\omega}) p_1^u(\underline{\omega})}{J_u}; \quad 0 \leq u \leq 1 \qquad (1)$$

The normalization factor J_u is a strictly convex function; hence, there exists a unique $0 < \nu < 1$ so that $J_\nu = \inf_{0 \leq u \leq 1} J_u$. The density p_ν has the property that it lies midway between p_0 and p_1 in the sense of Kullback-Leibler information. We have shown that choosing p_ν consistently yields significant importance sampling gain. Under the minimum probability of error criterion, the gain Γ is lower bounded by the reciprocal of the error probability raised to a constant power. Asymptotically, as performance improves, the importance sampling gain tends to infinity. Under the Neyman-Pearson criterion, the density p_1 is used in Monte-Carlo simulations. When the observations are IID, we have shown that computationally efficient estimates of the miss probability $\Pr[H_0|H_1]$ occur when importance sampling schemes employ the *other* nominal p_0. The importance sampling gain is bounded according to $\ln \Gamma / n > I(P_1^1|P_0^1)$, where $I(P_1^1|P_0^1)$ is the Kullback-Leibler information between observations' marginal distributions.

The biasing densities corresponding to some standard problems are very interesting. When the nominal biasing densities are equal-variance Gaussians with means m_0 and m_1, the geometric importance sampling biasing density equals a Gaussian with mean $\frac{m_0+m_1}{2}$ with the variance remaining the same. The geometric biasing density corresponding to nominals distributed with a Cauchy or the Generalized Gaussian density are multimodal. When the nominals are homogeneous Poisson densities with means λ_0 and λ_1, then a Poisson density with mean $\frac{(\lambda_1 - \lambda_0)}{\log(\lambda_1) - \log(\lambda_0)}$ is the biasing density. When the nominals are multivariate, shifted-mean (the two hypotheses correspond to deterministic signals in additive noise), densities symmetric about the mean, then $\nu = \frac{1}{2}$.

REFERENCES

[1] A G. Dabak. *A Geometry for Detection Theory*. PhD thesis, Rice University, Houston, Tx, 1992.

[2] A G. Dabak and Don H. Johnson. A geometry for detection theory. In *Proceedings Conf. Infor. Sc. Syst., Princeton, NJ, Princeton Univ.*, March 1992.

[3] G. Orsak and B. Aashang. On the theory of importance sampling applied to the analysis of detection systems. *IEEE trans. comm.*, 37(4):332-339, April 1989.

*Supported by ONR Grant N00014-89-J-3152.

Interference Channels with Correlated Sources

Masoud Salehi **Erozan Kurtas**
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115

Abstract

We investigate transmission of correlated information sources over an interference channel. A coding scheme for matching the source to the channel is developed and sufficient matching conditions between the source and the channel are derived

1 Introduction

So far, the problem of transmission of correlated sources over communication channels has been investigated for the multiple access channels[1], broadcast channels[2] and two-way channels[3]. In this work we study this problem for the interference channel. The existence of correlation between sources makes it possible for the encoders to partially cooperate and this partial cooperation results in better performance compared to the case of independent messages. An encoding scheme is proposed and based on this scheme sufficient conditions for reliable transmission of correlated sources over an interference channel are obtained. The results are then applied to two classes of interference channels for which the capacity region is already known.

2 Main Results

The interference channel is a mathematical model for communication between M transmitter-receiver pairs over a single communication medium. The existence of correlation between the two information sources makes it possible that the encoders, although located separately, cooperate to some extent. The cooperation between the encoders can be employed in designing improved codes. We employ an encoding scheme based on the using the covering lemma and the correlation preserving coding. In the first stage of encoding we use covering to represent the information about each source embedded in the other source by an auxiliary random variable. The next step is to provide partial cooperation between the encoders. The codewords generated in this step statistically depend on the information content of the each source output and the auxiliary random variable representing the information about the other source output. The decoding scheme and error analysis are based on using the properties of jointly typical sequences. Our main result is given the following theorem.

Theorem: Let a discrete-memoryless interference channel be denoted by input alphabets, $\mathcal{X}_1, \mathcal{X}_2$, output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$, and a probability transition matrix, $p^*(y_1, y_2|x_1, x_2)$, where $p^*(\mathbf{y}_1^n, \mathbf{y}_2^n|\mathbf{x}_1^n, \mathbf{x}_2^n) = \prod_{i=1}^n p^*(y_{1i}, y_{2i}|x_{1i}, x_{2i})$, and let two correlated information sources be generated according to independent drawings of random variables S and T with joint PMF $p^*(s,t)$. Then, if there exist two auxiliary random variables U and V such that

$$p(q,s,t,u,v,x_1,x_2,y_1,y_2) = p^*(s,t)p(q)p(u|s,q)p(v|t,q)p(x_1|s,u,q) \times \times p(x_2|t,v,q)p^*(y_1,y_2|x_1,x_2)$$

and

$$\begin{aligned}
&H(S|UVQ) < I(Y_1; SX_1|UVQ)\\
&H(S|VQ) < I(Y_1; SX_1|VQ)\\
&H(S|UVQ) + I(T;V|UQ) < I(Y_1; SVX_1|UQ)\\
&H(S|VQ) + I(T;V|Q) < I(Y_1; SVX_1|Q)\\
&H(T|UVQ) < I(Y_2; TX_2|UVQ)\\
&H(T|UQ) < I(Y_2; TX_2|UQ)\\
&H(T|UVQ) + I(S;U|VQ) < I(Y_2; TUX_2|VQ)\\
&H(T|UQ) + I(S;U|Q) < I(Y_2; TUX_2|Q)
\end{aligned}$$

the correlated sources (S,T) can be reliably transmitted via the interference channel.
These results are then applied to some special cases for which the capacity of the interference channel is already known.

*This work was supported by the National Science Foundation Grant NCR-9101560

References

[1] T. M. Cover, A. ElGamal, and M. Salehi, "Multiple-access channels with arbitrarily correlated sources," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 648--657, November 1980.

[2] T. S. Han and M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Transactions on Information Theory*, vol. IT-33, pp. 641--650, September 1987.

[3] M. Salehi, "Restricted two-way channels with correlated sources," in *Proceedings of the twenty-eighth annual Allerton Conference on communications control and computing*, October 1990.

Multiuser Water-Filling

ROGER S. CHENG
Department of Electrical and Computer Engineering
University of Colorado, Boulder, CO 80309
chengr@spot.colorado.edu

SERGIO VERDÚ
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544

Abstract

We find the capacity region of a two-user Gaussian multiaccess channel with intersymbol interference (ISI) where the inputs pass through respective linear systems and are then superimposed before being corrupted by an additive Gaussian noise process.

We give a novel geometrical method to obtain the optimal input power spectral densities and the capacity region. This method can be viewed as a nontrivial generalization of the single-user water-filling argument. We show that as in the traditional memoryless multiaccess channel, FDMA, with optimally selected frequency bands for each user, achieves the total capacity of the K-user Gaussian multiaccess channel with ISI. However, the capacity region of the two-user channel with memory is, in general, not a pentagon unless the channel transfer functions for both users are identical.

Summary

In a recent paper [1], a limiting expression for the capacity regions of multiaccess channels with memory was obtained. Such a limiting expression was explicitly evaluated for some channels with memory in [1] and [2]. In [3], we show that the limiting expression of [1] can be used to obtain a computable capacity region formula for Gaussian linear multiple-access channels with finite ISI. We extend the single-user water-filling argument to the two-user case and derive a geometrical method to obtain the optimal input power spectral densities (PSDs). We show that the optimal input PSDs of the users that maximize the total capacity do not overlap in the frequency domain. As in the traditional memoryless multiaccess channel, FDMA with optimally selected frequency bands and optimally shaped PSDs achieves the total capacity.

We consider a general Gaussian multiaccess channel with ISI

$$Z_i = \sum_{k=1}^{K} \sum_{j=0}^{n} H_{k,j} X_{k,i-j} + N_i,$$

where Z_i is the output of the channel, $X_{k,i}$ is the i^{th} symbol sent by user k, and N_i is a zero-mean m-dependent stationary Gaussian noise process (i.e., $R_n = 0, \forall |n| > m$). We assume that the k^{th} user has power constraint W_k and all the channels seen by the users have finite-length impulse responses with length less than or equal to n.

The capacity of a single-user Gaussian channel with ISI is obtained using the Karhunen-Loève expansion. This expansion decomposes the channel into independent parallel memoryless Gaussian channels whose capacities are well known; thereby reducing the problem to one of optimal power allocation into various channels. It is crucial to note that the kernel used in the Karhunen-Loève expansion depends on the ISI coefficients. In the two-user Gaussian channel with ISI, there are two sets of ISI coefficients, one for each user. If the channels seen by the users are identical, the traditional procedures can be applied and the capacity region has been obtained in [4, 5]. If the sets of ISI coefficients are not the same, a similar decomposition cannot be applied since no kernel can simultaneously decompose the signals from both users.

Therefore, in order to obtain the result in the multiuser case, a new approach based on the circular channel methods of [2] and [6] are employed. This approach enables an orthogonal decomposition of the channel using the discrete Fourier transform which is independent of the ISI coefficients. In this paper, we employ these ideas and the limiting expression for the capacity region of multiaccess channel with memory in [1] to obtain the capacity region of the Gaussian multiaccess channel with ISI.

This work was supported by the Office of Naval Research under Grant N00014-90-J-1734.

We also extend the single-user water-filling argument to the two-user case. We derive a geometrical method to obtain the optimal input PSDs. It turns out that this geometrical argument can be explained via two main ideas: the equivalent channel idea and the successive decoding idea (decode one user's information while treating the other user's information as noise first and then decode the remaining user's information). The equivalent channel idea bears some resemblance to the single-user water-filling argument in the sense that it obtains graphically the optimal input power distribution over the frequency domain. It can be applied directly to the single-user channel to obtain the optimal input PSD. Roughly speaking, in the two-user case, the equivalent channel idea determines graphically the optimal distribution of the *total* power over the frequency domain, while the successive decoding idea determines, again graphically, the optimal split of the total power among the users for each frequency.

In particular, we show that the optimal input PSD pair achieving the total capacity can be obtained graphically using the equivalent channel idea alone. Moreover, the optimal PSD pair do not overlap; hence, as in the memoryless multiaccess channel, FDMA, with optimally selected bands and optimally shaped PSDs, achieves the total capacity of the multiaccess channel with ISI.

Theorem 1

For any K-user m-dependent Gaussian multiaccess channel with finite intersymbol interference and power constraints $W_1, \ldots, W_K$, the total capacity can be achieved by FDMA with optimal input PSD K-tuple, $(S_1(w), \ldots, S_K(w))$, where

$$S_k(w) = \frac{\hat{S}_k(w)}{b_k}$$

$$\hat{S}_k(w) = \begin{cases} \left[1 - b_k T_k^{-1}(w)\right]^+ & \text{if } b_k T_k^{-1}(w) \le b_l T_l^{-1}(w) \text{ for } l \ne k, \\ 0 & \text{otherwise.} \end{cases}$$

$T_k(w) = |H_k(w)|^2/N(w)$ is the magnitude square of the transfer function over the noise PSD, and $b_1, \ldots, b_K$ are chosen such that

$$\frac{1}{\pi} \int_0^{\pi} \hat{S}_k(w) dw = b_k W_k$$

for $k = 1, \ldots, K$. □

References

[1] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Transactions on Information Theory*, vol. IT-35, no. 3, pp. 605–619, May 1989.

[2] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. IT-35, no. 4, pp. 733–751, July 1989.

[3] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: capacity region and multiuser water-filling," to appear in *IEEE Transactions on Information Theory*.

[4] C. W. Keilers, *The Capacity of the Spectral Gaussian Multiple-access Channel*. PhD thesis, Stanford University, May 1976.

[5] C. M. Zeng, N. He, and F. Kuhlmann, "Capacity region of a waveform Gaussian multiple-access channel," in *Book of Abstracts of the 1990 International Symposium on Information Theory*, San Diego, CA, p. 93, January 1990.

[6] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Transactions on Information Theory*, vol. IT-34, no. 3, pp. 380–388, May 1988.

THE SMALLEST LIST FOR THE ARBITRARILY VARYING CHANNEL

Brian Hughes
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, Maryland 21218

Abstract

The capacity $C(L)$ of the arbitrarily varying channel for deterministic list codes of *fixed* list size L is considered under the average probability of error criterion. When the random coding capacity C_r is positive, it is shown that

$$C(L) = \begin{cases} C_r, & L > L^*, \\ 0, & L \leq L^*, \end{cases}$$

where L^*, called the *symmetrizability*, is a computable, non-negative integer. Thus $L^* + 1$ is the smallest list size for which a positive rate is achievable.

Summary

At the 1990 IEEE Information Theory Workshop, M.S. Pinsker conjectured the following theorem: For an arbitrarily varying channel (AVC), every rate below the random code capacity C_r is achievable with *deterministic list codes of constant list size*, if the average probability of error criterion is used. Ahlswede and Cai [1] proved this conjecture by showing that codes exist with rate R, list size L, and average probability of error λ for all

$$L > \frac{\log |\mathcal{S}|}{\lambda E(R)}, \quad R < C_r, \tag{1}$$

where $E(R)$ is the random coding reliability function of the AVC, and $\mathcal{S}$ is the channel state alphabet. This list size depends on R, λ, and $\mathcal{S}$, and grows without bound as $\lambda \to 0$ or $R \to C_r$.

The contribution of this paper is to show that Pinsker's conjecture holds for a constant list size L that depends only on the channel and not on R, λ, or $\mathcal{S}$. Moreover, we determine the *smallest* list size for which the conjecture is valid.

Consider a discrete, memoryless AVC with transition probability $W : \mathcal{X} \times \mathcal{S} \to \mathcal{Y}$. Let X, S, and Y be random variables, with p.m.f. $P(x)Q(s)W(y|x,s)$.

Definition 1: An AVC is *m-symmetrisable* if there exists a channel $U : \mathcal{X}^m \to \mathcal{S}$ such that the channel $V : \mathcal{X}^{m+1} \to \mathcal{Y}$ defined by

$$V(y|x, x_1, \ldots, x_m) \equiv \sum_{s \in \mathcal{S}} W(y|x,s)U(s|x_1, \ldots, x_m)$$

is invariant under all permutations of the arguments $x, x_1, \ldots, x_m$ for all y, x, $x, x_1, \ldots, x_m$. By definition, we take all AVCs to be 0-symmetrisable. ◇

Remarks: (1) This definition generalises the symmetrisability condition of [2]. (2) If an AVC is m-symmetrisable then it is also m'-symmetrisable for all $0 \leq m' \leq m$.

Theorem 1: If an AVC is m-symmetrisable then

$$\begin{aligned} m &\leq \min_X \max_S \frac{I(S \wedge Y|X)}{I(X \wedge Y)} \\ &\leq (C_r)^{-1} \min\{\log|\mathcal{Y}|, \log|\mathcal{S}|\} . \end{aligned}$$

◇

Definition 2: The maximum m for which the AVC is m-symmetrisable is called the *symmetrizability* and is denoted by L^*. ◇

Theorem 1 and Definitions 1 and 2 imply a computable characterisation of L^*.

Theorem 2: The list-L capacity of the AVC under the average probability of error criterion is

$$C(L) = \begin{cases} C_r, & L > L^*, \\ 0, & L \leq L^*. \end{cases} \tag{2}$$

◇

Example: For $\mathcal{X} = \mathcal{S} = \{0,1\}$, $\mathcal{Y} = \{0,1,2\}$ and

$$W(y|x,s) \equiv \begin{cases} 1 & y = x + s, \\ 0 & y \neq x + s, \end{cases} \tag{3}$$

$C_r = 0.5$ bits/channel use, and $L^* = 1$. Consequently, there exist codes for all $R < 0.5$ such that the receiver can reliably narrow the transmitted message to one of two possibilities, but no further.

References

[1] R. Ahlswede and N. Cai, "Two proofs of Pinsker's conjecture concerning AV channels," *IEEE Transactions on Information Theory*, IT-37 (6), pp. 1647-1749, November 1991.

[2] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: positivity, constraints," *IEEE Transactions on Information Theory*, IT-34 (2), pp. 181-193, March 1988.

Supported in part by ARO Grant DAAL03-89-K-0130.

Coding Strategies for the Permuting Jammer Channel

by

Wah Keung Chan*
5411 Waverly, Montréal, Québec,
CANADA, H2T 2X8

SUMMARY

The difficulty of determining results on the capacity of finite-state channels with memory resides with the added complexity due to the memory. The natural approach is to study a simple model. The Permuting Channel is one such model. The best way to introduce the Permuting Jammer Channel is by an example. Consider Blackwell's Trapdoor channel [1].

Example. *Consider the case of two trapdoors with each door having the same probability of being opened, as shown in Figure 1. Initially (Fig. 1a), a ball labeled either 0 or 1 is present in one of the two slots. Then (Fig. 1b) a ball, either a 0 or 1, is placed in the empty slot, after which (Fig. 1c) one of the trapdoors opens. The ball lying above the open door then falls through. The door closes (as in Fig. 1a) and the process is repeated.*

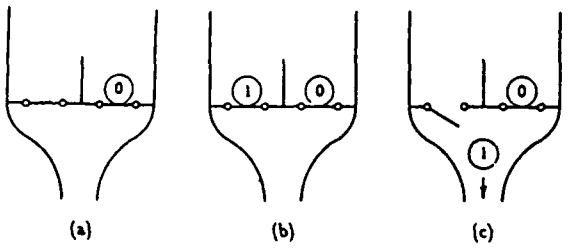


Figure 1. Blackwell's trapdoor channel [1]

We generalize this example as follows:

Definition. The *permuting channel* is a finite-state channel consisting of $\beta + 1$ trapdoors with each trapdoor compartment capable of holding only 1 character. At the start, β of these compartments are occupied. The alphabet consists of α characters. The initial state, S_0, represents the nature of the β characters at the start.

The operations of this channel, as introduced by Ahlswede and Kaspi [2], depends on three participants as shown

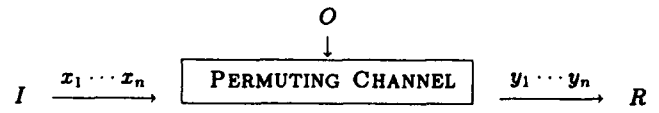


Figure 2. I – Sender, O – Trapdoor Selector, R – Receiver.

Definition. In the permuting *jammer* channel (PJ Channel), O acts as a jammer in frustrating I, the message sender, by scrambling the output. Here, we shall consider the condition that the initial state is known to all users.

In [2], Ahlswede and Kaspi found the capacity of the PJ channel in the binary case. Piret [3] solved the case of $\beta = 1$ memory. In this paper, the capacity for the case $\alpha = 3$ is established.

To find the capacity, we follow the lead of Ahlswede and Kaspi, of constructing maximal codes for particular input length n. Let us define certain properties of codes.

*Supported by a FCAR Post-Graduate Fellowship

Definition. A code is called *maximal* for particul : input length n if no other code has larger cardinality.

Definition. A code U is *packed* or is a *packed code* if it cannot accommodate any other codewords.

A maximal code is necessarily packed, but a packed code need not be maximal. Packedness is useful as a first test for maximality.

Definition. A repeated code is the cartesian product U^m of a code U.

Ahlswede and Kaspi [2] established their result by considering $U_1(\alpha,\beta) := \{\overbrace{11\cdots1}^{\beta+1}, \overbrace{22\cdots2}^{\beta+1}, \ldots, \overbrace{\alpha\alpha\cdots\alpha}^{\beta+1}\}$ and verifying that $U_1(2,\beta)^m$ is a maximal code for length $n = (\beta+1)m$.

We first note that

Lemma. *For $\alpha \geq 2, \beta \geq 1$ and arbitrary S_0, $U_1(\alpha,\beta)$ is the only maximal code for $n = \beta + 1$.*

Thus, in the binary case, the repeated code, $U_1(2,\beta)^m$, is maximal. We want to explore conditions on maximal codes which yield maximal repeated codes.

As a matter of notation, we let $S_0|x \rightarrow y$ to denote the output y is reachable from input x and initial state S_0. Furthermore, let $Y(S_0|x)$ denote the cover of x, i.e., the set of all possible y reachable from x, and $Y(S_0|U) = \cup_{x\in U} Y(S_0|x)$.

Definition. A set $U \subset C^n$ has the *Universal Property* if for all $S_1, S_2 \in \mathbf{S}$ and for all $x \in C^n$, $Y(S_1|x) \cap Y(S_2|U) \neq \emptyset$, i.e., there exist $u \in U$ and $y \in C^n$ such that

$$S_1|x \longrightarrow y \quad \text{and} \quad S_2|u \longrightarrow y.$$

A code U is a *Universal Code* or is said to be *universal* if it has the Universal Property.

Universal codes are packed. To see this, just let $S_1 = S_2$. Furthermore, repeated Universal codes are also Universal codes.

Definition. By *extending* a set $Z' \subset C^n$ to a code $Z \subset C^{n+k}$ we mean to augment each sequence $z_i \in Z'$ by adjoining to it attachment words $c_{ij} \in C^k$ such that $Z = \{z_{i,j} = z_i c_{ij}\}$ is a code. If there are in total a_i c_{ij}'s adjoint to z_i, we say that z_i *contributes* a_i codewords to Z.

Definition. Let $U \subset C^n$ be a code, $k = |U|$, and $Z' \subset C^n$ a set extended to a code $Z \subset C^{2n}$. A maximal code U is *strongly maximal* if for any set V of atmost k sequences with pairwise non-disjoint covers, i.e. $V = \{z_1, z_2, \ldots, z_l (\in Z')\}$ s.t. $Y(S_0|z_i) \cap Y(S_0|z_j) \neq \emptyset$, V contributes atmost k codewords.

We note that strongly maximal codes are universal codes. The truth of the converse statement is not known.

Result 1. *Repeated strongly maximal codes are maximal.*

This establishes *strongly maximal* as the condition sufficient for constructing maximal repeated codes. Is there a weaker condition? Examples of strongly maximal codes are hard to to find. For the ternary character set we found $U_1(3,\beta)$ to be strongly maximal, and thus $U_1(3,\beta)^m$ is maximal for $n = (\beta+1)m$. Thus

Result 2. *For $\alpha = 3$, $\beta \geq 1$ and for all S_0 the capacity of the permuting jammer channel is* $C_J(3,\beta,S_0) = \dfrac{\log 3}{\beta+1}$.

References

[1] R. Ash, *Information Theory*, New York: Wiley, 1965, pp. 211–229.

[2] R. Ahlswede and A. Kaspi, "Optimal Coding Strategies For Certain Permuting Channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 310–314, May 1987.

[3] Ph. Piret, "Two Results on the Permuting Mailbox Channel," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 888–892, July 1989.

SOURCE DIVISIBILITY AND RELATED PROBLEMS

V.N. Koshelev
Council for Cybernetics
Russian Academy of Sciences
Vavilov str. 40
117333 Moscow, Russia

Multi-level coding schemes reveal a special information-theoretical problem called the divisibility problem that is connected to a tree-like structure of the codes. It can be traced in source and channel coding as well as to hierarchical covering metric spaces. In the broadcast source coding scheme, divisibility looks like a problem of additional information necessary to achieve per-letter distortion $\epsilon_2 < \epsilon$, provided that the distortion level ϵ_1 has been already achieved. We say that the source is divisible for the pair (ϵ_1, ϵ_2) if this additional amount of information (per source letter) is equal to $R(\epsilon_2) - R(\epsilon_2)$, where $R(\epsilon)$ is distortion rate function of the source. It is easy to show that the source is divisible if and only if the matrix equation $q_{\epsilon_2} \cdot s = q_{\epsilon_2}$ holds, where q_ϵ is the test-channel matrix of the source, has a stochastic solution $s = s_{\epsilon_2 \epsilon_1}$. The divisibility equatioon was introduced by Koshelev (*Probl. Peeredachi Inform.*, 3, 1980, pp. 31–49, in Russian); recently it was also described by Equitz and Cover (*IEEE Trans. on IT*, 2, 1991, pp 269-275).

To solve the equation we propose a method of local divisibility letting $\epsilon_1 = \epsilon_2 + \delta$, $\delta \to 0$. Using it we find strong conditions under which equiprobable ternary memoryless source with balanced distortion measure is divisible. We discuss divisibility of metric spaces and construction of hierarchical ϵ-nets. Finally we consider multi-level channel coding and the problem of channel-input divisibility.

DATA COMPRESSION WITH SIDE INFORMATION AND GRAPH ENTROPY

Alon Orlitsky

Room 2C-361, AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974

Abstract

X and Y are random variables. A person who knows X wants to convey it to another person who knows Y. What is the minimum number, $L(X|Y)$, of bits he must transmit on the average? Let G be an appropriately defined *communication hypergraph* of X and Y, and let $H(G,X)$ be its *graph entropy*.[1] We show that for all (X,Y) pairs, $L(X|Y) \geq H(G,X)$ and that for a large class of (X,Y) pairs $L(X|Y) \leq H(G,X) + \log e + 1$.

Summary

X is a random variable. A person who knows X wants to convey it to another. What is the minimum number, $L(X)$, of bits he must transmit on the average? The well known answer to this question is:[2]

$$H(X) \leq L(X) \leq H(X) + 1 .$$

Surprisingly, the answer to the following, closely related question, is not known. X and Y are random variables. A person who knows X wants to convey it to another person who knows Y. Feedback is not allowed. What is the minimum number, $L(X|Y)$, of bits he must transmit on the average?

As usual, we assume that both persons agree in advance on an encoding of X that must be prefix free given the value of Y. Standard reasoning yields

$$H(X|Y) \leq L(X|Y) \leq H(X) + 1 .$$

Yet a simple example shows that neither bound is tight. For $\epsilon \in [0,1)$ define

$$p_\epsilon(x,y) \stackrel{\text{def}}{=} \begin{cases} \frac{1-\epsilon}{n} & \text{for } x = y, \\ \frac{\epsilon}{n^2-n} & \text{for } x \neq y, \end{cases}$$

and let (X,Y) be distributed according to $p_\epsilon(x,y)$. Then $H(X) = \log n$, while $H(X|Y) = h(\epsilon) + \epsilon \log(n-1)$. For $\epsilon = 0$ it is easy to show that $L(X|Y) = 0 = H(X|Y)$ whereas for $\epsilon > 0$ we have $L(X|Y) = \log n = H(X)$.

In this paper we provide a partial solution to the above question. We define G, the *communication hypergraph* of X and Y, and show that for many (X,Y) pairs, $L(X|Y)$ is roughly $H(G,X)$, the *graph entropy* of G and X. More precisely, we show that in the inequality

$$H(G,X) \leq L(X|Y) \leq H(G,X) + \log e + 1 ,$$

the lower bound holds for all (X,Y) pairs, and that the upper bound holds for a large class of (X,Y) pairs. We do not know whether the upper bound holds for all (X,Y) pairs.

Graph entropy was introduced by J. Körner [1]. It was recently applied to prove lower bounds on perfect hashing, circuit complexity, and sorting. Yet these proofs used technical properties of this formally-defined (see below) functional and did not shed light on its underlying meaning. Our bounds provide an intuitive interpretation for this increasingly prevalent measure.

We conclude the summary by defining the communication hypergraph and its graph entropy. Let (X,Y) be distributed over $\mathcal{X} \times \mathcal{Y}$ according to some probability distribution $p(x,y)$. The *communication graph* G of X and Y has $\mathcal{X}$ as its vertex set, and for every $y \in \mathcal{Y}$ it has the hyperedge $e_y \stackrel{\text{def}}{=} \{x : p(x,y) > 0\}$. This communication hypergraph is equivalent to a graph defined by [2] and used in [3] to analyze the number of bits needed in our problem if the two persons are allowed to communicate back and forth. This paper concerns only the one-way version of this problem.

Let G be a hypergraph and let X be a random variable ranging over its vertices (in our problem the vertices of G were defined by Y). A set of vertices of G is *independent* if no two of its members belong to the same edge. Denote by $\gamma(G)$ the collection of independent sets in G. The hypergraph entropy $H(G,X)$ of G and X is

$$H(G,X) \stackrel{\text{def}}{=} \min\{I(X;Z) : X \in Z \in \gamma(G)\} .$$

Namely, it is the minimum mutual information between X and any random variable Z ranging over independent sets of G and constrained to always contain X.

References

[1] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. *Proc. of the 6th Prague Conference on Information Theory*, pages 411–425, 1973.

[2] H. Witsenhausen. The zero-error side information problem and chromatic numbers. *IEEE Transactions on Information Theory*, 22(5):592–593, 1976.

[3] A. Orlitsky. Worst-case interactive communication I: Two messages are almost optimal. *IEEE Transactions on Information Theory*, 36(5):1111–1126, September 1990.

[1] All entropies and logarithms are to the base 2, and e is the base of the natural logarithm.

[2] Here and below, $H(X)$ is the binary entropy of X and $H(X|Y)$ is the conditional binary entropy of X given Y.

Multiple-User Distributed Information Storage

James R. Roche
Room 2C-254, AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974

Abstract

Suppose that we have an information storage network with m users and n disks, each disk having the same capacity. Different users are connected to arbitrary (perhaps overlapping) subsets of the n disks, and some of the disks might fail. We wish to encode a binary information sequence such that for a specified m-tuple $(X_1, \ldots, X_m)$, the ith user can reliably recover the first X_i bits of the sequence.

There is a natural upper bound on each individual X_i. If this bound can be attained simultaneously for each of the m users, we say that *sequential refinement* is possible. We find necessary and sufficient conditions for a storage network to admit sequential refinement.

Summary

Consider an information storage network with m users, $U_1, U_2, \ldots, U_m$, connected to arbitrary (perhaps overlapping) subsets $S_1, S_2, \ldots, S_m$ of a set of n disks, $\{D_1, D_2, \ldots, D_n\}$. For $1 \leq i \leq m$, let $n_i = |S_i|$. For simplicity, we assume that the n disks all have the same capacity, C bits. If user U_i is guaranteed access to just k_i disks of S_i at any given time ($k_i \leq n_i$), then he can hope to recover at most $k_i C$ bits of information reliably.

We wish to encode an r-bit information sequence $(y_1, \ldots, y_r)$ so that each user U_i $(1 \leq i \leq m)$ can reliably recover the first X_i bits of the sequence, $(y_1, \ldots, y_{X_i})$. For which m-tuples $(X_1, \ldots, X_m)$ is such an encoding possible?

Ideally, we might hope to achieve $(X_1, \ldots, X_m) = (k_1 C, \ldots, k_m C)$. In such a case, we say that *sequential refinement* is achievable. Unfortunately, this is not always possible. We find necessary and sufficient conditions for a network to admit sequential refinement. Before stating the conditions (Theorem 1), we give several definitions.

Definition 1 *Let* $k_{\max} = \max_{1 \leq i \leq m} k_i$.

Definition 2 *Let E be the set of* edges *between users and disks; i.e.,*

$$E = \{(i, j) : \textit{User } U_i \textit{ is connected to disk } D_j\}.$$

Definition 3 *For $1 \leq j \leq n$, associate with disk D_j the* disk degree

$$d_j = \min_{i\,(i,j) \in E} k_i.$$

Definition 4 *For $1 \leq i \leq m$ and $1 \leq s \leq n_i$, let $N_i(s)$ be the number of disks D_j connected to user U_i that satisfy the inequality $d_j \leq s$.*

Remark: It follows from Definitions 3 and 4 that $N_i(s) = n_i$ for $k_i \leq s \leq n_i$.

The conditions for sequential refinement can be shown to be related to the triangularity of certain incidence matrices determined by the network topology. We therefore make the following definition.

Definition 5 *User U_i $(1 \leq i \leq m)$ is* triangular *if $N_i(s) \leq s$ for all $s \leq k_i - 1$. A storage network is* fully triangular *if each user is triangular.*

Theorem 1 *Assume that each disk in a storage network of n disks has capacity C bits. If the storage network admits sequential refinement, it is fully triangular; the converse holds if $C \geq n$, or if $k_{\max} \leq n/2$ and $C \geq k_{\max} \log_2 n$.*

In practice the number of bits per disk, C, will generally be much larger than the number of disks, n. Thus in all cases of practical interest, a network admits sequential refinement if and only if the network is fully triangular. The necessity of this condition follows from simple information-theoretic arguments. The sufficiency can be established by a constructive scheme that encodes information using linear algebra over Galois fields.

Huffman Algebras for Independent Random Variables

Cheng-Shang Chang Joy A. Thomas
T.J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598

Abstract

The Huffman algorithm was originally devised to construct the minimal expected length instantaneous source code for a random variable. In this paper, we identify the Hardy-Littlewood-Polya inequality as the key step in the proof of the optimality of the Huffman algorithm and use this to provide an unified framework for various applications of the Huffman algorithm. Quantities that are analogous to entropy are identified for these applications. We also consider the case when the weights of the leaves of the tree are independent random variables. A Huffman algebra is defined, which provides conditions under which the Huffman algorithm is optimal for the case of random weights. In particular, it is shown that the "most balanced" tree is optimal for the case of independent and identically distributed weights for any arrangement increasing Huffman algebra.

The Huffman coding algorithm is a greedy bottom up tree building algorithm constructs the optimum source code for a random variable, i.e., it finds a binary tree that minimises the weighted lengths $\sum p_i l_i$, where p_i is the probability of symbol i and l_i is the depth of the corresponding leaf. The same greedy tree building algorithm was later applied to construct trees that minimise other tree functionals such as $\max_i(p_i + l_i)$, which has applications in circuit design and parallel processing[3].

In this paper, we provide an unified framework for the different applications of the Huffman algorithm and extend some of the results to the case when the weights of the leaves are independent random variables. In the proof of the optimality of the Huffman algorithm, the key step turns out to be a rearrangement inequality called the *Hardy-Littlewood-Polya inequality*, which states that for any real numbers, a, b, c and d, with $a \le b$ and $c \le d$, then $a \cdot d + b \cdot c \le a \cdot c + b \cdot d$. This inequality is used to show that the longest codewords are associated with the lowest weights.

Deterministic weights: Motivated by this inequality, we consider a two operator algebra $(S, \oplus, \otimes)$ that satisfies this rearrangement inequality with $+$ replaced by $\oplus$ and $\cdot$ replaced by $\otimes$. We call these algebras arrangement increasing Huffman algebras (AIHA) (extending an earlier definition by Knuth[4] of one operator Huffman algebras). We also define arrangement decreasing Huffman algebras (ADHA) when the inequality in the Hardy-Littlewood-Polya inequality is reversed. We show that if we define the cost of a tree T as $w(T) = \sum_{\oplus_{i=1}^m}(P_i \otimes \prod_{\otimes l_i}\gamma)$, then the Huffman algorithm on an AIHA (resp. ADHA) will produce an optimal tree that minimises (resp. maximises) the cost of the tree. (γ is a constant that represents the cost of one level of the tree.)

Table 1. Huffman algebras for deterministic weights

Algebra $(S, \oplus, \otimes)$	Type	Objective
$(\mathcal{R}^+, +, \cdot)$	AIHA	$\min \sum_{i=1}^m p_i \gamma^{l_i}$
$(\mathcal{R}, \max, +)$	AIHA	$\min \max_i(p_i + \gamma l_i)$
$(\mathcal{R}, \min, +)$	ADHA	$\max \min_i(p_i + \gamma l_i)$
$(\mathcal{R}, \max, \min)$	AIHA	$\min \max_i(\min(p_i, \gamma))$
$(\mathcal{R}^+, \max, \cdot)$	AIHA	$\min \max_i(p_i \cdot \gamma^{l_i})$
$(\mathcal{R}^+, \min, \cdot)$	ADHA	$\max \min_i(p_i \cdot \gamma^{l_i})$

Just as the entropy of a random variable is a fundamental lower limit to the expected length of the Huffman code, we show that the various examples of Huffman algebras in Table 1 have analogous fundamental limits. In particular, we define $I(p) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^m p_i^\alpha\right)$, where $\alpha = \frac{1}{\log_2 \gamma + 1}$, and $J(p) = \log_2\left(\sum_{i=1}^m 2^{p_i}\right)$. We can then show that[3, 5]

$$I(p) \le \min_T \sum p_i \gamma^{l_i} < I(p) + 1$$

$$J(p) - 1 < \max_T \min_i (p_i + l_i) \le J(p) \le \min_T \max_i (p_i + l_i) < J(p) + 1$$

The quantity $I(p)$ is a scaled version of the Renyi entropy of the distribution p. The quantity $J(p)$ can be identified as the fundamental limit in a model of parallel processing in which tasks of length $p_1, p_2, \ldots, p_m$ are executed in parallel and their results are combined two at a time. We also define the concept of tree extensions that is analogous to block coding for source codes, and show that it is possible to get arbitrarily close to the fundamental limits using tree extensions.

Random weights: When the weights of the leaves are independent random variables, we ask the question—when is the Huffman algorithm optimal in terms of expected cost? To answer this question, we define Huffman algebras for independent random variables. Random weights introduce many new difficulties. For example, there is no total order among random variables; instead, various partial orderings have been defined, such as likelihood ratio ordering, stochastic ordering, and increasing convex ordering. Chang and Yao [1] derived stochastic versions of Hardy-Littlewood-Polya inequalities using three different partial orderings for the random variables.

Motivated by the requirements for the Hardy-Littlewood-Polya inequality, we define an arrangement increasing Huffman algebra $(S, \oplus, \otimes, <_1, <_2, <_3)$ as a set with two operators $\oplus$ and $\otimes$ and three different partial orders $<_1, <_2, <_3$, which satisfies various consistency properties for the orderings and also satisfies the Hardy-Littlewood-Polya inequality[1]. For Huffman algebras that are closed under $\oplus$ and $\otimes$, we can show that the Huffman algorithm produces the optimal tree. However, most Huffman algebras are not closed, and in these cases, it is difficult to proceed with the Huffman algorithm after the first step, since the new weight produced by combining the lowest two weights is not directly comparable with the other weights.

However, in the special case when the weights are i.i.d., we can use the concepts of stochastic majorisation to prove that the most balanced tree minimises the expected cost $w(T)$ for arrangement increasing Huffman algebras.The "most balanced" tree can be constructed from the top down using a procedure called the "power of 2" rule[2]. However, the "most balanced" tree is not always optimal for ADHA's, and we conclude with counterexample to illustrate that our intuition about "balanced" trees is not always valid for the stochastic case.

References

[1] C.S. Chang and D.D. Yao. Rearrangement, majorisation and stochastic scheduling. *IBM RC 16250*, 1990.

[2] C.R. Glassey and R.M. Karp. On the optimality of Huffman trees. *SIAM J. Appl. Math.*, 31:368–378, 1976.

[3] M.C. Golumbic. Combinatorial merging. *IEEE Trans. Comput.*, C-25:1164–1167, 1976.

[4] D.E. Knuth. Huffman's algorithm via algebra. *Journal of Combinatorial Theory, Series A*, 32:216–224, 1982.

[5] D.S. Parker. Combinatorial merging and Huffman's algorithm. *IEEE Trans Comput.*, TC-28:365–367, 1979.

Minimum Average Cost Testing for Partially Ordered Components

Marc J. Lipman and Julia Abrahams
Mathematical Sciences Division
Office of Naval Research
Arlington, VA 22217-5660

The problem of designing a sequence of optimal binary tests for the identification of a single faulty component is addressed. For components in linear order this is equivalent to the classical alphabetic coding problem solved by Hu and Tucker. For partially ordered components the problem is more difficult. Here, the problem is solved by reduction to a minimization over a set of alphabetic problems.

Our problem is, for a given partial order and assignment of probabilities to the nodes in its Hasse diagram (which correspond to the components), to find a sequence of upper sets to test so as to minimize the average number of tests to locate the defective node. An upper set is defined by:

Definition: If X is a node, the upper set of X, $U(X) = \{Y : Y \geq X\}$, is X together with the set of all nodes greater than X in the partial ordering.

The sequence of upper sets to be tested can be represented as a binary search tree. A search tree which minimizes the average number of tests is called an optimal search tree.

Brute force examination of all search trees consistent with the partial order is, in general, not a feasible approach to minimizing the average number of tests. For some particular sets of probabilities assigned to the components there are known ways to restrict the search. A number of these are given by Gilbert and Moore in the linear order case. One method for partial orders is provided in

Theorem 1: In an optimal search tree, at least one node of minimum probability must form a sibling pair with a node to which it is linked in the Hasse diagram.

Nevertheless, we require a more powerful approach in order to make progress on the general partial order search problem. We can systematically decompose any partially ordered problem into a set of linearly ordered problems from which the original can be solved indirectly. We use the Hu-Tucker algorithm on each linear problem. The linear problems solved by the Hu-Tucker algorithm involve composite nodes, corresponding to sets of sibling pairs in the search tree, and these must be expanded back out in evaluating the average number of tests required by the particular candidate solution. The final tree is the minimum over the several Hu-Tucker solutions.

More specifically, each link, AB, in the Hasse diagram, consider the Hasse diagram formed by substituting the composite node A+B (with probability given by the sum of P_A and P_B in obvious notation) for A and replacing links of the form AC or BC by (A+B)C links, and replacing links of the form CA or CB by C(A+B) links. Newly redundant links are eliminated. One new Hasse diagram is created for each link in the original. The collection of search trees consistent with these diagrams with composite nodes may include some which are no longer consistent with the original diagram. These can be identified easily, since they contain upper sets which do not belong to the original partial order. Therefore we eliminate intermediate Hasse diagrams which introduce upper sets not found in the original Hasse diagram. It is thus clear that by construction the set of search trees over which we now optimize is exactly the set consistent with the original partial order.

We can repeat the process for the consistent intermediate Hasse diagrams whose solutions are not yet available to us. Once all intermediate problems are solved and their associated expected number of tests determined, the minimum expected number of tests solution is the solution to the original partial order problem.

There are two simple types of intermediate problem whose solutions are available to us. These are linear orders and "V - orders".

Definition: A partial order is a V - order if it the union of two linear orders which have precisely their minimal element in common.

Theorem 2: A V - order, $A_n > A_{n-1} > \ldots > A_{j+1} > A_1$; $A_1 > A_2 > \ldots > A_{j-1} > A_j$, has the same optimal search tree as the linear order, $A_n > A_{n-1} > \ldots > A_{j+1} > A_j > A_{j-1} > \ldots > A_2 > A_1$.

To summarize, the proposed approach consists of the following. For each link in the Hasse diagram form a new diagram with a composite node whose new probability is the sum of the previous node probabilities. If new upper sets not present in the original Hasse diagram have been created, do not consider the new diagram further. If the new diagram is a linear order or a V - order, find its optimal test tree by the Hu-Tucker algorithm. Split the composite node into a pair of sibling nodes in the final test tree, and calculate the cost of the tree. If the new diagram cannot be solved immediately, begin the process of forming composite nodes again with that diagram. The optimal test tree for the original problem is the minimum cost solution from among these candidate trees.

Extended Synchronizing Codewords for Binary Prefix Codes*

Wai-Man Lam and Sanjeev Kulkarni

Department of Electrical Engineering
Princeton University
Princeton, NJ 08544

Summary

When variable length codes such as Huffman codes are transmitted through a noisy channel, any bit error can lead to loss of synchronization at the decoder and cause error propagation. Synchronizing codewords (SCs) have been previously proposed to resynchronize the decoder regardless of any preceding synchronization slippage. However, SCs retain one significant disadvantage. Although the decoder will be synchronized after decoding a SC, the decoded symbols after the SC may be shifted since the number of decoded symbols before the SC may be different from the original number (due to decoding a variable-length code in the presence of errors). Also, even though the decoder will be synchronized after decoding a synchronizing codeword, the decoder may not realize it was ever out of synchronization.

To overcome the drawbacks of synchronizing codewords, we introduce the concept of *extended synchronizing codewords* (ESCs). ESCs can guarantee both codeword and symbol synchronization, so that the symbols after the ESC will be decoded correctly and will be put in their correct positions. Thus, ESCs can be used as markers in the bit stream to prevent propagation of both decoding errors and symbol shift errors. An application of this to image coding has been studied in [3]. We give a formal definition of ESCs and derive some bounds on the amount of overhead necessary in designing a binary prefix code with an ESC, and study relationships between ESCs and SCs.

In this paper, we consider only binary prefix codes. For a source S with symbols $(s_1, \cdots, s_n)$, denote the probability distribution of the symbols by $(p_1, \cdots, p_n)$, and assume $p_i \geq p_{i+1}$. A code $C = \{c_1, \cdots, c_n\}$ is associated with the source S if c_i is assigned to the symbol s_i. The length of c_i is denoted by l_i, and the average codeword length of C is denoted by $E(C) = \sum_{i=1}^{n} p_i l_i$.

Our notion of an ESC is formally defined as follows. A codeword c_{es} of a prefix code C is defined to be an ESC if it satisfies the following conditions: (1) for all $\alpha \neq c_{es}$, if $c_{es} = \alpha\beta$ and α is a suffix of some codeword of C, then $\beta = \gamma\delta$, where γ is empty or a sequence of codewords, and δ is not empty and is not a prefix of any codeword, and (2) c_{es} is not a substring of any other codeword. These two conditions guarantee that regardless of prior slippage the decoder will not decode the ESC as parts of other codewords (as long as there is no error in the ESC itself). Hence, the ESC will be correctly decoded, so that the decoder knows that there was an error and can resynchronize after the ESC.

The following bounds on the amount of overhead needed in designing a code with an ESC can be obtained. If a source S is designed to have a prefix code C that has at least one ESC, then $E(C) > E(H)$ where H is the Huffman code for S. On the other hand, there exists a prefix code C with an ESC of probability p_n, and $E(C) = E(H) + p_n l_n$. Also, if the depth of a maximal complete subtree in a Huffman code is d, then there exists a prefix code C with an ESC of probability p_n, and $E(C) = E(H) + p_n(d+1)$. Using results from [2] and [1] it follows that a source S admits a prefix code C with an ESC of probability p_n and $E(C) = E(H) + \frac{1}{2}$. But we can show that for all $n \geq 6$, $p_n(d+1)$ is a tighter bound than $\frac{1}{2}$ (the two bounds are equal only when $n = 6$ and $n = 8$). In fact, $p_n(d+1) \to 0$ as $n \to \infty$.

Finally, some relationships between ESCs and SCs can be obtained. For example, if a code C has a SC c_λ, then $c_\lambda c_i$ is an ESC for C', where C' has the same codewords except that C' leaves the codeword c_i unused. Also, if a source S can be designed to have a SC c_λ (probability p_λ) with no overhead, then S admits a code C with an ESC and $E(C) = p_i + p_\lambda(l_i + 1)$, where i can be any codeword index.

*This work was supported in part by a grant from Siemens Corporate Research of Princeton, NJ.

References

[1] T. Berger and R.W. Yeung, "Optimum 1-ended binary prefix codes," *IEEE Trans. Inforamtion Theory*, vol. IT-36(6), pp. 1435-1441, Nov., 1990.

[2] R.M. Capocelli, A.A.D. Santis, L. Gargano, and U. Vaccaro, "On the construction of statistically synchronisable codes," *IEEE Trans. Information Theory*, vol. IT-38(2), pp. 407-414, Mar., 1992.

[3] W.-M. Lam and A.R. Reibman, "Self-synchronising variable-length codes for image transmission," *Proc. ICASSP, vol 3*, pp. 477-480, Mar. 1992.

Arithmetic Code-like Variable-to-Variable Length Data Compression Code with a Fidelity Criterion for Binary IID Sources

Hisashi Suzuki
Dept. of Mechanical System Engineering
Kyushu Institute of Technology
Kawazu 680-4, Iizuka, Fukuoka 520, Japan

Suguru Arimoto
Dept. of Mathematical Engineering and Information Physics
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan

We consider an IID source that generates binary data with a distribution P on $\{0, 1\}$. $H(P)$ denotes the entropy function for P. Δ denotes an acceptable distortion that is a real constant in $[0, 1]$. W denotes a conditional probability on $\{0, 1\}\times\{0, 1\}$. $R(P, \Delta)$ denotes the rate function for P and Δ with Hamming distortion d that is measured by the normalized Hamming distance.

The proposed code system [1] encodes a binary message of variable length (VL) into the pair of 1) a binary VL codeword and 2) a binary VL sequence called a side information. A sequence concatenated codewords and another sequence concatenated side informations can be mixed in time sharing and be transmitted to a decoder as a single binary sequence.

Encoder

We suppose that P and W, where $W(1|0)+W(0|1)>1/2$ so that $m\geq 2$ may be guaranteed in $(*)$, are given beforehand. A counter c stores the number of the source data already processed by the encoder. Each of variables A and C can assume reals in $[0, 1]$. The encoder, after initialization

$c \leftarrow 1,\ A \leftarrow 1,\ C \leftarrow 0,$

begins the following Recursion.

Recursion: The encoder puts a source data x_c in. When $x_c=0$, one of the following three cases α_0, β_0 and γ_0 can occur.

In case α_0 of $[C,\ C+A)\subseteq[0,\ W(0|0))$, if $[C,\ C+A)\not\subseteq[0,\ W(0|1))$, then the processing, after update

$c \leftarrow c+1,\ A \leftarrow A\times PW(0)/W(0|0),\ C \leftarrow C\times PW(0)/W(0|0),$

reenters into Recursion. Otherwise, the processing escapes from Recursion and enters into Termination.

In case β_0 of $[C, C+A)\subseteq[W(0|0), 1)$, the processing escapes from Recursion and enters into Termination.

The third case γ_0 is $[C, C+A)\ni W(0|0)$, which includes α_0 and β_0 half and half. In this case, the encoder reduces $[C, C+A)$ into either one of subintervals $[C, W(0|0))$ and $[W(0|0), C+A)$. Definitely, the encoder selects a random real r uniformly in $[0, 1)$. (Transmission of r to the decoder side is unnecessary.) If $r<(W(0|0)-C)/A$, then $[C, C+A)$ is reduced into the first subinterval $[C, W(0|0))$ as

$A \leftarrow W(0|0)-C,\ C \leftarrow C,$

and the current case γ_0 returns to α_0. Otherwise, $[C, C+A)$ is reduced into the second subinterval $[W(0|0), C+A)$ as

$A \leftarrow C+A-W(0|0),\ C \leftarrow W(0|0),$

and γ_0 returns to β_0.

These cases α_0, β_0 and γ_0 can occur for $x_c=0$. Alternatively, when $x_c=1$, symmetric cases α_1, β_1 and γ_1 can occur (abbreviated).

Termination: If $[C, C+A)\subseteq[0, W(0|1))$, then c, A and C are updated as

$c \leftarrow c+1,\ A \leftarrow A\times PW(0)/W(0|1),\ C \leftarrow C\times PW(0)/W(0|1).$

Otherwise, that is, if $[C, C+A)\subseteq[W(0|0), 1)$, then c, A and C are updated as

$c\leftarrow c+1, A\leftarrow A\times PW(1)/W(1|0), C\leftarrow 1-(1-C)\times PW(1)/W(1|0).$

Next, finding the *minimum* length l such that

$[0.\mathbf{y}, 0.\mathbf{y}+2^{-l})\subseteq[C, C+A) \quad \text{where} \quad 0.\mathbf{y}=\sum_{j=1}^{l} 2^{-j}y_j$

for at least a binary sequence $\mathbf{y}=y_1\cdots y_l$, the encoder puts $\mathbf{y}$ out. This $\mathbf{y}$ is the codeword for a message $\mathbf{x}=x_1\cdots x_k$, where k denotes the final value of c.

The encoder also puts out a side information $\mathbf{z}$ that is composed of i) $\lceil\log_2(\lfloor k/m\rfloor+1)\rceil$-length juxtaposition of 1s where

$m=\lceil 1/(W(1|0)+W(0|1))\rceil,$ $(*)$

ii) a 0, iii) $\lceil\log_2(\lfloor k/m\rfloor+1)\rceil$-bit integer expression of the value of $\lfloor k/m\rfloor$, and iv) $\lceil\log_2 m\rceil$-bit integer expression of the value of $k-m\times\lfloor k/m\rfloor$.

Decoder

The decoder receives from the encoder a concatenation of codewords. However, no boundary is there between the first codeword $\mathbf{y}$ and the second codeword $\mathbf{y}'$. Therefore the decoder cannot but reproduce data by using some $\mathbf{y}$-prefixed sequence $\hat{\mathbf{y}}=\mathbf{y}\mathbf{y}'\cdots$ instead of using $\mathbf{y}$ directly.

The decoder also receives a concatenation of side informations. The first side information $\mathbf{z}$ is separable in the following way. 1) Read the length $n+1$ of a segment $1\cdots10$. 2) Regard the n-length sequence ζ prefixed by $1\cdots10$ as the n-bit integer expression of the value of $\lfloor k/m\rfloor$, where m is given by $(*)$. 3) Regard the $\lceil\log_2 m\rceil$-length sequence prefixed by ζ as the $\lceil\log_2 m\rceil$-bit integer expression of the value of $k-m\times\lfloor k/m\rfloor$.

Now, let a counter $\hat{c}$ store the number of the data already reproduced. A variable $\hat{A}$ can assume reals in $[PW(1)/W(1|0)-1,\ PW(0)/W(0|1))$, and another variable $\hat{C}$, in $[1-PW(1)/W(1|0),\ PW(0)/W(0|1))$. The decoder, after initialization

$\hat{c} \leftarrow 0,\ \hat{A} \leftarrow 1,\ \hat{C} \leftarrow 0,$

retraces the Recursion and Termination of encoder in the following way.

Retrace of Termination: In the Termination of encoder, before update of A and C, either $[C, C+A)\subseteq[0, W(0|1))$ or $[C, C+A)\subseteq[W(0|0), 1)$ is true. By updating A and C in the former case, it follows that $[C, C+A)\subseteq[0, PW(0))$, that is, $0.\hat{\mathbf{y}}\in[0.\mathbf{y}, 0.\mathbf{y}+2^{-l})\subseteq[C, C+A)\subseteq[0, PW(0))$. By updating A and C in the latter case, it follows that $[C, C+A)\subseteq[PW(0), 1)$, that is, $0.\hat{\mathbf{y}}\in[0.\mathbf{y}, 0.\mathbf{y}+2^{-l})\subseteq[C, C+A)\subseteq[PW(0), 1)$. These cases can occur alternatively and the following rule for retrace of Termination can discriminate them.

If $0.\hat{\mathbf{y}}\in[0,\ PW(0))$, then the decoder, after update

$\hat{c} \leftarrow \hat{c}+1,\ \hat{A} \leftarrow \hat{A}\times PW(0)/W(0|1),\ \hat{C} \leftarrow \hat{C},$

puts a data $\hat{x}_{\hat{c}}=0$ out. Otherwise, the decoder, after update

$\hat{c}\leftarrow\hat{c}+1,\ \hat{A}\leftarrow\hat{A}\times PW(1)/W(1|0),\ \hat{C}\leftarrow\hat{C}+\hat{A}\times(1-PW(1)/W(1|0)),$

puts $\hat{x}_{\hat{c}}=1$ out.

After execution of either one of the above operations, the decoder retraces the Recursion of encoder in the following way.

Retrace of Recursion: If $0.\hat{\mathbf{y}}\in[\hat{C},\ \hat{C}+\hat{A}\times PW(0))$, then the decoder, after update

$\hat{c} \leftarrow \hat{c}+1,\ \hat{A} \leftarrow \hat{A}\times PW(0)/W(0|0),\ \hat{C} \leftarrow \hat{C},$

puts a data $\hat{x}_{\hat{c}}=0$ out. Otherwise, the decoder, after update

$\hat{c}\leftarrow\hat{c}+1,\ \hat{A}\leftarrow\hat{A}\times PW(1)/W(1|1),\ \hat{C}\leftarrow\hat{C}+\hat{A}\times(1-PW(1)/W(1|1)),$

puts $\hat{x}_{\hat{c}}=1$ out. After execution of either one of the above operations, if the counter $\hat{c}$ is less than k, then the processing reenters into retrace of Recursion. Otherwise, the processing escapes from retrace of Recursion.

After the escape, regard the reverted sequence $\hat{\mathbf{x}}^{-1}=\hat{x}_k\ldots\hat{x}_1$ as the reproduced message that corresponds to the original message $\mathbf{x}=x_1\cdots x_k$. Further, $\mathbf{y}$ is reproducible only from $\hat{x}_k\ldots\hat{x}_1$ without using the random numbers. Hence the decoder can separate uniquely $\mathbf{y}$ from $\hat{\mathbf{y}}$, and can proceed to processing of the next codeword.

Compression Efficiency

We consider an arbitrary W achieving the minimum mutual information $I(P, W)$ on condition that

$$\sum_{(x,y)\in\{0,1\}\times\{0,1\}} P(x)W(y|x)d(x,y) = P(0)W(1|0)+P(1)W(0|1) \leq \Delta.$$

If Δ is small so that the condition

$$\max\left(\frac{PW(0)}{W(0|0)}, \frac{PW(1)}{W(1|1)}\right) < \left(\frac{W(1|0)+W(0|1)}{6}\right)^{\frac{2(W(1|0)+W(0|1))}{1-2(W(1|0)+W(0|1))}}$$

may be satisfied, then the redundancy ρ defined by

$$\frac{\left(\begin{array}{c}\text{the expectation of}\\ \text{codeword length}\end{array}\right)+\left(\begin{array}{c}\text{the expectation of}\\ \text{side information length}\end{array}\right)}{\text{(the expectation of message length)}}$$

is bounded as

$$\rho < R(P,\Delta)+(W(1|0)+W(0|1))\times\left(-\sum_{x\in\{0,1\}}P(x)W(x|x)\log_2\frac{W(x|x)}{W(x|\text{not }x)}+\left\lceil\log_2\left\lceil\frac{1}{W(1|0)+W(0|1)}\right\rceil\right\rceil+2\right),$$

the right hand side of which is close to $R(P, \Delta)$ for small Δ.

Reference

[1] H. Suzuki and S. Arimoto, "Arithmetic Code-Like Variable-to-Vari-able Length Source Code with a Fidelity Criterion for Binary IID Sources," *IEICE Trans. Fundamentals*, vol. E75-A, no. 9, pp. 1148-1158, Sep. 1992.

Failure Detection for Communication Networks Using Finite-State Models and Viterbi Decoding

Ender Ayanoğlu
AT&T Bell Laboratories
101 Crawfords Corner Road 4F-507
Holmdel, NJ 07733-3030, USA

In any communication network, facility failures need to be timely detected so that necessary protection switching functions can be initiated without much delay and loss of customer data. Older transmission systems such as T1 trunks base this decision on a count of the severely errored seconds (SES). When "many" consecutive SES's occur, a facility failure detection scheme concludes that the line is not reliable, and initiates an alarm, which in turn initiates a changeover of the transmission link to spare capacity. For transmission systems that employ a packet format with total or partial cyclic redundancy check (CRC) fields, the information on the line errors is available at the receiver after performing a CRC check. Such transmission systems include many data link layer transmission protocols such as HDLC, SDLC, LAPD, and LAPB, or the physical layer schemes such as the SONET transmission frame, the ATM cell structure, etc. Alternatively, the corrupted flags of a data link layer protocol can be used, or special sampling frames can be transmitted for facility monitoring purposes. For example, in Signaling System No. 7 (SS7), fill-in signal units (FISU's) are transmitted for this purpose over the signaling network. For many transmission functions, in particular for signaling, this detection should be very fast. Existing algorithms for this purpose usually involve some integration so that intermittent SES's or bad CRC's do not cause alarms when the facility is healthy, or conversely, intermittent non-severely errored seconds or good CRC's do not prevent a changeover when the facility is faulty. Usually, this filter is designed with deterministic specifications, without any stochastic modeling of the source.

The deterministic filter designed for tracking the low-frequency trends in SES's or CRC's is a linear and time-invariant system. On the other hand, the underlying process one would like to estimate is a nonstationary or time-varying stochastic process. A linear and time-invariant system would be inferior to an adaptive nonlinear scheme based on a good stochastic model of the source, designed to optimize an objective performance criterion. Even if its parameters are not adjusted adaptively, a stochastic nonlinear source model can greatly improve the performance of the failure detection system. One such model is a finite-state model, or a Markov model. It is possible to generate detailed Markov models, but in its simplest form, the Markov model for the facility consists of a "good state" and a "bad state" where good state refers to a healthy facility, and bad state refers to a failed facility. Associated with each state is a probability of a bad checksum, equal to p for the good state, and equal to $1\text{-}q$ for the bad state. The system is in good state at time k, conditioned that it was in good state at time $k\text{-}1$ with probability $1\text{-}a$, and it is in bad state at time k, conditioned that it was in bad state at time $k\text{-}1$ with probability 1. Typically, $p \ll 1$, $q \ll 1$, $a \ll 1$. Such Markov models, known as hidden Markov models, are used in many fields such as ecology, cryptoanalysis, and most importantly, speech recognition.

Fitting a time series of good and bad checksums, i.e., a sequence of 0's and 1's to this model can be achieved via dynamic programming or the Viterbi algorithm, provided that a meaningful performance criterion is chosen. The most commonly used performance criterion for this purpose is known as the *maximum likelihood* criterion. Based on the observation of a good (G) or bad (B) checksum, the Viterbi algorithm associated with the model above proceeds as follows.

1. $D_G(0) = 0$, $D_B(0) = -\infty$, $k = 1$.

2.

$$D_G(k) = D_G(k\text{-}1) + \log(1\text{-}a) + \begin{cases} \log(1\text{-}p) & \text{if G} \\ \log p & \text{if B} \end{cases}$$

$$D_B(k) = \max[D_G(k\text{-}1) + \log a, D_B(k\text{-}1)] + \begin{cases} \log q & \text{if G} \\ \log(1\text{-}q) & \text{if B} \end{cases}$$

3. If $D_B(k) > D_G(k)$, initiate a changeover. Else, set $k \leftarrow k+1$, go to 2.

The probabilities p and q, and the conditional probability a can be estimated from real data, obtainable from transmission statistics. Alternatively, hidden Markov model training techniques can be used.

The performance evaluation for a failure detection scheme should be based on how long it takes to declare a line failed after an actual failure, as well as how frequently the scheme declares a healthy line failed. We will call the first of these quantities the *detection delay*, T_{det}, and the second, the *probability of false alarm*, P_{FA}. A good failure detection scheme minimizes both of these quantities.

In order to benchmark the performance of the stochastic model, we use the conventional leaky integration scheme known as the *leaky bucket*. In this scheme, there is a counter, which is incremented every time a bad checksum occurs, and decremented whenever the count is positive for every T_L good checksums received. A changeover is initiated when the counter reaches T_B. The values used for T_L and T_B for the SS7 network are 256 and 64, respectively.

Modeling the temporal behavior of a channel at the receiver end as a Markov source is quite common. With this motivation, we simulated the data at the receiver as the output of a Markov source (different than the one used for failure detection). Picking a block size of $B = 16$, we assumed the source to consist of 17 states, each roughly corresponding to the number of errors in the block, from 0 to 16. The transition from one state to its neighbors is governed by a geometric rule, the farther a neighbor is, it is geometrically harder to get to it. States 8–16 are bad states, there is no way the system can go back to the good states 0–7, once it enters one of the bad states. For the observations, we introduce another degree of randomness into the model. The number of errors that the model generates when at state i is a random variable whose mean is i, and whose standard deviation is proportional to i. This model is relatively arbitrary, however, it captures the important features of facility failure. A Markov source is commonly used for modeling receiver data. The block size 16 is chosen as a compromise between performance and complexity. This size is sufficient for this scheme to have better performance than the leaky bucket, but smaller block sizes may have even better performance. We would like the system to exhibit an absorbing state, or a super-state, and we pick states 8–16 for this purpose. The geometrical rule is one way to assign higher transition probabilities to closer neighbors, other possibilities exist, but it is not much likely that this will change the overall result. Making the number of bad checksums at each state random introduces an extra degree of randomness so that the failure detection algorithm does not simply learn to keep track of the number of errors and determine the state of the model, i.e., it hides the underlying model from the observer. By design, we pick the mean of the errors at each state equal to the state. In order to increase the uncertainty at the larger states, we also pick the standard deviation proportional to the state. Other models are possible, but again, this is not expected to change the overall result significantly.

Simulations show that the large values of T_B are associated with small values of the false alarm probability, and conversely, small values of T_B are associated with small values of the detection delay. Although the effect is more minor, small values of T_L can be observed to be associated with a large probability of false alarm probability, and large values with small detection delay. In summary, it is not possible to optimize either T_B or T_L for minimizing both P_{FA} and T_{det}. The best solution seems to be reaching a compromise between P_{FA} and T_{det}. Our simulations showed that a value of $T_B = 64$, and $256 \le T_L \le 32$ yields the best solution, in line with the SS7 standard. On the other hand, all the hidden Markov models studied substantially outperform the leaky bucket, achieving *both* P_{FA} and T_{det} results significantly better than the leaky bucket.

Leaky Bucket: P_{FA}

T_B \ T_L	256	128	64	32	16	8	4
128	0.08	0.08	0.07	0.07	0.06	0.03	0.00
64	0.31	0.31	0.32	0.30	0.29	0.19	0.05
32	0.60	0.60	0.60	0.60	0.58	0.50	0.26
16	0.83	0.84	0.85	0.84	0.81	0.75	0.94
8	0.94	0.93	0.93	0.93	0.92	0.88	0.78

Leaky Bucket: T_{det}

T_B \ T_L	256	128	64	32	16	8	4
128	134	134	134	135	137	145	170
64	63	63	62	62	63	68	80
32	34	34	34	34	35	37	41
16	22	21	21	22	22	23	32
8	17	17	16	17	17	17	17

Finite-State Model			
Initial		Trained	
P_{FA}	T_{det}	P_{FA}	T_{det}
0.12	22	0.06	16

Asymptotic Non-stationary Behavior of Statistical Multiplexing with Multiple Types of Traffic

Qiang Ren and Hisashi Kobayashi
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544

Abstract

The cell loss probability is a major performance factor in designing an ATM (asynchronous transfer mode) network for broadband integrated services of multi-media communications. The dynamic behavior of an ATM network needs to be well understood because of its extremely high speed and diversity of traffic. We analyse the transient buffer overflow probability of a statistical multiplexer with *multiple* types of traffic by taking a spectral representation approach. The joint distribution is obtained in the Laplace transform domain in analytic form. An asymptotic behavior is characterised by simple parameters of what we term the "dominant" type traffic.

Summary

We assume that there are M types of sources, and the traffic of type m is characterized by the arrival of "bursts" with Poisson rate λ_m. The burst length is exponentially distributed with mean $\frac{1}{\beta_m}$, and each burst generates cells at the rate of R_m [cells/sec]. The output link capacity is denoted by C [cells/sec]. To make the system stable, we require $\sum_{m=1}^{M} \frac{\lambda_m R_m}{\beta_m} < C$.

Let $J_m(t)$ be the number of type m burst at time t. The aggregate rate of cell arrivals at the multiplexer is then given by $R(t) = \sum_{m=1}^{M} R_m J_m(t)$. When $R(t)$ exceeds C [cells/sec], all the cells cannot be handled immediately. Let $Q(t)$ denote the number of cells outstanding in the output buffer, and define

$$P_{\mathbf{j}}(t,x) = Prob\{J_m(t) = j_m,\ 1 \le m \le M;\ \text{and } Q(t) \le x\}. \quad (1)$$

Let $\mathbf{P}(t,x)$ be the column vector that consists of all the $P_{\mathbf{j}}(t,x)$. Following [1], we can derive a matrix differential equation for $\mathbf{P}(t,x)$:

$$\frac{\partial \mathbf{P}(t,x)}{\partial t} + \mathcal{D}\frac{\partial \mathbf{P}(t,x)}{\partial x} = \mathcal{M}\mathbf{P}(t,x) \quad (2)$$

where

$$\mathcal{M} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \cdots \oplus \mathcal{M}_M$$
$$\mathcal{D} = \mathcal{R}^{(1)} \oplus \mathcal{R}^{(2)} \oplus \cdots \oplus \mathcal{R}^{(M)} - C \cdot I.$$

Here $\otimes$ and $\oplus$ represent Kronecker product and Kronecker sum, respectively, and I is the identity matrix of infinite dimension. And

$$\mathcal{M}_m = \begin{bmatrix} -\lambda_m & \beta_m & 0 & \cdots \\ \lambda_m & -(\lambda_m+\beta_m) & 2\beta_m & \cdots \\ 0 & \lambda_m & -(\lambda_m+2\beta_m) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\mathcal{R}^{(m)} = \text{diag}[0, R_m, \cdots, j_m R_m, \cdots].$$

In order to solve Eq.(2), we first take the double Laplace transform $(t,x) \leftrightarrow (s,u)$ on $\mathbf{P}(t,x)$, i.e., $\mathbf{P}(t,x) \leftrightarrow \mathbf{P}^{**}(s,u)$, and use $\mathbf{P}^{*}(s,0)$ and $\mathbf{P}^{*}(0,u)$ to denote the Laplace transforms of $\mathbf{P}(t,0)$ and $\mathbf{P}(0,x)$, respectively. Equation (2) then becomes

$$\mathbf{P}^{**}(s,u) = (u\mathcal{D} + sI - \mathcal{M})^{-1}[\mathbf{P}^{*}(0,u) + \mathcal{D}\mathbf{P}^{*}(s,0)]. \quad (3)$$

Let us solve the eigenvalues with respect to u, i.e., $u\mathcal{D}V(s) = (\mathcal{M} - sI)V(s)$, and let $V(s)$ and $V(\mathbf{z};s)$ be the corresponding right eigenvector and its generating function. we assume that $V(\mathbf{z};s)$ can be decomposed as $V(\mathbf{z};s) = \prod_{m=1}^{M} V_m(z_m;s)$, then it follows

$$\sum_{m=1}^{M}[(uR_m + \beta_m)z_m - \beta_m]\frac{\partial}{\partial z_m}\{lnV_m(z_m;s)\} = -s + uC + \sum_{m=1}^{M}\lambda_m(z_m - 1). \quad (4)$$

The solution of Eq.(4) should have the following form:

$$V_{\mathbf{k},m}(z_m;s) = exp\{\frac{\lambda_m z_m}{u(s)R_m + \beta_m}\}[\beta_m - (u(s)R_m + \beta_m)z_m]^{k_m}, \quad (5)$$

where k_m is the mth element of an integer vector $\mathbf{k} = [k_1, \cdots, k_M]$: $k_m = 0, 1, 2, \cdots$ and $m = 1, 2, \cdots, M$.

The eigenvalue $u(s)$, involved in the above equation, is given by solving

$$\sum_{m=1}^{M} \frac{k_m(uR_m + \beta_m)^2 + u\lambda_m R_m}{uR_m + \beta_m} = uC - s. \quad (6)$$

If we denote $u_{\mathbf{k}}(s)$ the eigenvalue for the integer vector $\mathbf{k}$, and let $V_{\mathbf{k}}(s)$ and $U_{\mathbf{k}}(s)$ be the corresponding right and left eigenvectors, respectively, then they can be represented as

$$V_{\mathbf{k}}(s) = V_{k_1}(s) \otimes \cdots \otimes V_{k_M}(s)$$
$$U_{\mathbf{k}}(s) = (\mathcal{Q}^{-1})^2 V_{\mathbf{k}}(s)$$

with

$$U_{\mathbf{l}}'(s)\mathcal{D}V_{\mathbf{k}}(s) = \delta_{\mathbf{lk}} \quad (7)$$

and $\mathcal{Q}$ is a diagonal matrix of infinite dimension given by

$$\mathcal{Q} = \mathcal{Q}_1 \otimes \cdots \otimes \mathcal{Q}_M, \quad \mathcal{Q}_m = \text{diag}[1, \cdots, \sqrt{(\tfrac{\lambda_m}{\beta_m})^j \tfrac{1}{j!}}, \cdots]$$

It can be shown from [4] that the number of the positive eigenvalues, denoted by $u_{\mathbf{k}}^{+}(s)$ and derived from Eq.(6), is equal to the number of $\mathbf{k}$'s that satisfy

$$\sum_{k=1}^{M} R_m k_m < C \quad (8)$$

Thus, the unknown transient boundary condition $\mathbf{P}^{*}(s,0)$ can be determined by a set of linear constraint equations

$$U_{\mathbf{k}}'(s)[\mathbf{P}^{*}(0, u_{\mathbf{k}}^{+}(s)) + \mathcal{D}\mathbf{P}^{*}(s,0)] = 0 \quad (9)$$

Since the dimension of this matrix equation is infinite, we should be interested in the most dominant (largest negative) eigenvalue $u_{dom}(s)$, which is obtained by setting $\mathbf{k} = 0$ in Eq.(6). This dominant term will be of practical importance when we consider an asymptotic buffer behavior, i.e., when x is large enough. The transient probability that the buffer content $Q(t)$ exceeds some predetermined buffer capacity B [cells] is approximately given, for larger B, by

$$G^{*}(s,B) \stackrel{def}{=} \mathcal{L}_t\{Prob\{Q(t) > B\}\} \approx b(s)exp\{u_{dom}(s)B\} \quad (10)$$

where $b(s) = -U_0'(s)[\mathbf{P}^{*}(0, u_{dom}(s)) + \mathcal{D}\mathbf{P}^{*}(s,0)]V_0(1;s)$.

We can show that the dominant eigenvalue $u_{dom}(s)$ lies between $\max_m\{-\frac{\beta_m}{R_m}\}$ and 0 for all $s \ge 0$.

References

[1] Kobayashi, H. and Q. Ren [1992], "Non-stationary Behavior of Statistical Multiplexing for Multiple Types of Traffic", *Proc. Twenty-Sixth Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, N.J., March 18-20, 1992.

[2] Kobayashi, H. [1991], "A Spectral Representation Approach to Statistical Multiplexing of Multiple Types of Traffic", *Proc. 1991 IEEE International IT Symposium* , p.156, Budapest, Hungary.

[3] Kosten, L. [1984], "Stochastic Theory of Data handling Systems with Group of Multiple Sources". In H. Rudin and W. Bux (eds.), *Performance of Computer-Communication Systems*, pp. 321-331. North-Holland Publishing Co.

[4] Lancaster, P. and M. Tismenetsky [1985], *The Theory of Matrices*, Academic Press.

THE THROUGHPUT REGION OF NETWORKS WITH TIME-VARYING TOPOLOGY

Leandros Tassiulas

Department of Electrical Engineering

Polytechnic University

6 Metrotech Center, Brooklyn, NY 11201

Abstract

A communication network with time-varying topology is considered and the region of achievable throughputs under any transmission control strategy is characterized. The topology of the network is arbitrary. The topological property that varies with time is the connectivity and/or the capacity of the links. An underlying network state process is considered that reflects the physical characteristics of the network that affect the link transmission capacity. The capacities of the links is a function of the state process which has Markovian statistics. The transmissions are scheduled dynamically based on information about the link capacities and the backlog in the network. The maximum achievable throughput is characterized and a scheduling policy that obtains it is specified. The model of changing topology that is considered here applies to TDMA and CDMA networks with mobile users and networks with meteor-burst communication channels.

Summary

The network model consists of N transmitters and M receivers. Each transmitter may transmit to every receiver. The transmission of transmitter i to receiver j at slot t is successful with some probability $Q_{ij}(t)$. A network with arbitrary topology can be represented with the above model. The transmitters and the receivers correspond to the network nodes and the connectivity is mapped in the probabilities of successful transmissions. If there is no communication link from transmitter i to receiver j then $Q_{ij}(t) = 0$.

The time varying topology is represented by the variation with time of the probabilities of successful transmission $Q_{ij}(t)$. These probabilities depend on certain physical characteristics of the network that change with time. In addition these probabilities depend on which transmitters attempt transmission towards which receivers at each slot. The physical characteristics of the network that affect the probabilities of successful transmission are captured by the *underlying network state variable* $\mathbf{s}(t)$ which takes values in the set $\mathcal{S} = \{1,..,L\}$. In the case of a network of mobile nodes, the underlying network state denotes the geographical position of the nodes, while in a meteor-burst network the state denotes the existence or absence of meteor-bursts for the various links. Each transmitter at any slot may attempt to transmit to one of the receivers or to idle. The transmission attempts at slot t are denoted by the binary transmission vector $\mathbf{R}(t) = (R_{ij}(t) : i = 1,..,N \; j = 1,..,M)$ where $R_{ij}(t)$ is equal to 1 if transmitter i attempts transmission to receiver j at that slot and 0 otherwise. Let $Q_{ij} : \mathcal{S} \times \{0,1\}^{MN} \rightarrow [0,1]$ be the function that determines the probability of success in the transmission from i to j at t based on $\mathbf{R}(t)$, $\mathbf{s}(t)$; that is $Q_{ij}(t) = Q_{ij}(\mathbf{s}(t), \mathbf{R}(t))$.

If the number of packets $X_{ij}(t)$ in transmitter i with destination the receiver j at the end of slot t is nonzero then a packet is transmitted successfully to j in slot $t+1$ with probability $Q_{ij}(t+1)$ and independently of the past. At transmitter i, $A_{ij}(t)$ packets are generated with destination the receiver j during slot t. The arrival processes are Markov modulated. The transmission vector $\mathbf{R}(t)$ is determined according to some transmission scheduling policy. In this work we characterize the *throughput region* of the time-varying network. That is the set of arrival rates $a_{ij} = E[A_{ij}(t)]$, $i = 1,..,N$, $j = 1,..,M$ for which the system is stable under some scheduling policy where the network is defined to be stable if the expectation of the total number of packets in the system is uniformly bounded. The underlying network state process is assumed to be a finite state space irreducible Markov chain. The probability of state s_l under the stationary distribution is denoted by $p^s(s_l)$. The two main results are the following.

Theorem 1: The necessary and sufficient condition for a vector $\mathbf{a} = (a_{ij} : i = 1,..,N, \; j = 1,..,M)$ to belong to the throughput region of the system is that there exist nonegative numbers c_{lm}, $l = 1,..,L$, $m = 1,..,2^{NM}$ such that

$$\sum_{m=1}^{2^{NM}} c_{lm} < 1, \; l = 1,..,L$$

for which we can express the arrival rate vector as

$$\mathbf{a} = \sum_{l=1}^{L} p^s(s_l) \sum_{m=1}^{2^{NM}} c_{lm} Q(s_l, \mathbf{r}^m)$$

where $\mathbf{r}^m$, $m = 1,..,2^{NM}$ are all the binary vectors with NM elements and $Q(s_l, \mathbf{r}^m) = (Q_{ij}(s_l, \mathbf{r}^m) : i = 1,..,N, \; j = 1,..,M)$.

Theorem 2: The policy that schedules at slot t the transmission vector

$$\mathbf{R}(t) = \arg \max_{\mathbf{r} \in \{0,1\}^{NM}} \sum_{i=1}^{N} \sum_{j=1}^{M} Q_{ij}(S(t), \mathbf{r}) X_{ij}(t)$$

stabilizes the network under the necessary and sufficient condition of theorem 1.

The necessity in theorem 1 follows from the fact that in stable mode the long time average number of packets successfully transmitted equals to the arrival rate. The sufficiency in theorem 1 is proved by showing that under the policy of theorem 2 the system is stable when the condition of theorem 1 holds. The state of the system is represented by the vector of packet backlogs in the nodes and the underlying topology state. First it is shown that the drift of a quadratic function of the backlog when it is averaged by the underlying topology state stationary distribution is negative. Then it is shown that if for a multidimensional Markov chain with infinite and finite valued components, the drift of a Liapunov function of the infinite valued components is negative when averaged by the stationary distribution of the finite valued components then the chain is ergodic.

Loss Probability Approximation for General Stationary Input Traffic

Kenji NAKAGAWA, Nagaoka University of Technology

1 Introduction

Queueing problem is investigated for a very wide class of input traffic models and a good loss probability approximation is obtained.

We first consider ATM (Asynchronous Transfer Mode) queueing, i.e., G/D/1, and then extend the approximating method to general queueing problems. The only essential assumption is the stationarity of customer's arrival process and the service process.

2 Preliminaries

We first consider the discrete time ATM queueing system. Every cell which arrives at a multiplexer is served by a single server with constant service time by FIFO (First-In-First-Out) discipline. The multiplexer is assumed to have a buffer of infinite length. The unit of time is taken to be the service time of one cell.

Let us denote by a_t the number of cells which arrives at t th time slot, and Q_t the queue length at the end of t th time slot. We assume $Q_{t_0} = 0$ for some t_0. The queue length Q_t satisfies the well-known recursion formula $Q_t = \max(Q_{t-1} - 1, 0) + a_t$ by Lindley [1]. By solving this recursion, we have the direct expression of Q_t [1]:

$$Q_t = \max_{0 \le i < t - t_0} \Big(\sum_{j=t-i}^{t} a_j - i \Big). \tag{1}$$

Now, we assume that the arriving cell number $\{a_t\}$ is a stationary process. Let the initial time t_0 tend to $-\infty$ and denote by Q the stationary queue length, and N_i the stationary number of arriving cells in an i-interval. (Here, we call an interval of length i an i-interval.) Thus, by (1), the queue length Q is written as $Q = \max_{i \ge 0}(N_{i+1} - i)$, and the cell loss probability $P[Q > q]$ is given by

$$P[Q > q] = P[\max_{i \ge 1}(N_i - i) \ge q]. \tag{2}$$

In general, however, it is difficult to calculate the exact value of the right-hand side of (2). The purpose of this paper is to give a good approximation of (2).

3 Loss Probability Upper Bounds

It is readily seen from (2) that the following inequality holds.

Lemma 1

$$P[Q > q] \le \sum_{i \ge 1} P[N_i \ge i + q]. \tag{3}$$

Each term $P[N_i \ge i + q]$ in the right-hand side of (3) is approximated with the aid of the Chernoff bound technique.

Lemma 2 (Chernoff bound) Let N be a random variable taking on non-negative integral values and $\Psi(z)$ the probability generating function (PGF) of N. Then $P[N \ge r] \le \alpha^{-r}\Psi(\alpha)$ holds for any integer r and any real number $\alpha \ge 1$.

By applying Lemma 2 to (3), we have

Theorem 1 Let $\Psi_i(z)$ be the PGF of the N_i, $i = 1, 2, \cdots$. Then, we have $P[Q > q] \le u(q)$, where $u(q) = \sum_{i \ge 1} \alpha_i^{-(i+q)}\Psi_i(\alpha_i)$, and α_i is the number that minimizes $\alpha^{-(i+q)}\Psi_i(\alpha)$, $\alpha \ge 1$, $i = 1, 2, \cdots$.

4 Application to Several Input Models

4.1 M/D/1

The PGF $\Psi_i(z)$ in Theorem 1 for Poisson traffic of rate ρ is $\Psi_i(z) = e^{\rho i(z-1)}$, $i = 1, 2, \cdots$. Hence, we have the upper bound $u(q)$ of the M/D/1 queueing system as $u(q) = \sum_{i \ge 1} (\frac{\rho i}{i+q})^{i+q} e^{i+q-\rho i}$ for any integer $q \ge 0$.

4.2 AR(1)/D/1

Suppose the input process $\{a_t\}$ is represented approximately as $a_t = \rho + \xi_t$, $\xi_t = b\xi_{t-1} + \epsilon_t$, where b is a constant, $|b| < 1$, and $\{\epsilon_t\}$ are i.i.d. Gaussian random variables such that $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We have the upper bound $u(q)$ for AR(1)/D/1 as $u(q) = \sum_{i \ge 1} e^{-(q+i-\rho i)^2/2\sigma_i^2}$, where $\sigma_i^2 = \frac{\sigma^2}{1-b^2}\frac{1}{(1-b)^2}((1-b^2)i - 2b(1-b^i))$. (we omit details)

5 General Input and Service Time

The above idea can be extended to obtain upper bounds in the cases where the cell inter-arrival time distribution is specified in the ATM queueing, and more generally, extended to G/G/1 queueing.

5.1 Inter-arrival Time Distribution

Let w_n be the waiting time of the n th cell, τ_n the arrival time of the n th cell. The unit of time is, of course, taken to be the service time of the server. Denote by $t_n = \tau_n - \tau_{n-1}$ the inter-arrival time between the n th and $n-1$ th cells. Assume that the random variables $\{t_n\}$ are mutually independent and identically distributed. Let $p(k) = \text{Prob}[t_n = k]$, $k = 0, 1, \cdots$, denote the probability distribution of t_n and $\Psi(z)$ the PGF of t_n. Lindley's recursion $w_{n+1} = \max(0, w_n + 1 - t_{n+1})$ leads to the direct expression $w = \sup_{n \ge 0} V_n$, where $w = \lim_{n \to \infty} w_n$ and $V_n = \sum_{i=0}^{n-1}(1 - t_{i+1}) = n - \sum_{i=1}^{n} t_i$, $n > 0$, $V_0 = 0$. We have an upper bound of the cell loss probability as $P[w > q] \le \sum_{n > q} \alpha_n^{n-q-1}\Psi^n(\alpha_n^{-1})$, where α_n is the number that minimizes $\alpha^{n-q-1}\Psi^n(\alpha^{-1})$, $\alpha \ge 1$.

5.2 G/G/1

We consider an extension of our method to general input and service time distribution. No specific stochastic nature of the input traffic and the service time are assumed except for stationarity.

Let us denote by Q, N_i and L_i the stationary number of packets in the queue, that of arriving packets during an i-interval and that of packets served during an i-interval, respectively. Further, denote by $\Psi_{N_i}(z)$ and $\Psi_{L_i}(z)$ the PGF's of N_i and L_i, respectively. Then we have $P[Q > q] \le \sum_{i \ge 1} \alpha_i^{-(q+1)}\Psi_{N_i}(\alpha_i)\Psi_{L_{i-1}}(\alpha_i^{-1})$, where α_i is the number that minimizes $\alpha^{-(q+1)}\Psi_{N_i}(\alpha)\Psi_{L_{i-1}}(\alpha^{-1})$, $\alpha \ge 1$.

Furthermore, we can eliminate the independence assumption of the input traffic and the service time if the joint distribution of the input and the service time is given.

6 Heuristic Modification of $u(q)$

From detailed and extensive numerical comparison between $u(q)$ and exact formulas or simulation of loss probability, it seems that $\log P[Q > q]$ is approximated well by $\log u(q)$+constant. Since $P[Q > 0] = \rho$, we modify $u(q)$ to define $\tilde{u}(q) = \frac{\rho}{u(0)}u(q)$, where ρ is the link utilization.

We show the numerical calculation results of $u(q)$ and $\tilde{u}(q)$ in Figures 1-4 to compare them with the exact loss probability or computer simulation.

References

[1] Bhargava,A., et al., IEEE GLOBECOM, 1989, pp.903-907.

[2] Roberts,J.W., et al., IEEE Trans on Com, Feb. 1991, pp.298-303.

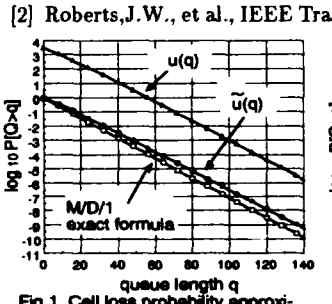


Fig.1 Cell loss probability approximation for Poisson input(load=0.921)

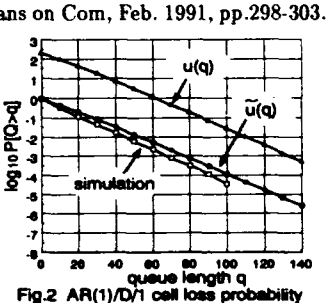


Fig.2 AR(1)/D/1 cell loss probability approximation (load=0.8, σ^2=1, b=0.5)

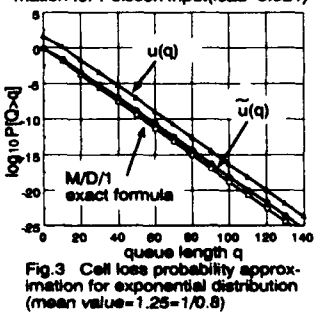


Fig.3 Cell loss probability approximation for exponential distribution (mean value=1.25=1/0.8)

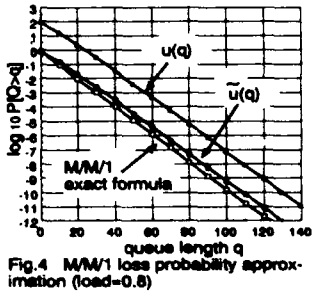


Fig.4 M/M/1 loss probability approximation (load=0.8)

Performance Analysis for Two Manhattan Street Network Routing Algorithms

Zheng Chen and Toby Berger
School of Electrical Engineering
Cornell University
Ithaca, NY 14853, U.S.A.

The MSN is a two-connected, regular network with unidirectional links. The links are arranged in a structure that resembles the one-way system of streets and avenues in midtown Manhattan; a 16-node MSN is shown in Figure 1. It has the same number of connections per node as a bidirectional loop, namely two inputs and two outputs. The node numbered (i, j) belongs to row level ring i and column level ring j.

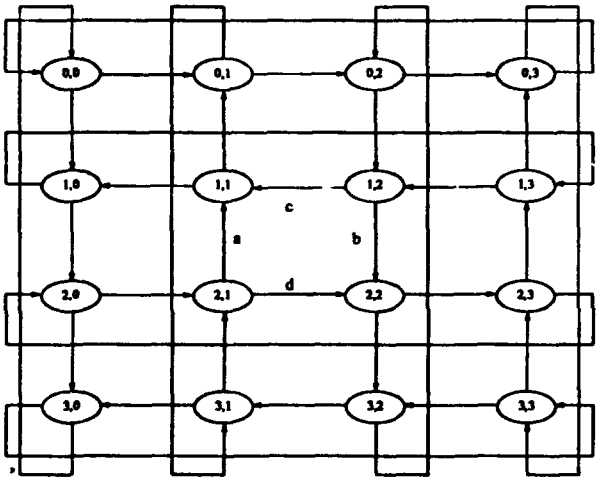


Figure 1: 16-node MSN structure

The choice of the routing procedure in a MSN strongly impacts performance. In particular, the shortest path routing technique minimizes the transmission capacity used by each packet; if the load is balanced among the node pairs, it may be effective in maximizing the throughput of the network. However, it necessitates extra time for a routing table look-up for each packet at each node. For high speed networks (e.g., optical nets) or heavily loaded networks, this can cause long waiting time as many packets queue in buffers at the nodes. Also the routing table at each node needs to be updated when network irregularities occur because of link failures or network expansion. Maxemchuk's routing rules need to decide routes (besides checking the message headers) at each node; some of them are applicable to irregular networks.

We consider two routing algorithms that eliminate the above drawbacks–random routing and a hierarchical deflection routing. The random routing algorithm, selects the output link of every packet randomly at each node. It can be performed fast enough to copy packets using high speed lines. It does not require any knowledge about the current topology of nodes and totally eliminates queuing delay; it is always possible to switch the input packets to the output links without conflicts provided the links have the same speed. Moreover, no memory space is needed at nodes for saving routing tables. For the random routing algorithm we give the theoretical steady state delay and throughput analysis for MSNs via a single node approximation Markov Chain model. A simple iterative formula used for calculating the related distributions is derived, and its accuracy is verified by simulation. Not surprisingly, the network throughput of this random routing is quite low compared to that of the shortest path routing. To address this weakness we propose a hierarchical deflection routing procedure that retains many of the advantages of the random routing (for example, simplicity and no need for topological knowledge) yet achieves an efficient network throughput. We analyze approximate analytical models of this routing algorithm under the conditions of infinite buffers, finite buffers and no buffers at each node and give a simple, iterative formula to calculate the steady state performance parameters. Copious simulations have been done, and the results match well with the theory. We conclude with a comparison of the performances of shortest path, hierarchical and random routing, describing their individual characteristics and how they can be combined for enhanced performance in practical applications.

*This work was supported in part by NSF grants NCR-8903288 and IRI-9005849, and by the K C. Wong Education Foundation in Hong Kong.

SCHEDULING TRANSMISSIONS IN A MULTICAST PACKET SWITCH WHEN CALL SPLITTING IS ALLOWED

by
Charutosh Dixit[1] and Galen Sasaki[2] [3]

Abstract: Multicast packet switching has recently received considerable attention [1, 2, 4, 5]. A multicast packet is one that has a set of destinations; hence, it has applications in multiparty communication (e.g. voice conference calls, video conferencing, and video distribution). The switch considered has N inputs and N outputs, as shown in Figure 1, and transports packets at the inputs that are destined for the outputs. Time is assumed to be slotted and packets at inputs are transmitted at slot boundaries. Packets transmitted in a given slot arrive at all their destinations in the same slot. The switch has the *call splitting* ability. This means that a multicast packet can be duplicated at an input so the destinations of the copies form a partition of the destinations of the original packet. The copies can then be transmitted at different times. The call splitting operation is assumed to require negligible time. Call splitting allows higher throughput. In fact, it has been shown experimentally that it provides near optimal throughput [4]. We consider the problem of finding a schedule of transmissions that delivers a set of multicast packets (or their copies) to their destinations in the minimum number of time slots.

Summary: The scheduling problem will be defined more precisely. It is assumed that there are V packets initially in the switch. These packets are called the *original set*. A set of packets resulting from call splitting none, some, or all of the original set is called a *refinement*. Note that a packet π that has been call split into a set of packets $\pi_1, ..., \pi_n$ has the property that the destinations of $\pi_1, ..., \pi_n$ form a partition of the destinations of π. If Π is a refinement and σ is a mapping from Π to $\{1, 2, ...\}$ then (Π, σ) is called a *schedule*. A schedule (Π, σ) is called *feasible* if for each $\pi_1, \pi_2 \in \Pi$, $\sigma(\pi_1) = \sigma(\pi_2)$ implies that packets π_1 and π_2 do not share a common destination. The *length* of schedule (Π, σ) is $\max_{\pi \in \Pi} \sigma(\pi)$.

Let I_i be the number of packets in the original set at input i; O_i be the number of packets in the original set destined for output i; $I_* := \max_{1 \le i \le N} I_i$; $O_* := \max_{1 \le i \le N} O_i$; and $\tau := \max\{I_*, O_*\}$. Note that τ is a lower bound on the number of slots required to deliver the packets to their destinations.

Multicast Scheduling with Call Splitting (MSCS) Problem. Find a feasible schedule with minimum length.

The problem can be shown to be NP-Complete by modifying a proof used in [2] for another NP-Complete scheduling problem. However, for a restricted class of instances, the problem has polynomial time complexity.

Theorem 1. Consider the MSCS Problem with the following additional conditions: each input has either (*i*) at most one multicast packet or (*ii*) a set of *unicast* packets (i.e., packets with one destination). Then the problem has time complexity $O(N^5)$ and minimum schedule length τ.

The theorem can be proven by transforming the problem into a polynomial scheduling problem described in [3].

The MSCS Problem is NP-Complete but our simulations show that simple scheduling strategies are likely to find schedule lengths of τ. The next theorem is an attempt to explain this under the following probabilistic assumptions.

Assumptions 1. Each packet of the original set is equally likely to be at one of the inputs and these locations are independent of one another. A packet will choose an output as one of its destinations with probability p and the probability is independent of other outputs and packets. Hence, the average number of destinations of a packet is pN.

Theorem 2. There is an Algorithm A that produces a feasible schedule of length τ_A in $O((V+N)N^2)$ time with the following property. Suppose Assumptions 1 are true, $V \ge \frac{c}{\epsilon^2} N \log N$, and $p \ge \frac{c(1+\epsilon)}{\epsilon^2} \frac{\log N}{N}$, where $c \ge 1$ and $\epsilon \in (0, 1]$ are constants. Then $P\left[\frac{\tau_A}{\tau} \ge 1 + f_{\epsilon,N}\right]$ is $O(VN^{-c})$, where $f_{\epsilon,N} = \left(\frac{1+\epsilon}{1-\epsilon/\sqrt{N}}\right)^2 \left(1 + \frac{1}{N}\right)$. (Note that $f_{\epsilon,N} \to (1+\epsilon)^2$ as $N \to \infty$.)

Algorithm A. First, divide the V packets in the original set into I_* subsets, called *bins*. The size of bins are required to be in $\left\{\lfloor \frac{V}{I_*} \rfloor, \lceil \frac{V}{I_*} \rceil\right\}$ and each bin has at most one packet for each input. Second, find a minimum length schedule for each bin. *Greedy* scheduling finds the minimum length schedule because there are at most one packet per input. Finally, concatenate the schedules together to form the final schedule. ■

The proof of Theorem 2 has two parts. First, $P[\tau \le pV(1 - f_{\epsilon,N})]$ is shown to be $O\left(e^{-cN}\right)$ using the Chernoff and union bounds. Finally, $P\left[\tau_A \ge pV\frac{(1+\epsilon)^2(1+1/N)}{1-f_{\epsilon,N}}\right]$ is shown to be $O(VN^{-c})$. This is derived by showing that with probability $1 - O(N^{-c+1})$, $-\frac{V}{N} f_{\epsilon,N} \le I_* - \frac{V}{N} \le \frac{V}{N}\epsilon$, and with probability $1 - O(N^{-c+1})$, a bin has schedule length at most $\frac{1+\epsilon}{1-f_{\epsilon,N}} p(N+1)$. The last two probabilistic results can be derived using the Chernoff bound, the union bound, and Theorem 1.

References

[1] R.P. Bianchini, Jr. and H.S. Kim, "Design of a nonblocking shared memory copy network for ATM," *Proc. of IEEE Infocom '92*, pp. 876-885.

[2] W.T. Chen, P.R. Shen, and J.H. Yu, "Time slot assignments in TDM multicast switching systems," *Proc. of IEEE Infocom '91*, pp. 1296-1305.

[3] R. Jain and G. Sasaki, "Scheduling packet transfers in a class of TDM hierarchical switching systems," *Proc. of ICC '91.*

[4] C. K. Kim and T. Lee, "Call scheduling algorithms in a multicast switch," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 625-635, March 1992.

[5] T.H. Lee and S.J. Lin, "A fair high speed copy network for multicast packet switch," *Proc. of IEEE Infocom '92*, pp.886-894.

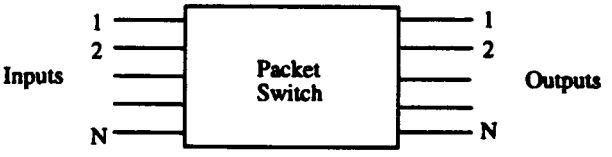


Figure 1

[1]Department of Electrical and Computer Engineering; The University of Texas at Austin; Austin, TX 78712-1084

[2]Department of Electrical Engineering; University of Hawaii at Manoa; 2540 Dole St. Holmes Hall 483; Honolulu, HI 96822; Email: sasaki@spectra.eng.hawaii.edu

[3]Research supported in part by the National Science Foundation under grant NCR-8958556.

MINIMAL STANDARD-PATH SWITCHING NETWORKS

Chris J. Smyth and Liam Halpenny
Department of Mathematics and Statistics, Edinburgh University, JCMB, KB
Mayfield Road, Edinburgh EH9 3JZ, UK

In this paper we investigate n-inlet, n-outlet networks having, at each node, a two-inlet, two-outlet (2×2) switch with two states

possessing the standard-path property. This property is the following:
given any inlet i and outlet j of the network, there is a fixed path through the network, denoted $i \to j$ and called a standard path, from inlet i to outlet j, which is always free (carrying no signal from any other inlet to any other outlet) provided inlet i and outlet j are free, and existing connections have been made using standard paths. We call a network with this property a standard-path network (SPN).

An SPN is therefore a special kind of (wide-sense) non-blocking network: a free inlet can certainly be connected to a free outlet without disturbing existing connections. However, an SPN also has the advantage that a signal can be routed through the network without regard to the network state: only the free inlet and free outlet to be connected need be known. The signal can therefore be self-routing.

We have found SPNs, staged SPNs and staged planar SPNs having the minimal numbers of switches:

THEOREM 1. Any $n \times n$ standard-path network has at least $n^2 - \lfloor \frac{3n}{2} \rfloor$ switches. This number is best possible: there are $n \times n$ SPNs with $n^2 - \lfloor \frac{3n}{2} \rfloor$ switches achieving these lower bounds. (See Figure for $n = 8$).

THEOREM 2. (see also [1]) Any $n \times n$ staged standard-path network has at least n^2-n-1 (n even) and n^2-n (n odd) 2×2 switches. Furthermore, these lower bounds are always achieved.

THEOREM 3. Any $n \times n$ planar staged standard-path network has at least

$$M_n = n^2 - \lfloor \tfrac{n}{2} \rfloor - 2$$

2×2 switches. Furthermore, these numbers M_n of switches are achieved for $n = 2, 3, 6, 7, 10, 11, \ldots$, while for $n = 4, 5, 8, 9, 12, 13, \ldots$, $M_n + 1$ is achieved.

Proofs of the theorems appear in [2].

For an $n \times n$ SPN with inlets $1, 2, \ldots, n$ and outlets $1, 2, \ldots, n$, we choose a typical inlet i, outlet j and standard path $i \to j$. All edges of the network on this path $i \to j$ will be given the label element $i:j$. The label of an edge is the set of all such label elements $\{i:j \mid i \to j$ uses the edge$\}$. For any integers m, ℓ with $1 \leq m, \ell \leq n$, let $\bar{m}$ or $\bar{\ell}$ denote some m- or ℓ-element subset of $\{1, 2, \ldots, n\}$, and $i:\bar{m}$ (respectively $\bar{\ell}:j$) denote the labels $\{i:j \mid j \in \bar{m}\}$ (resp. $\{i:j \mid i \in \bar{\ell}\}$). Then

LEMMA Every edge label in an SPN is either of the form $i:\bar{m}$ or $\bar{\ell}:j$ for some subsets $\bar{\ell}$ or $\bar{m}$ of $\{1, 2, \ldots, n\}$.

One can then show that in any SPN we can divide the switches of the network into five types ($0, \pm 1, 2, 3$) according as to how they alter, on output, the labels on their inlets.

The design of minimal staged networks can be formulated in terms of a game:

The match game and minimal staged networks

We now describe a solitaire game, played on a grid of n^2 points (i,j) $(i, j = 1, 2, \ldots, n)$, with $n(n-1)$ matches. Initially, the matches are all in a vertical position, i.e. with endpoints $\{(i,j), (i,j+1)\}$ $(i = 1, 2, \ldots, n;$ $j = 1, \ldots, n-1)$. The aim is to position all the matches horizontally (i.e. with endpoints $\{(i,j), (i+1,j)\}$ $(i = 1, \ldots, n-1;$ $j = 1, \ldots, n)$) using as few moves as possible.

The allowable moves are as follows:

1. Remove a vertical match, or add a horizontal match.
2. Rotate a vertical match through 90° about one of its endpoints.
3. Replace an adjacent vertical pair of matches by an adjacent horizontal pair of matches with the same four endpoints.

These moves may only be used in positions where no "right angle of matches" (a horizontal match and a vertical match with a common endpoint) is created.

The correspondence between this game and the $n \times n$ staged network is as follows: The point (i,j) represents the label $i:j$, and the connected components of the graph defined by the matches represent the current label sets. Each move corresponds to a switch:

move 1 corresponds to a type 1 or type -1 switch;
move 2 corresponds to a type 2 switch;
move 3 corresponds to a type 3 switch.

There is a corresponding game using coins which corresponds to minimal staged *planar* networks.

[1] H.D.L. Hollmann and J.H. van Lint Jr, "Nonblocking self-routing switching networks", Discrete Applied Mathematics, vol. 37/38, pp.319-340, 1992.
[2] Liam Halpenny and C.J. Smyth, "A classification of minimal standard-path 2 x 2 switching networks, Theoretical Computer Science, vol. 102, pp.329-354, 1992.

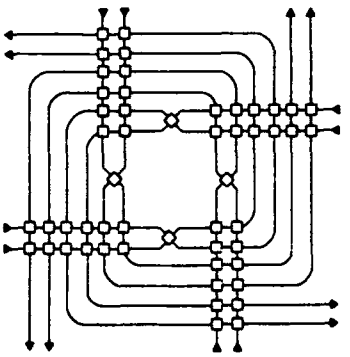


Figure. A minimal 8×8 SPN with 52 switches.

BINARY TREES FOR CLASSIFICATION, REGRESSION, AND CLUSTERING, WITH APPLICATIONS TO LOSSY DATA COMPRESSION

Richard A. Olshen
Division of Biostatistics
Stanford University School of Medicine
Stanford, California 94305-5092

ABSTRACT

This talk is a survey of binary tree-structured methods for classification, regression, survival analysis, and clustering. The discussion will include a survey of unifying themes, together with applications, and an introduction to mathematical issues that arise in studying their asymptotic properties. There will be special emphasis on the CART™ algorithms of Breiman et al., and on applications of the clustering algorithms to predictive, pruned, tree-structured vector quantization (predictive PTSVQ). The talk is a summary of collaborations with many authors over an eighteen year period.

Binary tree-structured statistical methods have found wide applicability in recent years. Areas of application have included computer-aided diagnosis and prognosis in medicine ([1], [5]); speech recognition, ship recognition [1]; prediction in economics and finance; the search for promoter regions in genetics; particle identification in physics; and, perhaps especially for this audience, lossy data compression [3] in digital radiography [2]. There are predictors (features), $\mathbf{X}$, and a response, Y. There is a learning sample $\mathcal{L} = \{(\mathbf{X}_i, Y_i) : i = 1, ..., n\}$, possibly independent and identically distributed, or at least with the $Y_i's$ conditionally independent given the $\mathbf{X}_i's$ (see Chapter 12 of [1]). We use $\mathcal{L}$ to infer an unknown future Y^* from its corresponding known $\mathbf{X}^*$. In some cases we estimate the entire conditional distribution $F(\cdot|\mathbf{X}^* = x^*)$ of Y^*, even when for some $(\mathbf{X}_i, Y_i)$ pairs in $\mathcal{L}$, Y_i is "censored." If the range of Y^* is finite and the goal is to predict its value, then the problem is one of "classification" (or "discrimination"). If Y^* is real, the problem is "regression." If Y^* is real, and the goal is to estimate $P(Y^* \leq y|X^* = x)$, then the problem is "survival analysis." PTSVQ can be viewed as an approach to successive 2-means clustering; $\mathbf{X}^* = Y^*$ is Euclidean, and we want to predict Y^*; but the complexity (bit rate) of the predictor is constrained.

Algorithms involve successively partitioning, that is to say "splitting," the range of $\mathbf{X}$ (the "feature space"). At least when $\mathbf{X} \in \mathcal{R}^k$ the partitioning is by hyperplanes. Results of this "recursive partitioning" can be summarized by a binary tree; $\mathbf{X}^*$ is passed from the root node successively to a terminal node. There is a rule, that typically is constant on each terminal node, by which Y^* is predicted. The rule can be an estimated Bayes rule, as in classification, or a centroid of members of the terminal node, as in PTSVQ. Splitting is always "greedy," one node at a time. In order to obtain some possible benefits of "lookahead," which these algorithms do not have, we grow larger trees than we intend to use and prune them back [1] on the basis of a validation sample; internal validation (typically cross-validation); or, in the case of PTSVQ, a bit rate constraint ([3], [2]).

Versions of these algorithms can be shown to be "consistent" in various senses: Bayes risk consistent for classification, L^P and almost surely consistent for regression, clustering, and survival analysis. See, for example, [1], [4], and [6].

The talk will include a report on applications of PTSVQ to problems of data compression in digital radiography. The studies were undertaken by a group of engineers, radiologists, and statisticians at Stanford University (see [2]).

REFERENCES

[1] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*. Belmont, California: Wadsworth, 1984.

[2] P.C. Cosman, C. Tseng, R.M. Gray, R.A. Olshen, L.E. Moses, H.C. Davidson, C.J. Bergin, and E.A. Riskin, "Tree-structured vector quantization of CT chest scans: Image quality and diagnostic accuracy." Submitted for publication.

[3] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer, 1992.

[4] L. Gordon and R.A. Olshen, "Almost surely consistent nonparametric regression from recursive partitioning schemes," *Journal of Multivariate Analysis*, vol. 15, pp. 147-163, October 1984.

[5] L.W. Kwak, J. Halpern, R.A. Olshen, and S.J. Horning, "Prognostic significance of actual dose intensity in diffuse large-cell lymphoma: Results of a tree-structured survival analysis," *Journal of Clinical Oncology*, vol. 8, pp. 963-977, June 1990.

[6] A.B. Nobel and R.A. Olshen, "Boundedness and consistency of greedy growing for tree-structured vector quantizers," in *Proceedings of the 1993 IEEE International Symposium on Information Theory*, 1993, to appear.

PRIVATE-KEY BURST CORRECTING CODE ENCRYPTION

F. M. R. Alencar
A. M. P. Léo
R. M. Campello de Souza

Communications Research Group - CODEC
Departamento de Eletrônica e Sistemas - UFPE
CP. 7800, 50740-530, Recife - PE, Brasil

Abstract - In this paper, a private-key encryption tecnique is proposed, which makes use of binary linear block burst-error-correcting codes and is based on the fact that such codes have an error control capacity related to bursts which is, in general, larger than its random error control capacity.

Summary

Encryption techniques based on algebraic codes have been proposed for both public and private-key cryptosystems. Specifically, McEliece [1] introduced a public-key cryptosystem based on t-error correcting Goppa codes and in the fact that efficient decoding algorithms exist for such codes while that is not true for a linear code in general. To be effective, the system needs a large blocklength code ($n \sim 10^3$) which is capable of correcting a large number of random errors ($t \sim 40$). Besides that, McEliece's scheme results in a substantial data expansion. More recently, Rao and Nam [2], [3], introduced a private-key encryption similar to McEliece's, the difference being the fact that the code generator matrix was kept secret. This allowed the use of simpler codes, while keeping the security level.

In this paper a private-key cryptosystem is introduced, which makes use of binary linear block burst-error-correcting codes. The main goal is to construct a secure system which employs simple error control codes, based on the fact that the burst-correcting capacity of a code is, in general, larger than its random error correcting capacity.

In what follows, B(n,k,d,b) denotes a binary linear block burst-correcting code with length n, dimension k, minimum Hamming distance d, capable of correcting bursts of length up to b. By a burst of lenght b, it is meant a binary error vector of length n whose nonzero components are confined to b consecutive positions, the first and last of which are nonzero. It is assumed that $b > t$, where $d = 2t+1$.

Denoting by G the generator matrix of B(n,k,d,b), the system designer chooses an n x n permutation matrix P. From theses, the enciphering and deciphering operations related to the plaintext M proceed as follows :

i - Enciphering

$$C = (MG + E_b)P$$

where E_b denotes a burst of length b and Hamming weight > t, randomly generated at the transmitter. M and C (the ciphertext) are binary vectors of length k and n, respectively.

ii - Deciphering

step 1 - Compute

$$C' = C P^T$$

where P^T is the transpose of P, to obtain

$$C' = MG + E_b$$

step 2 - Decode C', i. e., remove E_b via some decoding algorithm. This recovers the plaintex M.

Comments

Considering that the system implementation is simple, it is necessary to focus on questions related to its security. Apparently, the cryptanalist has, at his disposal, at least two main approaches to attack the system, namely

i) To find the matrices G and P from known pairs (M,C).

ii) To recover M, after intercepting C, from chosen pairs (M,C).

The unfeasibility of (i) via any exhaustive search type of procedure is clear, because of the large number of choices for the matrices G and P. Besides that, the known plaintext attack is also difficult to implement, since the solution of the equations related to the column vectors of G (or G' = GP) requires a large number of known pairs and this may be prevented by timely changes on the keys used.

To recover M from C, as suggested in (ii), means, first, to find G' from a sufficient number of chosen pairs and, second, considering

$$M = M_1 \ldots M_k$$

$$C = C_1 \ldots\ldots\ldots C_n$$

$$E_b = e_{b_1} \ldots\ldots\ldots e_{b_n}$$

and

$$G = [g_{ij}], \quad 1 \le i \le k, \quad 1 \le j \le n$$

to solve, for M, the system

$$\begin{aligned} C_1 &= m_1 g_{11} + \ldots + m_k g_{k1} + e_{b_1} \\ C_2 &= m_2 g_{12} + \ldots + m_k g_{k2} + e_{b_n} \\ &\vdots \quad\quad \vdots \quad\quad\quad \vdots \quad\quad \vdots \\ C_n &= m_1 g_{1n} + \ldots + m_k g_{kn} + e_{b_n} \end{aligned}$$

The point here is not only that the solution of the above system requires a computational complexity of the order of k, but also, and perhaps more important, the fact that finding this solution is equivalent to decode, using a t-error-correcting code, a received vector corrupted by an error vector of weight greater than t. Therefore, the system security relies not only on the difficulty of decoding a general linear code, as in the McEliece scheme, but also on the difficulty of correcting a number of errors which is beyond the error-correcting capacity of a given code.

Acknowledgements

This work received partial support from the Brazilian Science and Research Council - CNPq and Banco do Brasil.

References

[1] R.J. McEliece, A Public-Key Cryptosystem Based on Algebraic Coding Theory, DSN Progress Report 42-44, pp. 114-116, Jet Propulsion Laboratory, CA, January and February 1978.

[2] T.R.N. Rao and K. Nam, Private-Key Algebraic Cryptosystems, in Advances in Cryptology-Crypto 86, pp. 35-48, Springer- Verlag, 1986.

[3] T.R.N. Rao and K. Nam, Private-Key Algebraic-Code Encryptions, IEEE Transactions, Vol. IT-35, Nº 4, pp. 829-833, july 1989.

UNIVERSAL HASHING AND UNCONDITIONAL AUTHENTICATION CODES

Tran van Trung
Institute for Experimental Mathematics, University of Essen
Ellernstrasse 29, 4300 Essen 12, Germany

Universal classes of hash functions were introduced by Carter and Wegman [1] and were studied further by Sarwate [2], Wegman and Carter [8] and recently by Stinson [6], [7]. Stinson has found the connections between combinatorial designs and universal hashing. He has proved new lower bounds on the size of universal$_2$ classes of hash functions.

In this paper we study universal$_t$ classes of hash functions for $t \geq 2$. The case $t = 2$ has been investigated in [1], [2], [6], [7], [8]. We present some characterizations of universal$_t$ classes of hash functions in term of combinatorial designs and orthogonal arrays and the application of universal$_t$ classes of hash functions to the construction of authentication codes.

Let A and B be finite sets, where $|A| \geq |B|$. A function $h : A \to B$ will be called a ***hash function***. Let h be a hash function and let $t \geq 2$ be an integer. For a set of t pairwise distinct elements $x_1, \ldots, x_t \in A$, define $\delta_h(x_1, \ldots, x_t) = 1$ if $h(x_1) = \ldots = h(x_t)$, and $\delta_h(x_1, \ldots, x_t) = 0$ otherwise. For a finite set H of hash functions, define $\delta_H(x_1, \ldots, x_t) = \sum_{h \in H} \delta_h(x_1, \ldots, x_t)$. *We now give two definitions of classes of hash functions.*

1. Let ϵ be a positive real number. H is *ϵ-almost universal$_t$* (or $\epsilon - AU_t$) if $\delta_H(x_1, \ldots, x_t) \leq \epsilon |H|$ for all t pairwise distinct elements $x_1, \ldots, x_t \in A$.

2. Let ϵ be a positive real number. H is *ϵ-almost strongly-universal$_t$* (or $\epsilon - ASU_t$) if the following two conditions are satisfied:

(a) for every $x \in A$ and for every $y \in B$, $|\{h \in H : h(x) = y\}| = |H|/|B|$.

(b) for every set of t pairwise distinct elements $x_1, \ldots, x_t \in A$, and for every $y_1, \ldots, y_t \in B$,
$|\{h \in H : h(x_1) = y_1, \ldots, h(x_t) = y_t\}| \leq \epsilon |H|/|B|^{t-1}$.

First, we state a bound on $\delta_H(x_1, \ldots, x_t)$, that is a generalization of Theorem 1.1 [ ].

Theorem 1. *For any class H of hash functions from A to B and for any integer $t \geq 2$, there exist t pairwise distinct elements $x_1, \ldots, x_t \in A$, such that*

$$\delta_H(x_1, \ldots, x_t) \geq |H| b \binom{a/b}{t} / \binom{a}{t},$$

where $a = |A|$ and $b = |B|$.

Two special cases of our definitions are *optimally-universal$_t$* and *strongly-universal$_t$* classes of hash functions, which are defined as follows:

H is *optimally-universal$_t$* (or OU_t) if

$\delta_H(x_1, \ldots, x_t) = |H| b \binom{a/b}{t} / \binom{a}{t}$, for all t pairwise distinct elements $x_1, \ldots, x_t \in A$.

H is *strongly-universal$_t$* (or SU_t) if for every set of t pairwise distinct elements $x_1, \ldots, x_t \in A$ and for every $y_1, \ldots, y_t \in B$,

$$|\{h \in H : h(x_1) = y_1, \ldots, h(x_t) = y_t\}| = |H|/|B|^t.$$

A $t - (v, k, \lambda)$ design is a pair $(\mathcal{X}, \mathcal{B})$, where $\mathcal{X}$ is a set of v elements (called *points*) and $\mathcal{B}$ is a family of k-subsets of $\mathcal{X}$ (called *blocks*) such that every t-subset of $\mathcal{X}$ is contained in exactly λ blocks. A $t - (v, k, \lambda)$ design is *resolvable* if the blocks can be partitioned into $r = \lambda \binom{v-1}{t-1} / \binom{k-1}{t-1}$ parallel classes, each of which consists of v/k blocks that partition the set of points.

An orthogonal array $OA_\lambda(t, n, k)$ is a $\lambda n^t \times k$ array of n symbols such that every set of t columns contains every ordered t-set of symbols exactly λ times.

For descriptions of unconditional authentication codes, we refer to the papers of Simmons and Stinson (see e.g. [3], [4],[5],[7])

Our further results are presented in the following theorems.

Theorem 2. *If there exists a resolvable $t - (v, k, \lambda)$ design, then there exists an OU_t class H of hash functions from A to B, where $|A| = v$, $|B| = v/k$ and $|H| = r = \lambda \binom{v-1}{t-1} / \binom{k-1}{t-1}$. Conversely, if there exists an OU_t class H of hash functions from A to B, where $a = |A|$ and $b = |B|$, then there exists a resolvable $t - (v, k, \lambda)$ design, where v=a, k=a/b and $\lambda = |H| \binom{k-1}{t-1} / \binom{a-1}{t-1}$.*

Theorem 3. *If there is an orthogonal array $OA_\lambda(t, n, k)$, then there exists an SU_t class H of hash functions from A to B, where $|A| = k$, $|B| = n$ and $|H| = \lambda n^t$. Conversely, if there exists an SU_t class H of hash functions from A to B, where $a = |A|$ and $b = |B|$, then there exists an $OA_\lambda(t, n, k)$, where n=b, k=a and $\lambda = |H|/n^t$.*

Theorem 4. *If there exists an $\epsilon - ASU_t$ class H of hash functions from A to B, then there exists an authentication code for $|A|$ source states, $|B|$ authenticators and $|H|$ encoding rules, such that $Pd_0 = 1/|B|$ and $Pd_i \leq \epsilon$, $i = 1, \ldots, t$.*

Using combinatorial designs many families of hash functions in the above theorems have been constructed.

References

[1] J. L. Carter and M. N. Wegman, Universal classes of hash functions, *J. Computer and System Sci.*, **18**(1979), 143-154.

[2] D. V. Sarwate, A note on universal classes of hash functions. *Information Processing Letters*, **10** (1984), 41-45.

[3] G. J. Simmons, Message authentication: a game on hypergraphs, *Congressus Numerantium*, **45** (1984), 161-192.

[4] G. J. Simmons, A survey of information authentication, *Proceedings of the IEEE*, **76** (1988), 603-620.

[5] D. R. Stinson, The combinatorics of authentication and secrecy codes, *J. Cryptology*, **2** (1990), 23-49.

[6] D. R. Stinson, Combinatorial techniques for universal hashing, *preprint*.

[7] D. R. Stinson, Universal hashing and authentication codes, *preprint*.

[8] M. N. Wegman and J. L. Carter, New hash functions and their use in authentication and set equality, *J. Computer and System Sci.*,**22** (1981), 265-279.

THRESHOLD SCHEMES WITH DISENROLLMENT

Bob Blakley
Entry Systems Division
IBM Corporation
Austin, TX 78758

G.R. Blakley
Dept of Mathematics
Texas A&M University
College Station
TX 77843

Agnes Hui Chan
The MITRE Corporation
Bedford MA 01736 and
Northeastern University
Boston, MA 02115

James L. Massey
Swiss Federal Institute
of Technology
Zurich 8092
Switzerland

Abstract

When a shadow of a threshold scheme is publicized, new shadows have to be reconstructed and redistributed in order to maintain the same level of security. In this paper we consider threshold schemes with disenrollment capabilities where the new shadows can be created by broadcasts through a public channel. We establish a lower bound on the size of each shadow in a scheme that allows L disenrollments. We exhibit three systems that achieve the lower bound on shadow size.

Summary

In safeguarding a secret, there are many situations where two or more guardians provide more security than only one. Common examples can be found in safe deposit boxes and in the control of nuclear weapons. In these cases, two keys are needed to activate the control mechanism; the ability to exercise shared control is lost if either key is lost or either key's owner is incapacitated. To guard against such a loss, copies of keys or instructions may be made and distributed to different parties. However, increasing the number of distributed copies increases the risk of some copy being compromised, reducing the security of the system. By distributing "shadows" of a shared secret (which can be used as a key), threshold schemes allow shared control without risking compromise of the secret.

A (t, n) threshold scheme distributes partially redundant shadows $S_1, ..., S_n$ among n users so that any t or more shadows uniquely determine the secret K. Using the entropy function $H(X)$ introduced by Shannon, we have the following definitions.

DEFINITION 1. *A (t, n) threshold scheme is a collection of random variables $(K, S_1, ..., S_n)$ such that for any $1 \leq i_1 < i_2 < ... < i_j \leq n$,*

$$H(K|S_{i_1}, ..., S_{i_j}) = 0 \qquad \forall j \geq t, \tag{1}$$
$$H(K|S_{i_1}, ..., S_{i_j}) > 0 \qquad \forall j < t. \tag{2}$$

Condition (1) says that every set of t or more shadows determines the secret uniquely, whereas condition (2) indicates that the secret cannot be uniquely determined by fewer than t shadows. A (t, n) threshold scheme is said to be *perfect* if

$$H(K|S_{i_1}, ..., S_{i_j}) = H(K) \qquad \forall j < t. \tag{3}$$

Condition (3) says that knowledge of fewer than t shadows does not reduce one's uncertainty about the secret.

The disclosure of a shadow decreases the security against collusion of a threshold scheme since every $t - 1$ remaining shadows, together with the disclosed shadow, determine the secret. Thus, the threshold is reduced from t to $t - 1$. In order to maintain the same threshold t, the key must be changed and the shadows modified. One way to do this is to design a new (t, n) scheme where shadows are then distributed through secure channels. The security of the new system is not compromised if the new shadows are independent of the disclosed shadow. However, setting up the secure channels for distributing shadows can be expensive. This paper considers schemes which distribute modifications to existing shadows through *insecure channels*. Such a scheme is said to have a *disenrolling capability*.

DEFINITION 2. *A (t, n) threshold scheme with L-fold disenrollment capability is a collection of random variables $(K_0, K_1, ..., K_L, S_1, ..., S_n, P_1, ..., P_L)$ such that for each $i, i = 0, ..., L$,*

$$H(K_i|\Delta_i(k), P_1, ..., P_i) \neq 0 \qquad \forall k \geq t, \tag{4}$$
$$H(K_i|\Delta_i(k), P_1, ..., P_i, S_1, ..., S_i) > 0 \qquad \forall k < t, \tag{5}$$

where $\Delta_i(k) = \{S_{i_1}, ..., S_{i_k}\} \subseteq \{S_{i+1}, S_{i+2}, ..., S_n\}$.

In order to minimize the cost of distributing shadows through secure channels, we wish to minimize the number of bits required to encode each shadow. It is conceivable that a (t, n) threshold scheme with higher disenrollment capability requires higher overhead for encoding the shadows. We show that this is indeed the case by establishing a lower bound on the number of bits required to encode a shadow that grows linearly with the number L of disenrollments.

THEOREM. *Let $(K_0, K_1, ..., K_L, S_1, ..., S_n, P_1, ..., P_L)$ be a perfect (t, n) threshold scheme with L-fold disenrollment capability. If $H(K_i) = m$, for $i = 0, ..., L$, then*

$$H(S_j) \geq (L + 1)m \qquad \forall j = 1, ..., n.$$

We consider three examples of optimal threshold schemes with L-fold disenrollment capability, each of which achieves the above lower bound. The Brickell-Stinson scheme [3] makes use of one-time pads, the nonrigid hyperplane scheme [2] is based on geometric properties of hyperplanes and the Martin scheme [5] employs threshold schemes with higher thresholds.

References

1. G.R. Blakley, "Safeguarding Cryptographic Keys," *Proceedings AFIPS 1979 Nat. Computer Conf.* **48** (1979), 313-317.
2. Bob Blakley, G. R. Blakley, A. H. Chan and J. L. Massey, "Threshold Schemes With Disenrollments," *Proceedings of CRYPTO92* (to appear).
3. Brickell and Stinson, *oral communication.*
4. Karnin, Greene and Hellman, "On Secret Sharing Systems," *IEEE Trans. on Information Theory* **IT-29** (1983), 35-41.
5. K. M. Martin, "Untrustworthy Participants in Perfect Secret Sharing Schemes," *preprint.*
6. A. Shamir, "How to Share a Secret," *Communications ACM* **22-11** (1979), 612-613.

Acknowledgement

Agnes Chan's work was supported by MITRE Sponsored Research Program.

AN ADAPTIVE HOMOFONIC ALGORITHM

Christian Gehrmann
Department of Information Theory
Lund University, Box 118
S-221 00 Lund, Sweden

Abstract: Günther gave an algorithm for homofonic coding of messages for cryptographic purpose which was based on known source statistics. In this paper we give an adaptive homofonic algorithm with short delay for a discrete memoryless source with unknown statistics based on Günthers algorithm. We give a formula for the individual redundancy as well as a bound for the max redundancy. Finally a comparison with universal source coding and classical ciphers is made.

Summary— The unicity distance is an important measure for the strength of a cipher. The unicity distance depends on the redundancy and a possible strengthening of a cryptoalgorithm is source coding. Another approach is homofonic coding.

Let $\mathcal{M}$ denote a discrete memoryless source emitting symbols from the alphabet M of $L = \mid M \mid$ letters; $\underline{m}^n = m_1, m_2, ..., m_n$ a sequence of n letters from the alphabet M; $\underline{M}^n$ the set of all possible $\underline{m}^n$; $p(m)$ the probability of occurrence of the letter $m \in M$ of the source $\mathcal{M}$; $\hat{p}(m \mid \underline{m}^{n-1})$ an estimation for the probability of the letter $m \in M$ given the sequence $\underline{m}^{n-1}$; E a way for calculating $\hat{p}(m \mid \underline{m}^{n-1})$. Denote by $R(\underline{m}^n \mid E)$ the total individual redundancy after coding given E, when the sequence $\underline{m}^n$ is coded by the proposed algorithm; $R_n(\underline{m}^n \mid E)$ the per letter redundancy for the same sequence; similar $R(\mathcal{M} \mid E)$ respectively $R_n(\mathcal{M} \mid E)$ the ordinary average redundancy for a source $\mathcal{M}$.

Given an estimation E, the following algorithm realizes a letter by letter adaptive binary homofonic coding for $\mathcal{M}$ according to Günthers algorithm [1].

$s := 0$

1) $i := 1, s := s + 1$, read a new letter $m' \in M$, $\forall m \in M$, calculate $\hat{p}(m|\underline{m}^{s-1})$, and let $\hat{p}^{(0)}(m \mid \underline{m}^{s-1}) := \hat{p}(m|\underline{m}^{s-1})$.

2) $\kappa_i := \lceil -\log(\max_{m \in M} \hat{p}^{(i-1)}(m|\underline{m}^{s-1})) \rceil$.

3) $\forall m \in M, n_m^{(i)} := \lfloor 2^{\kappa_i} p^{(i-1)}(m|\underline{m}^{s-1}) \rfloor$, let the $n_m^{(i)}$ (natural chosen) symbols $\beta_m^{(i,1)}, ..., \beta_m^{(i,n_m^{(i)})}$ represent the letter m.

4) The remaining $n^{(i)} := 2^{\kappa_i} - \sum_{m \in M} n_m^{(i)}$ symbols $\sigma^{(i,1)}, ..., \sigma^{(i,n^{(i)})} \in 2^{\kappa_i}$ are chosen as prefix symbols.

5) Choose a random number $r \in [0, \hat{p}^{(i-1)}(m'|\underline{m}^{s-1})]$.

6) If $2^{\kappa_i} r \leq n_{m'}^{(i)}$, transmit $\beta^{(i,\lceil 2^{\kappa_i} r \rceil)}$ and go to 1), else

7) $\forall m \in M, \hat{p}^{(i)}(m|\underline{m}^{s-1}) := (2^{\kappa_i} \hat{p}^{(i-1)}(m|\underline{m}^{s-1}) - n_m^{(i)})/n^{(i)}$. Transmit $\sigma^{(i,\lceil \frac{1}{\hat{p}^{(i)}(m'|\underline{m}^{s-1})}(2^{\kappa_i} r - n_{m'}^{(i)}) \rceil)}$, $i := i + 1$, go to 2).

Using the results by Jendal, Kuhn and Massey [2] it is possible to give the following theorem.

Theorem: The individual per letter redundancy after coding a sequence $\underline{m}^n$ with the algorithm above given E is:

$$R_n(\underline{m}^n \mid E) = \frac{1}{n} \sum_{s=1}^{n} -\log(\frac{\hat{p}(m_s \mid \underline{m}^{s-1})}{p(m_s)}) \qquad (1)$$

The theorem states that the redundancy for a given source only depends on the choice of E. Hence it is then easy to calculate the average redundancy. Next we select an adequate E. According to results from Shtarkov [3] a natural choice is

$$\hat{p}(m \mid \underline{m}^{s-1}) = \frac{t_m(\underline{m}^{s-1}) + 1/2}{s - 1 + L/2} \qquad (2)$$

where $t_m(\underline{m}^{s-1})$ is the number of occurrences of the letter m in $\underline{m}^{s-1}$. This choice ensures uniform convergence towards zero max redundancy as n grows towards infinity and a proof for the next theorem can be given.

Theorem: For the algorithm the following inequality holds for the maximum redundancy $\max R_n(\mathcal{M} \mid E)$ over all possible sources $\mathcal{M}$ when E is determined by (2):

$$\max R_n(\mathcal{M} \mid E) \leq \frac{L-1}{2n} \log(n) + \frac{c}{n} \qquad (3)$$

where c is a positive constant that is small compared to $\frac{L-1}{2} \log(n)$ for large n.

We analyze the performance of the given algorithm and compare it with results by Davisson [4] for binary memoryless sources and the optimal homofonic algorithm [2]. The algorithm is then analyzed according to Shannon and Hellmans theory for secrecy. Examples are given for classical stream ciphers with key entropy exceeding $\frac{L-1}{2} \log(n) + c$, the minimal key entropy for a unicity distance greater than n.

This work was supported by the TFR grant 222 92-662

References

[1] Ch.G. Günther, "A Universal Algorithm for Homofonic Coding", pp. 405-414 in *Advances in Cryptology - Eurocrypt '88*, Lect. Notes in Comp. Sci. No. 330. New York and Heidelberg: Springer 1988.

[2] H.N. Jendal, Y.J.B. Kuhn and J.L. Massey, "An Information-Theoretic treatment of Homophonic Substitution", pp. 382-394 in *Advances in Cryptology - Eurocrypt '89*, Lect. Notes in Comp. Sci. No. 435. Berlin: Springer 1989.

[3] Y.M. Shtarkov, "Universal Sequential Coding of single message", Problemy Peredachi Informatsii(english trans.), Vol 23, No 3, pp. 3-17, July-Sept., 1987.

[4] L.D. Davisson, R.J. McElice, M.B. Pursley and M.S. Wallace, "Efficient universal noiseless source codes", IEEE Trans. on Information Theory, IT-27, No. 3, 1981, pp.269-279.

LOWER BOUNDS ON THE PROBABILITY OF DECEPTION IN AUTHENTICATION WITH ARBITRATION

Thomas Johansson
Department of Information Theory
Lund University, Box 118
S-221 00 Lund, Sweden

Abstract — Lower bounds on the probability of success for the different kinds of attacks in authentication with arbitration are derived. These bounds give rise to combinatorial lower bounds on the number of encoding rules and on the number of messages necessary in an authentication code with arbitration.

Summary — In the model for normal authentication the transmitter and the receiver are using the same encoding rule and are thus trusting each other. However, it is not always the case that the two communicating parties want to trust each other. Inspired by this problem Simmons has introduced an extended authentication model, here referred to as the authentication model with arbitration, [1]. In this model caution is taken against deception from both outsiders (opponent) and insiders (transmitter and receiver). The model includes a fourth person, called the arbiter. The arbiter has access to all key information and is by *definition* not cheating. The arbiter does not take part in any communication activities on the channel but has to solve disputes between the transmitter and the receiver whenever such occur.

There are essentially five different kinds of attacks to cheat which are possible. The attacks are the following:
I, Impersonation by the opponent. The opponent sends a message to the receiver and succeeds if the message is accepted by the receiver as authentic.
S, Substitution by the opponent. The opponent observes a message that is transmitted and substitutes this message with another. The opponent succeeds if this other message is accepted by the receiver as authentic.
T, Impersonation by the transmitter. The transmitter sends a message to the receiver and denies having sent it. The transmitter succeeds if the message is accepted by the receiver as authentic and if the message is not one of the messages that the transmitter could have generated due to his encoding rule.
$\mathbf{R_0}$, Impersonation by the receiver. The receiver claims to have received a message from the transmitter. The receiver succeeds if the message could have been generated by the transmitter due to his encoding rule.
$\mathbf{R_1}$, Substitution by the receiver. The receiver receives a message from the transmitter but claims to have received another message. The receiver succeeds if this other message could have been generated by the transmitter due to his encoding rule.
In all these possible attacks to cheat it is understood that the cheating person is using an optimal strategy when choosing a message. For each way of cheating, we denote the probability of success with P_I, P_S, P_T, P_{R_0} and P_{R_1}. The overall probability of deception is denoted P_D and is defined to be $P_D = \max(P_I, P_S, P_T, P_{R_0}, P_{R_1})$.

For unconditionally secure authentication codes we derive the following lower bounds on the probability of success for the different kinds of deceptions:

$$P_I \geq 2^{-I(E_R;E_T)+I(E_R;E_T|M)}$$
$$P_S \geq 2^{-I(E_R;E_T|M)}$$
$$P_T \geq 2^{-H(E_R|E_T)}$$
$$P_{R_0} \geq 2^{-I(E_T;M|E_R)}$$
$$P_{R_1} \geq 2^{-H(E_T|M,E_R)}$$

Here E_R is the receiver's encoding rule and E_T is the transmitter's encoding rule. The bounds are valid for all authentication codes with $|S| > 1$ except for a class of degenerate codes which all have $P_{R_0} = 1$ and hence not very interesting.

From the above bounds we also derive lower bounds on the number of encoding rules and on the number of messages to be used in an authentication code with arbitration. Assume that the number of source states for a symmetric source is $|S|$ and let $P_D = 1/q$ for an authentication code with arbitration. Let $\mathcal{E}_\mathcal{R} \circ \mathcal{E}_\mathcal{T}$ denote the set of possible pairs (E_R, E_T). Then the following lower bounds are valid on the number of encoding rules and on the number of messages that are necessary in the code,

$$|\mathcal{E}_\mathcal{R}| \geq q^3$$
$$|\mathcal{E}_\mathcal{T}| \geq q^4$$
$$|\mathcal{E}_\mathcal{R} \circ \mathcal{E}_\mathcal{T}| \geq q^5$$
$$|\mathcal{M}| \geq q^2|\mathcal{S}|.$$

Using these combinatorial lower bounds it is for example possible to show that the cartesian product construction for authentication codes with arbitration does not meet all lower bounds with equality, [1].

References

[1] G. Simmons, "A Cartesian Product Construction for Unconditionally Secure Authentication Codes that Permit Arbitration", *Journal of Cryptology*, Vol. 2, no 2, 1990, pp. 77-104.

This work was supported by the TFR grant 222 92-662

ON THE SPECIFICATION OF PERMUTATIONS FOR BLOCK CIPHERS

Peter Mathys
Department of Electrical and Computer Engineering
Campus Box 425
University of Colorado
Boulder, CO 80309-0425, USA

One way to encrypt data for secrecy is to use a block cipher which maps ℓ q-ary message symbols into ℓ q-ary cipher symbols, i.e., $\underline{c}=s_k(\underline{m})$, where $\underline{m}$ denotes a message block of length ℓ, $\underline{c}$ denotes a ciphertext block of length ℓ and s_k denotes a one-to-one transformation under the control of a secret key k. Without loss of generality s_k can be regarded as an element of the set of permutations on q^ℓ objects. There are $q^\ell!$ such permutations and in practice s_k needs to be restricted to a subset of all permutations to obtain a manageable keyspace size. The "art" of designing block ciphers then is to find a simple way to specify permutations from a subset of all possible permutations, without simplifying the job of the cryptanalyst which tries to break the system without knowledge of the secret key k.

A straightforward way to specify any permutation of q^ℓ objects is to specify the vector $\underline{s}=(s(0),s(1),...,s(q^\ell-1))$, which describes the effect of the permutation s(.) for each input value. In this case $\underline{s}$ is the secret key, but for any practical values of q^ℓ this key is impractically large (e.g., $2^{70}\approx10^{21}$ bits for $q^\ell=2^{64}$). Another way to specify any permutation if $q^\ell>2$ is a prime or a prime power is in the form of a polynomial, i.e., $s(x)=s_t x^t+...+s_1x+s_0$, where $t=q^\ell-2$ and $s_i\in GF(q^\ell)$. Here the key is the set of coefficients (s_i) and it is readily seen that this description is as cumbersome as the $\underline{s}$ vector given above. However, any permutation which can be written in polynomial form can also be obtained from a series of elementary building-block permutations, each of which is easy to specify and to compute.

Our new method of specifying permutations over $GF(q^\ell)$ is based on the function $f(x)=mx^e+c$. It specifies a permutation if $m,c\in GF(q^\ell)$, $m\neq0$, and $\gcd(q^\ell-1,e)=1$. Clearly, f(x) is quite easy to specify and to compute, but by itself offers little security. However, consider a sequence of n building-block permutations $f_i(x)=m_i x^{e_i}+c_i$, i=1,2,..,n, with randomly chosen m_i, e_i, and c_i (satisfying the restrictions on m,e,c given above). Successive application of these building-blocks yields an overall permutation $s(x)=f_n(f_{n-1}(..f_1(x)..))$, which is easy to specify and compute for suitably chosen n and which (based on our simulation results) appears to be undistinguishable from a randomly chosen permutation. In this case the key is given by $k=((m_i,e_i,c_i);\ i=1,..,n)$, and it is easy to see that the size of the key can be adapted to a wide range of needs by varying n. The deciphering function is also very easy to obtain by using the inverse permutations $f(y)^{-1}=((y-c)/m)^{1/e}$, starting with $f_n(y)^{-1}$ and ending with $f_1(y)^{-1}$. Note that the exponent 1/e is computed modulo $q^\ell-1$.

Example: Let $q^\ell=23$ and let $f_1(x)=11x^7+14$, $f_2(x)=8x^{13}+3$, $f_3(x)=2x^9+6$. Then $\underline{s}=(s(0),..,s(22))=(4,14,12,17,1,2,21,19,10,5,18,7,11,20,9,6,0,22,8,3,16,13,15)$ and $s(x)=f_3(f_2(f_1(x)))=2(8(11x^7+14)^{13}+3)^9+6=2x^{21}+7x^{20}+14x^{19}+12x^{18}+7x^{17}+19x^{16}+16x^{15}+12x^{14}+15x^{13}+11x^{11}+9x^{10}+22x^9+22x^8+2x^7+12x^6+5x^5+2x^4+5x^3+19x^2+4x+4$. The key is $\underline{k}=((11,7,14),(8,13,3),(2,9,6))$ and the deciphering function is $f_1(f_2(f_3(y)^{-1})^{-1})^{-1}$, where $f_1(y)^{-1}=(21y+5)^{19}$, $f_2(y)^{-1}=(3y+14)^{17}$, $f_3(y)^{-1}=(12y+20)^5$.

A SOURCE OF CRYPTOGRAPHICALLY STRONG PERMUTATIONS FOR USE IN BLOCK CIPHERS

Lothrop Mittenthal
Teledyne Electronics
649 Lawrence Drive
Newbury Park, CA 91320

This paper suggests a scheme in which cryptographically strong permutations can be randomly selected from a large proper subset of the permutations on blocks of binary numbers which have certain properties of cryptographic strength that are independent of the underlying Boolean functions.

* * *

Let Z_2^n be the group of n-bit binary numbers under coordinatewise addition modulo 2. An orthomorphism is a 1-to-1 mapping R: $Z_2^n \to Z_2^n$ so that $\{x \oplus R(x) \mid x \in Z_2^n\} = Z_2^n$.

Now, let $S(x) = x \oplus R(x)$, then $S(x)$ is a permutation on Z_2^n. $S(x)$ will have a single fixed point, assumed here to be $\Theta = 00 \cdots$ and otherwise maximal. It can be represented as a permutation $(\Theta)(x_1, x_2, \cdots, x_m)$ or a set of $m = 2^n - 1$ equations:

$$\begin{array}{lcll} x & \oplus & R(x) & = & S(x) \\ \Theta & \oplus & \Theta & = & \Theta \\ x_m & \oplus & x_1 & = & z_1 \\ x_1 & \oplus & x_2 & = & z_2 \\ & & \vdots & & \\ x_{m-1} & \oplus & x_m & = & z_m \end{array}$$

Figure 1

where $R(x_{k-1}) = x_k$ and $S(x_{k-1}) = z_k$.

As in any mapping on Z_2^n, the orthomorphic mappings $x \to S(x)$ can be linear, affine, or nonlinear. The linear version is actually linear only at the bit level, but nonlinear at the integer level. It is common to express block substitutions in the form of Boolean functions on $x \in Z_2^n$ where $f_i(x) = 0, 1$ is the value in the ith bit position of the encrypted image of x. Orthomorphic mappings are generated by other means but can be expressed by Boolean functions if desired: the Strict Avalanching Criterion (SAC), the Bit Independence Criterion (BIC), and other desirable properties in a block substitution or permutation depend on the defining Boolean functions. The process of changing such substitutions raises questions as to the strength of the replacement. Because it would be useful to have a class of block substitutions which possess some property of cryptographic strength that is invariant under change of the defining Boolean functions, a class of so-called balanced block substitutions is offered. The definition of balanced block substitutions is based on the fact that Z_2^n is an additive group of order 2^n and has $2^n - 1$ maximal subgroups [2], e.g., the even numbers.

A permutation or block substitution on Z_2^n is said to be balanced if it maps each maximal subgroup half into itself and half into its complement. This can be expressed in terms of Shannon's information theory: If M is any maximal subgroup of the n-bit numbers, for a balanced mapping, $x \to y$, then the uncertainty may be expressed as: $H(x \in M \mid y \in M) = -\log_2 P(x \in M \mid y \in M) = -\log_2 (0.5) = 1$.

A permutation or block substitution on Z_2^n is shown to be balanced if and only if it is an orthomorphism, irrespective of whether it is linear, affine, or nonlinear.

Because an orthomorphic permutation can be described by a set of 2^n equations, an approach is to generate these randomly with constraints, taking advantage of the balance property. However, this is a rather inefficient process.

An alternate method is to construct an orthomorphism which is linear at the bit level and modify it to be nonlinear. In that case, the equations in Figure 1 take the form:

$$x_{k-1} \quad \oplus \quad x_k \quad = \quad x_{k-p} \qquad (1)$$

for some integer p and for all indices k. The permutation $(\Theta)(x_1, x_2, \cdots, x_m)$ represented by the order of the $\{x_{k-1}\}$, $\{x_k\}$, and $\{x_{k-p}\}$ in the set of equations is also the same as that specified by the mapping $x \to R(x)$. Additional orthomorphic permutations are defined by any power s of this permutation. The result is a new orthomorphic linear permutation $S'(x)$ defined by $x \to R^s(x)$ represented by a set of m equations:

$$x_{k-s} \quad \oplus \quad x_k \quad = \quad x_{k-p_s} \qquad (2)$$

The integer p_s is a function of s. This holds for $1 \le s \le 2^n - 2$, so that for one basic orthomorphic permutation, a family of $2^n - 2$ orthomorphic permutations, all linear at the bit level, is generated. This is a transitive group of permutations [3]. This property is also invariant under change of the Boolean functions defining the basic block substitution if it is a linear orthomorphism. Any or all of these can now be converted to nonlinear permutations by suitable modifications to a subset of the equations, or, equivalently, by altering the order in two of the three columns of numbers.

References

1. C. Adams & S. Tavares, "The Structured Design of Cryptographically Good S-Boxes," J. Cryptology (1990) 3:27 – 41.

2. M. Jacobson, Lectures in Abstract Algebra, Vol 1, D., VanNostrand Co., 1964.

3. L. Mittenthal, "Block Substitutions Using Orthomorphic Mappings," Advances in Math (to appear).

A New Bound for Substitution Attack

L. Tombak

Department of Computer Science
University of Wollongong
PO Box 1144, Wollongong 2500
Australia

R. Safavi-Naini

Department of Computer Science
University of Wollongong
PO Box 1144, Wollongong 2500
Australia

Abstract

We define two classes of strategies for substitution attack and derive lower bounds on the probability of deception for each class for codes with perfect protection for impersonation. We show that the equality of the two bounds uniquely determines the number of encoding rules and forces the incidence matrix of the code to be that of a BIBD. It also implies that random selection from the remaining cryptograms gives the same probability of deception to the enemy as random selection from the set of keys that are incident with the intercepted codeword.

1 Preliminaries

We consider an authentication scenario in which a *transmitter* wants to send a message to a *receiver* over a publicly exposed channel and an *enemy* who tries to deceive the receiver in accepting a fraudulent message as genuine. An *authentication code* (A-code) is a collection Z, $|Z| = E$, of mappings from the set X, $|X| = k$, of the source states into the set Y, $|Y| = M$, of codewords. Let Y_z denote the subset of codewords that are authentic under the key $z \in Z$ and Z_y denote the subset of encoding rules that are incident with $y \in Y$. The *incidence matrix* of an A-code is a zero-one matrix of size $E \times M$ in which $a_{zy} = 1$ only if $y \in Y_z$.

The communicants use a probability distribution $\pi = (\pi_1, \cdots, \pi_E)$ on the key space as their *strategy*. In a *substitution attack* the enemy intercepts a codeword and tries to substitute it with a fraudulent one.

2 Lower Bound for Substitution Attack

We consider two possible courses of action (classes of strategies) for the enemy and find bounds on the probability of deception in each case.

Assume the enemy intercepts a cryptogram $v \in Y$. In a class $\mathcal{K}_1$ strategy the enemy chooses a probability distribution p^v on Z which is nonzero only on Z_v. The enemy uses this distribution to select a key $z \in Z_v$ and then randomly chooses a codeword of Y_z for substitution. Let P_0 and P_1 denote the best probability of deception in impersonation and substitution attack respectively.

Proposition 2.1 *If the source is uniform and $P_0 = k/M$ we have*

$$P^{k_1} \geq \frac{M}{kE}. \tag{1}$$

The bound is achieved if p^v is uniform for all v.

In class $\mathcal{M}_1$ strategies the enemy chooses a probability distribution q^v on the reduced cryptogram set $Y \backslash v$ with $q^v_v = 0$, and uses it to choose a cryptogram for substitution. We have [1]

$$P^{m_1} \geq \frac{k-1}{M-1} \tag{2}$$

and the bound is achieved when $q^v = 1/(M-1)$ for all v. Combining the bound 1 and 2 we have theorem 2.1.

Theorem 2.1 *For a uniform source, if $P_0 = k/M$ the probability of deception is lower bounded by*

$$max\,(\frac{M}{kE}, \frac{k-1}{M-1}).$$

The two bound are equal if $E = E_0 = \frac{M(M-1)}{k(k-1)}$ in which case the incidence matrix of the code corresponds to a BIBD.

This result is in accordance with Stinson's [2].

Bounds 1 and 2 are achieved for the random strategies of class $\mathcal{K}_1$ and $\mathcal{M}_1$ respectively. If $E > E_0$ random strategy of class $\mathcal{K}_1$ gives a higher probability of success and if $E < E_0$ random strategy of class $\mathcal{M}_1$ is superior.

References:

1. J.L. Massey, *Cryptography, A Selective Survey*, Digital Communi cations, ed.E. Biglieri and G. Pratti, Elsevier Science Publ., 1986, 3-25.

2. D.R. Stinson, *Combinatorial Characterizations of Authentication Codes*, Proceedings of Crypto '92.

On RSA Signatures

Thijs Veugen
Group on Information and Communication Theory, Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: thijs@ei.ele.tue.nl

When analysing the security of electronic cash systems [1,3] one comes up with a question concerning RSA signatures. This question arises when looking at the cut and choose method of the withdrawal protocol. Previous results [2] cannot be applied. Using some assumptions this problem is solved.

Introduction

A way to protect the privacy of the user in electronic cash systems is to use the cut and choose idea in the withdrawal protocol. At the end of the withdrawal protocol [1,3] the user obtains an RSA root [4] on the product of the numbers that were not opened by the bank. The numbers in the signed product that represents money, are supposed to contain the identity of the user which prevents the user from double-spending his money. Since the bank cannot verify all numbers before signing the user can try to cheat. For the bank it is important to know what kind of money-representing signatures a cheating user can obtain.

This question is formalized next. After that the assumptions that are used to solve this problem are given and finally the results are presented. For more details and proofs see [5].

Formal statement of the problem

Let n be an RSA modulus [4], e and d integers such that $e \cdot d \equiv 1 \pmod{\varphi(n)}$ and C a set of numbers coprime with n. The numbers e and n are public while the number d and the factorization of n are not. The elements of C are images of a one way function F. The sentence "Choose an element a from the domain of F and compute the image $x = F(a)$" is for convenience abbreviated to "Let $x \in C$". Similarly with subsets of C. For subsets W of C the number $\underline{W}$ is defined as the product of all the elements of W modulo n.

In the case of the electronic cash system each element from the domain of F determines the identity of some user.

An honest user doing the withdrawal protocol would choose some set $W \subset C$ with all elements corresponding to his identity. The bank would ask him to open some numbers $x \in W$ and the user would obtain $\underline{X_i}^d$ for some subset X_i of W.

A cheating user however chooses at least one number not containing his identity. Instead, he chooses a number z modulo n in a clever way. So w.l.o.g. he chooses some set $W \subset C$ with all elements corresponding to his identity, and a number z modulo n. The bank then asks him to open some of these numbers and if he is not caught (i.e. the number z is not chosen by the bank), he obtains $(\underline{X_i} \cdot z)^d$ for some subset X_i of W. Since this signature does not represent money he tries to compute $\underline{Y_i}^d$ from it for some subset Y_i of C. Note that a cheating user can allways (try to) obtain a signature on the *opened* numbers by computing $z \equiv \underline{W}^{-1} \pmod{n}$.

The central problem in this paper can now be stated as:

Let $l \geq 1$. Let X_i and Y_i be subsets of C $(i = 1, \ldots, l)$.
Is it feasible to compute, without knowing the factorization of n, a number z coprime with n such that for each $1 \leq i \leq l$ it is feasible to compute $\underline{Y_i}^d$ from $(\underline{X_i} \cdot z)^d$ modulo n?

Assumptions

Four assumptions are made (their interpretation follows below):

Prime: The root e is a fixed prime, at least 5 (for $e = 3$ the results are different).

Subset: The sets X_1 to X_l are not subset-related i.e. there are no two sets X_i and X_j $(i \neq j)$ such that $X_i \subset X_j$.

Rootcomputability: Let x and y be coprime with n. If it is feasible to compute x^d from $\{x, y, y^d\}$ modulo n without knowing the factorization of n, then it is feasible to compute a number $r \in \{0, \ldots, e-1\}$ and a number s coprime with n from $\{x, y\}$ such that $x \equiv y^r s^e \pmod{n}$.

Rootinfeasibility: Let $k \geq 1$ and let x_1 to x_k be k different elements of C. Then it is infeasible to compute numbers $r_1, \ldots, r_k \in \{0, \ldots, e-1\}$ not all zero, and a number s coprime with n such that $x_1^{r_1} \cdot \ldots \cdot x_k^{r_k} \equiv s^e \pmod{n}$.

In the case of the electronic cash system the subset assumption is satisfied because the number of opened numbers is fixed. The rootcomputability assumption means that if an RSA-root is computable from another RSA-root, this computation can be done using only multiplications, divisions and exponentiations. Note that this excludes cases like $x \equiv (\mathrm{DES}(y^d))^e \pmod{n}$, but for randomly chosen x this seems to be a reasonable assumption. The rootinfeasibility assumption means that it is infeasible to compute (non-trivial) e^{th} roots on products of elements of C. The essential restriction on the $r_1, \ldots, r_k$ is that at least one is not zero. Realizing that the numbers in the set C are images of a one way function makes this assumption reasonable.

Results

Using all four assumptions the problem is solved. W.l.o.g. all Y_i are not empty and $l \geq 2$. The sentence "a number z coprime with n such that for each $1 \leq i \leq l$ it is feasible to compute $\underline{Y_i}^d$ from $(\underline{X_i} \cdot z)^d$ modulo n?" is for convenience abbreviated to "such a number z".

Theorem 1. If it is feasible to compute such a number z, then (the sets Y_1 to Y_l are not subset-related) or (there is a number $j \in \{1, \ldots, l\}$ such that the sets Y_i for $i \neq j$ are not subset-related and $Y_j \subseteq Y_i$ for every i).

Theorem 2. Suppose the first case of Theorem 1 is satisfied. Define the set U as the union of all X_i, the set X as the intersection of all X_i and the set Y as the intersection of all Y_i $(1 \leq i \leq l)$. Then it is feasible to compute such a number z if and only if $\forall_{1 \leq i \leq l}[Y_i = (U \setminus X_i) + Y]$ or $\forall_{1 \leq i \leq l}[Y_i = (X_i \setminus X) + Y]$.†

Theorem 3. Suppose the second case of Theorem 1 is satisfied. W.l.o.g. $j = 1$. Define the set U as the union of all X_i, the set X as the intersection of all X_i and the set Y as the intersection of all Y_i $(2 \leq i \leq l)$. Then it is feasible to compute such a number z if and only if $\{\forall_{2 \leq i \leq l}[Y_i = (U \setminus X_i) + Y]$ and $Y = (X_1 \div U)$ and $Y_1 = (U \setminus X_1)\}$ or $\{\forall_{2 \leq i \leq l}[Y_i = (X_i \setminus X) + Y]$ and $Y = (X_1 \div X)$ and $Y_i = (X_1 \setminus X)\}$.†

From Theorem 1 follows that if such a number z is computable, the $Y_i (1 \leq i \leq l)$ are related in only two possible ways. The first possibility is treated in Theorem 2. The second possibility is treated in Theorem 3.

Observation of the proofs shows that if such a number z is computable, it is easy to compute such a number z. Furthermore, in the proofs is not used that the elements of C are images of a one way function, although these numbers have to satisfy the rootinfeasibility-assumption.

When applying the results to an electronic cash system one has to realize that a signature $\underline{Y_i}^d$ can only represent money if the cardinality of Y_i has some specific value. Therefore Theorem 2 can be applied. When translating the results of Theorem 2 to cheating-user-strategies it follows that a cheating user can try to replace some not- opened-numbers that contain his identity (the elements of X) by other numbers that do not contain his identity (the elements of Y). The remaining signed numbers can be either the *opened* numbers or the other not-opened-numbers. Note that a cheating user is caught with probability 0.5 independent of his strategy, although the probability that his strategy succeeds is generally less.

Further research can be done on the area of cheating users who want to combine their obtained signatures to produce other signatures.

Finally I would like to thank David Chaum, Matthijs Coster, Hendrik Jan Evertse, Eugène van Heyst and Henk van Tilborg for their useful comments and discussions.

References

[1] Chaum, D., A. Fiat and M. Naor, Untraceable electronic cash, Advances in Cryptology-CRYPTO '88, S. Goldwasser ed., Springer-Verlag, pp. 319-327.
[2] Evertse, J.H. and E. van Heyst, Which new RSA-signatures can be computed from certain given RSA-signatures?, Journal of Cryptology, Vol. 5, No. 1, 1992, pp. 41-52.
[3] Okamoto, T. and K. Ohta, Universal electronic cash, Advances in Cryptology-CRYPTO '91, J. Feigenbaum ed., Springer-Verlag, pp. 324-337.
[4] Rivest, R.L., A. Shamir and L.Adleman, A method for obtaining digital signatures and public key cryptosystems, Comm. ACM, Vol. 21, February 1978, pp. 120-126.
[5] Veugen, P.J.M., Some mathematical and computational aspects of electronic cash, Master's thesis, Eindhoven University of technology, November 1991, pp. 43-66.

† The + operator denotes the union of two disjunct sets. The symmetrical difference $\div$ of two sets A and B is defined as the union of $A \setminus B$ and $B \setminus A$.

AN ATTACK ON XINMEI'S DIGITAL SIGNATURE SCHEME

Yuan-Xing Li

(P.O.Box 145, Dept. of Information Engineering,
Beijing University of Posts and Telecommunications, Beijing, 100088 P.R.China)

SUMMARY

Xinmei [1] first proposed a digital signature scheme based on error-correcting codes. Later, Harn and Wang [2] modified Xinmei scheme to improve the security and performance of it. Recently, Alabbadi and Wicker [3] cryptanalysed the Xinmei scheme amd H-W's modified scheme, pointed out that under a chosen message attack, the private keys of these schemes can be obtained in polynomial time. Furthermore, in this paper, we show that if the public keys W, H, J, and T designed for the schemes satisfy that the condition: the matrix $[W, H^T]$ or $[J, T]$ is nonsingular, then the private keys are easy to be infered in polynomial time under a known signature attack. Finally, some examples are given to illustrat our conclusions.

1. Descriptions of Xinmei signature scheme:

User A of the scheme chooses an (n, k, d) binary Goppa code C with a $k \times n$ generator matrix G, an $(n-k) \times n$ parity check matrix H, and t-error-correcting capacity. The public keys of the schems are

$$J = P^{-1}G^*S^{-1} = P^{-1}W,$$
$$W = G^*S^{-1},\ T = P^{-1}H^T,\ \text{and H, t, } t_a.$$

While the private keys are SG and P. Where G^* is the matrix which satisfies

$GG^* = I_k$, H^T is the transposed matrix of H.

The signature C_i of a k-bit message M_i is as follows:

$$C_i = (E_i + M_i SG)P.$$

where E_i is an n-bit random vector with the Hamming weight $w(E_i) = t_a < t$ chosen by user a.

2. An attack on the scheme:

If $[W, H^T]$ is full-rank (then $[J, T]$ is full-rank too, vice versa), then

$$P = [W, H^T][J, T]^{-1},$$

the computational complexity of calculating P is $O(n^3)$. The next step is to get the other private key SG, knowing P, under a known-signature attack As

$$C_i = (E_i + M_i SG)P,$$

$$C_i P^{-1} - E_i = M_i SG.$$

Suppose that we know k signature-message-error pattern triplets $\{(C_i, M_i, E_i)\}$, then we can produce a matrix equation:

$$C'^T = [C_1P^{-1} - E_1, \ldots, C_kP^{-1} - E_k]$$
$$= (SG)^T [M_1, \ldots, M_k]$$

If k messages $M_1, \ldots, M_k$ are linearly independent, then

$$(SG)^T = C'^T [M_1, \ldots, M_k]^{-1}.$$

This step can be fulfilled in $O(k^3)$ operations. So the total computation complexity of this attack is $O(n^3)$.

This attack is also effective to Harn-Wang's modified scheme.

To avoid this attack, the scheme designer must pick such public keys W, H, J, and T that they satisfy the requirement: matrix $[W, H^T]$ or $[J, T]$ is not full-rank.

3. Examples (omitted)

This work was supported in part by the China National Nature Science Foundation and the China National Information Security Key Lab Foundation.

REFERENCES

(1) Wang Xinmei: 'Digital signature scheme based on error-correcting codes,' Electronics Letters, 1990, 26(13), pp898-899.
(2) L. Harn, and D-C, Wang: 'Cryptanalysis and modification of digital signature scheme based on error-correcting code,' Electron. Lett., 1992, 28(2), pp157-158.
(3) M. Alabbadi, and S.B. Wicker: 'Security of Xinmei digital signature scheme,' Electron. Lett. 1992, 28(9), pp890-891.

ASYMPTOTIC BOUNDS ON THE RATES OF RUNLENGTH-LIMITED CODES

Shih-Hsuan Yang and Kim A. Winick
Electrical Engineering and Computer Science Department
University of Michigan, Ann Arbor, MI 48109

Abstract

Runlength-limited (RLL) codes are widely used in magnetic and optical recording systems to aid in bit synchronization and reduce the effects of intersymbol interference. Asymptotic lower bounds on the size of these codes as a function of minimum distance have been recently reported by Kolesnik and Krachkovsky. These lower bounds are generalized to include cost-constraints. Asymptotic upper bounds for the size of runlength-limited codes are also investigated, and two separate bounds are presented. Finally, the maximum rate at which information can be transmitted across a noiseless channel, using sequences produced by a nondeterministic graph, is lower bounded. The bound is derived using generating function techniques.

Introduction

In order to minimize the effects of intersymbol interference and aid in bit synchronization, many digital transmission and recording systems restrict the set of allowable binary channel sequences. A commonly used constraint imposes limits on the minimum and maximum number of consecutive zeros, which may follow a "1" in the binary channel stream. Sequences which satisfy these runlength conditions are said to be (d,k) runlength-limited (RLL), where d and k are the minimum and maximum zero runlengths, respectively.

Channels are in general noisy, and it may be desirable to incorporate an error-correcting capability into the communication or storage system at the expense of data rate. The error-correcting capability of a code is a function of the distance distribution between codewords. The minimum distance between any two codewords, d_{min}, is a parameter of particular interest. A fundamental problem in coding theory is to determine the maximum size of a code having a given d_{min}. This problem has been studied extensively for unconstrained codes and remains unsolved, although upper and lower bounds have been derived. Recently, Kolesnik and Krachkovsky have reported an asymptotic lower bound for runlength-limited codes.[1] They obtained their result using a sphere packing type argument combined with a generating function technique. Asymptotic upper bounds for runlength constrained codes, however, have not been reported.

The purpose of this paper is threefold. First, Kolesnik and Krachkovsky's asymptotic lower bound is extended to include costly runlength-limited sequences.[1,2] Second, two asymptotic upper bounds are derived for the maximum size of RLL-codes as a function of the codes minimum distance. Third, the maximum rate at which information can be transmitted across a noiseless channel, using sequences produced by a nondeterminisitc graph, is lower bounded.

Costly RLL-Codes

It is known that (d,k)-RLL sequences can be described by a finite state machine having (k+1)-states It is convenient to represent the finite state machine by a graph with (k+1)-vertices. The edges between the vertices indicate the possible state transitions, and the labels on those edges give the corresponding output bits in the RLL-sequence. Costs can be assigned to RLL-sequences by attaching a cost to each edge in the graph.[2] We show that the maximum rate, R, at which information can be transmitted across a noiseless channel using only those sequences with average cost per bit less than α, and minimum distance greater than δ, is lower bounded by

$$R(\delta, \alpha) \geq R(0, \alpha) - \min_{0<x,y,z\leq 1} [\log_2\lambda_1(x,y,z) - \alpha\log_2(xy) - \delta\log_2(z)]$$

where λ_1 is the largest positive eigenvalue of a $(k+1)^2$ by $(k+1)^2$ transition matrix which can be computed from the graph. It is also shown that $R(0,\alpha)$ is given by

$$R(0, \alpha) = \min_{0<x,y\leq 1} [\log_2\lambda_1(x,y,1) - \alpha\log_2(xy)]$$

and is equal to the maximum entropy of the Markov source generated by assigning transition probabilities to the edges of the graph.

Asymptotic Upper Bounds on the Size of RLL-Codes

Two asymptotic upper bounds are derived for the maximum size of a runlength-limited code as a function of its minimum distance δ. Let $R(\delta)$ equal the maximum rate at which information can be transmitted across a noiseless channel using RLL-codes with minimum distance δ. Runlength-limited codes can divided into constant weight subsets, and the code rate, $R^*(w)$, of the weight w subset is computed using combinatorical arguments. It is shown that

$$R(\delta) \leq \max_{0\leq w\leq 1} \min[R^*(w), R^{**}(\delta, w)]$$

where $R^{**}(\delta,w)$ is the McEliece-Rodemich-Rumsey-Welch linear programming bound for the maximum rate of (unconstrained) constant weight codes having minimum distance δ.[3] A second asymptotic upper bound is also derived from upper bounds on the capacity of input-constrained discrete memoryless channels. The simplest version of this bound yields

$$R(\delta) \leq C(\delta/2)$$

where $C(\delta/2)$ is the capacity of a runlength-limited, input-constrained binary symmetric channel, which has cross-over probability $\delta/2$. A tight upper bound for the right-hand side of Eq. (4) is then obtained using techniques developed by Shamai and Kofman.[4]

Bounds on the Size of Nondeterministic Finite State Codes

If a distance constraint is not imposed, then the asymptotic rate of a RLL-code is given by the logarithm of largest eigenvalue of the graph's adjacency matrix. This result, is generally true for any finite state code provided the graph is deterministic, that is, the edges leaving any given vertex have unique labels. A nondeterministic graph having m-vertices can always be mapped into an equivalent deterministic graph, but the new graph may have as many as 2^m-1 vertices.[2] Thus, when m is large it may be computationally difficult to determine the largest eigenvalue of the new graph's adjacency matrix. A new lower bound has been developed for the maximum rate at which information can be transmitted across a noiseless channel using sequences produced by a nondeterministic graph. The lower bound is expressed in terms of the largest eignevalues of a pair of m^2 by m^2 matrices, and these matrices are easily found from the original graph.

[1].V.Yu.Krachkovsky, Generating Functions and Lower Bounds on Rates for Limited Error-Correcting Codes," IEEE Trans. Inform. Theory, 37,pp. 778-788, (1991).
[2].Z. A. Khayrallah, "Finite-State Codes and Input-Constrained Channels," Ph.D. Dissertation, University of Michigan (1989).
[3].R. J. McEliece, E. R. Rodemich, H.C. Rumsey, Jr. and L.R. Welch, "New Upper bounds on the Rate of a Code via the Delsarte-MacWilliam Inequalities Inequalities,"IEEE Trans.Inform. Theory, 23, pp. 157-166, (1977).
[4].S. Shamai and Y. Kofman, "On the Capacity of Binary and Gaussian Channels With Run-Length-Limited Inputs," IEEE Trans. Commun., 38, pp. 584-594 (1990).

Runlength Limited Trellis Codes for Partial Response Recording Channels[1]

by

Mignon Belongie[2] and Chris Heegard[3]

Summary

Partial response models for recording channels go back many years [1]. This method of channel modeling has led to an interest in developing trellis codes geared to such channels [2-7]. This talk describes a comparison of certain techniques for code construction that combines run-length limiting constraints with constraints that improve the free distance (and thus the noise tolerance) over common models for magnetic recording channels. Three partial response channels are considered, "1 - D", "1 - D^2" and "1 +D - D^2 - D^3".

We consider two techniques to define codes and find encoders with run-length constraints and coding gain. In the first constraint type, a convolutional code is used to constrain the locations of the transitions of the signal [4]. In this case, a convolutional code over the ring of integers modulo q, Z_q is specified. The code is used to constrain the sequence of transition times, modulo q; the sequence must be a member of the fixed convolutional code. It has been shown how this is an effective technique for finding codes with a non-trivial d constraint (i.e., d > 0) [4, 7]. The other construction is based on a recent result of Siegel and Karabed, [5-7], and shows more promise in the d=0 case. Their result shows that matching the channel null with a null in the codebook leads to a coding gain.

In this talk we present a comparison of runlength codes we have constructed. The resulting combined constraints are specified by labeled directed graphs. Using Mathematica, we automatically construct a code given a graph specifying a constraint and a rate less than or equal to its capacity. Thus, the main problem is finding interesting constraints. A constraint is said to have a certain free distance d_{free}, for a given partial response channel, if the distance between any two runlength sequences satisfying the constraint have distance at least d_{free} at the output of the channel (and at least one pair of codewords has distance d_{free}). An important issue is the number of states, both in the original constraint and in the final code. The number of states in the constraint determines the complexity of the decoder; we decode these codes by finding the signal satisfying the constraint that is closest to the received signal. (It is possible that a received signal could thus be decoded to something that isn't actually a codeword from the encoder, the probability of making an error of this type is no more than that of picking a codeword that is incorrect.) The number of states is important for encoding and uncoding. A table is presented that summarizes of the codes that we have found.

[1] H. Kobayashi and D. T. Tang, "Application of Partial Response Channel Coding to Magnetic Recording Systems," *IBM Journal of Research and Development*, Vol. 14, pp. 368-375, July 1970.

[2] A. Robert Calderbank, Chris Heegard and Ting-Ann Lee, "Binary Convolutional Codes with Application to Magnetic Recording," *IEEE Transactions on Information Theory*, Vol. IT-32, No. 6, pp. 797-815, November 1986.

[3] Jack K. Wolf and Gottfried Ungerboeck, "Trellis Coding for Partial Response Channels," *IEEE Transactions on Communications*, Vol. COM-34, pp. 765-773, August, 1986.

[4] Chris Heegard, "Trellis Codes for Recording," 1988 IEEE Military Communications Conference, San Diego, October 23-26, 1988.

[5] Razmik Karabed and Paul Siegel, "Matched Spectral Null Trellis Codes for Partial Response Channels," *IEEE Transactions on Information Theory*, Vol. IT-37, May 1991.

[6] Mignon Belongie and Chris Heegard, "Pairwise Charge Constrained Run Length Codes," 1991 Conference on Information Sciences and Systems, John Hopkins University, March, 1991.

[7] Mignon Belongie, "Run-Length Codes Based on Variable Length Graphs," PhD Thesis, Cornell University, January, 1992.

1. This work was supported in part by NSF grants NCR-8903931, NCR-9207331 and IBM.

2. JPL, Pasadena, CA (formerly School of Electrical Engineering, Cornell University, Ithaca, NY).

3. School of Electrical Engineering, Cornell University, Ithaca, NY.

ON TRELLIS CODES FOR PEAK SHIFT MAGNETIC RECORDING CHANNEL

Ephriam Zehavi and Aaron Biniashvili
Department of Electrical Engineering
Technion — Israel Institute of Technology
Haifa 32000, Israel

Peak detectors are a common practice in standard high density magnetic recording systems [1]-[3]. It has been noticed that one of the major impairments in these systems is the so-called bit shift (peak shift) [1]-[5]. In this talk we propose a new coding technique for correcting peak shifts.

A coded system that operates over a multi- peak (bit) shift channel (PSC) is best formulated in terms of phrase lengths, where a phrase is uniquely defined by a consecutive sequence of bits starting with none, one or more zeros ("0") and terminating with the first single one ("1") [6]. Any binary sequence of zeros and ones is uniquely decomposable into a concatenated sequence of phrases.

The new coding scheme is described as follows. A sequence of phrases $U = \{..., U_i, ...\}$, $U_i \in S = \{d+1, ..., k+1\}$ is mapped to the sequence of symbols $V = \{..., V_i, ...\}$, $V_i \in GF(q)$, according to

$$V_i = U_i - (d+1) \bmod q. \qquad (1)$$

The sequence V is passed through a rate k_0/n systematic convolutional code over $GF(q)$ that converts k_0 input symbols into n output symbols. The encoder output sequence C is mapped back to sequence of phrases in S according to

$$X_{in+r} = \begin{cases} U_{ik_0+r} & , r = 0, ..., k_0-1 \\ C_{i\,k_0+r} + d+1 & , r = k_0, ..., n-1 \end{cases} \quad , i = 0, 1, ..., \qquad (2)$$

The combined coding and mapping defines a trellis code with $(k-1-d)^{\nu}$ encoder states, where ν is the constraint length of the convolutional code. Note that the parity check phrases of the trellis code are in the set $\{d+1, ..., d+q\}$.

The output of the trellis encoder is passed through a peak (bit) shift channel (PSC). A r- position bit shifts cause the "1" symbol terminating the phrase to wonder by r positions to its right (right shifts) or to its left (left shifts). The bit shift effect, either shrinks or expands the input phrase. Of course the phrase lengths are not modified if no bit shift has taken place. We restrict our discussion to $d \geq 2$, $k \geq d+q$. Therefore, in this case additional phrases are neither generated nor existing phrases are destroyed, and the parity check symbols are not violating the (d,k) constraint.

Let X_i stands respectively for the i –th channel input phrase length and Y_i for the corresponding channel output phrase length. Then, the peak shift channel is described by

$$Y_i = X_i + e_i - e_{i-1}. \qquad (3)$$

Here ,e_i is a random variable taking on $(0, \pm 1, ..., \pm j, ..., \pm t)$ values designating whether a left $(-j)$, a right $(+j)$ or no (0) bit shifts has occurred at the end of the i –th phrase. The set of legitimate peak shifts is $S' = (0, \pm 1, ..., \pm j, ..., \pm t)$.. We assume here that $\{e_i\}$ is an independent identically distributed (i.i.d.) sequence with some probability distribution.

The output phrase length is in the set $\{d-2t+1, ..., k+2t+1\}$. At the receiver the decoder produces a maximum likelihood estimate of the sequence U which we write as $\hat{U}$.

In the talk we will analyze the main properties of the codes, the error correction capability of the proposed system, and practical decoding techniques. An optimal decoder as well as sub optimal decoder will be introduced. Interesting upper and lower bound on the maximum data rate that can be transferred as a function of the convolutional code rate that is used is also worked out and discussed.

For example it will be shown that the maximum information rate that can be transmitted using this technique over a noiseless channel is bound by

$$R((d+1+t)\frac{n-k_0}{n})) \leq R \leq R((d+1)\frac{n-k_0}{n})).$$

Here, $R(x)$ is the solution to the equation

$$\sum_{i=d+1}^{k+1} 2^{-iR(x)} = 2 - xR(x).$$

References

[1] P. H. Siegel, "Application of a Peak Detection Channel Model", IEEE Transactions on Magnetics, Vol. MAG-18, Nov. 1982, pp. 1250-1252.

[2] P. H. Siegel, "Recording Codes for Digital Magnetic Storage", IEEE Transactions on Magnetics, Vol. MAG-21, No. 5, Sept. 1985, pp. 1344-1349.

[3] R. Wood,"Magnetic and Optical Storage Systems: Opportunities for Communications Technology", Int. Conf. on Comm., ICC-89, pp. 53.1.1-53.1.8, Boston, MA, June 1989.

[4] E. R. Katz and T. G. Cambell, "Effect of Bitshift Distribution on Error Rate in Magnetic Recording", IEEE Transactions on Magnetics, Vol. MAG-15, No. 3, May 1979, pp. 1050-1053.

[6] S. Shamai (Shitz) and E. Zehavi, "Bounds on the Capacity of the Peak Shift Magnetic Recording Channel", Trans. Info. Theory, Vol. IT-37, No. 3, Pt. I, , pp. 863-871, May 1991.

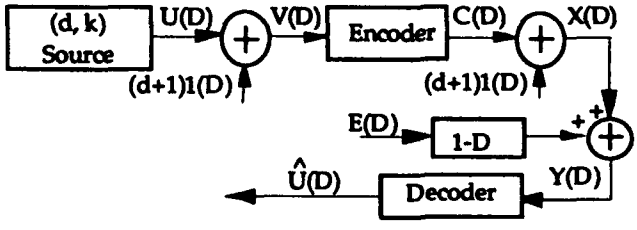


Figure 1:Description of the Coded System using D-Transform

A CLASS OF DC-FREE SUBCODES OF CONVOLUTIONAL CODES

M. Nasiri-Kenari and C. K. Rushforth
Department of Electrical Engineering
University of Utah
Salt Lake City, Utah 84112

Abstract

We describe a class of dc-free subcodes of convolutional codes that satisfy certain runlength constraints and that also possess error-correcting capability. The running disparity and the maximum runlength of these codes are bounded by quantities that are independent of the free distance. Decoding is accomplished using a Viterbi decoder for the underlying convolutional code.

1. Introduction

Deng and Herro [1] describe a method for constructing a dc-free coset code from a given block code. Their codes satisfy

$$R_{max} \le W + \left\lfloor \frac{W}{2} \right\rfloor$$

and

$$K \le 2W + \left\lfloor \frac{W}{2} \right\rfloor - 1$$

or

$$K \le 2\left\{ W + \left\lfloor \frac{W}{2} \right\rfloor \right\} - 1,$$

where R_{max} is the maximum running disparity, K+1 is the maximum runlength, W is a quantity that is greater than or equal to the minimum Hamming distance of the code, and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x [1][2]. Thus, these bounds tend to increase as the error-correcting capability of the code increases.

Using a procedure similar to that of Deng and Herro [1], we have developed a method for finding a subcode of a convolutional code that has a spectral null at dc and at the same time satisfies certain runlength constraints. We describe this construction procedure and state the basic properties of the codes it produces in Section 2. We describe some representative codes obtained using this procedure in Section 3.

2. Construction of the Codes

We begin with a convolutional code with rate (m-1)/m whose m-1 input sequences and m output sequences are $\{a_{1,i}\}..\{a_{m-1,i}\}$ and $\{b_{1,i}\} \ldots \{b_{m,i}\}$, respectively. The (m-1)×m generator matrix is

$$G(D) = \begin{bmatrix} G_1^1(D) & G_1^2(D) & \ldots & G_1^m(D) \\ G_2^1(D) & G_2^2(D) & \ldots & G_2^m(D) \\ \vdots & \vdots & \vdots & \vdots \\ G_{m-1}^1(D) & G_{m-1}^2(D) & \ldots & G_{m-1}^m(D) \end{bmatrix},$$

where

$$G_i^j(D) = g_{0i}^j + g_{1i}^j D + g_{2i}^j D^2 + \ldots$$

is the generator polynomial determining the relationship between input sequence i and output sequence j.

Assume that for one of the (m-1) inputs, the following requirement is satisfied:

$$g_{0i}^j = 1 \quad \text{for} \quad 1 \le j \le m. \tag{1}$$

Without loss of generality, we take the input satisfying (1) to be input 1. We use input bit $a_{1,n}$ not for transmitting information, but for controlling the value of the running disparity. We denote the running disparity prior to the encoding of input block n by R_{n-1}.

It can easily be verified that by complementing the value of $a_{1,n}$ we complement all m output bits in block n. This property enables us to control the value of R_n through the following encoding procedure:

1. Choose $a_{1,n} = 0$ and compute the disparity of the m output bits in block n; denote this disparity by r_n.
2. If $R_{n-1} \cdot r_n < 0$, encode the m-2 information bits; $R_n = R_{n-1} + r_n$.
3. If $R_{n-1} \cdot r_n = 0$, choose the value $a_{1,n}$ to reduce the runlength and then encode the m-2 information bits; if $a_{1,n} = 0$, $R_n = R_{n-1} + r_n$; otherwise $R_n = R_{n-1} - r_n$.
4. If $R_{n-1} \cdot r_n > 0$, change $a_{1,n}$ to 1 and then encode the m-2 information bits; $R_n = R_{n-1} - r_n$.

The disparity r_n of the m output bits at time n is upperbounded by m; moreover,

$$R_{max} \le m + \left\lfloor \frac{m}{2} \right\rfloor$$

and

$$K \le 2m + \left\lfloor \frac{m-1}{2} \right\rfloor - 2, \tag{2}$$

respectively.

3. Example

We obtained a family of convolutional codes of rate 3/4 that satisfy condition (1) by performing row operations on the generator matrices of the corresponding codes given in Table 11.1 of [3]. These codes have constraint lengths ranging from 3 to 9 and free Hamming distances ranging from 4 to 8. The dc-free subcodes obtained by applying our construction procedure to these codes have rate 2/4, with upper bounds on R_{max} and K equal to 6 and 7, respectively.

4. Conclusions

Using an approach similar to that described in [1], we have developed a procedure for constructing dc-free codes with error-control capabilities. The codes produced by this procedure are subcodes of a convolutional code with bounded running disparities and runlengths. The decoder for one of these codes is simply a Viterbi decoder for the underlying convolutional code. Whereas the bounds in [1] increase with the minimum distance of the code, our bounds are independent of distance. The codes given in [1], however, generally have higher rates.

References

[1] R. H. Deng and M. A. Herro, " DC-Free Coset Codes," IEEE Trans. Inform. Theory, vol. 34, pp. 786-792, July 1988.

[2] J. J. O'Reilly and A. Popplewell, " A Further Note on DC-Free Coset Codes," IEEE Trans. Inform. Theory, Vol. 36, pp. 675-76, May 1990.

[3] S. Lin and D. J. Costello, Jr., *Error Control Coding*, Prentice-Hall, Englewood Cliffs, N. J., 1983.

CONCATENATED CODING FOR BINARY PARTIAL-RESPONSE CHANNELS

Giovanni Cherubini and Sedat Ölçer

IBM Research Division, Zurich Research Laboratory
CH-8803 Rüschlikon, Switzerland

ABSTRACT

Concatenation of outer conventional rate-k_c/n_c binary convolutional coding with inner rate-k/n special trellis coding for binary partial-response 1-D channels is investigated. In the considered scheme, the code sequence generated by the convolutional code is time-interleaved prior to trellis encoding. Decoding for the outer convolutional code takes into account the reliability of individual code symbols provided by the inner trellis-decoding stage. The trellis code is designed to achieve large minimum Euclidean distance. The reliability of the decoded symbols is obtained in a computationally efficient way. The construction of the trellis code is based on the partitioning of a set of noiseless channel-output sequences into subsets which are assigned to the branches of a combined encoder and channel trellis. An algorithm for constructing the subsets of channel-output sequences is discussed. Trellis codes with various rates, minimum distances, and complexities are described and the performance achieved by concatenation with different convolutional codes is presented. It is shown that substantial coding gains can be achieved by this method as compared to the maximum-likelihood sequence detection of uncoded binary signals.

SUMMARY

It is well known that large coding gains can be achieved by concatenated coding [1]. When two concatenated codes are employed in conjunction with time-interleaving, decoding for the inner code takes place first. If decoding for the outer code is based on soft decisions, increased immunity against noise in the order of 2 dB can be obtained over decoding schemes using hard decisions only [2].

This paper investigates concatenated coding for binary transmission over partial-response channels described by the time-discrete transfer function $1-D$. A conventional rate-k_c/n_c binary convolutional code is employed as an outer code. The inner rate-k/n trellis code is designed to achieve large minimum Euclidean distance and is constructed so that the reliability of the decoded symbols is obtained in a computationally efficient way.

The sequence of information bits is mapped into a sequence of binary symbols by convolutional encoding. The binary symbols are then time-interleaved and the obtained sequence $\{b_n\}$, $b_n \in \{0,1\}$, is input to the trellis encoder. The trellis encoder provides a sequence of binary channel-input signals $\{a_n\}$, $a_n \in \{-1,+1\}$. At the output of the channel,

$$z_n = a_n - a_{n-1} + w_n, \tag{1}$$

where $\{w_n\}$ is a sequence of white Gaussian noise samples. Decoding for the inner trellis code is performed using the sequence of unquantized channel-output signals $\{z_n\}$. In this decoding stage, reliability information associated with the binary symbols b_n is computed employing the combined encoder and channel trellis. The sequence of reliability information is then deinterleaved and used to perform soft-decision decoding for the outer convolutional code.

A v-state, rate-k/n trellis code is constructed as follows. Let $k = k_1 + k_2$. Consider a set of noiseless $1-D$ channel-output sequences of length n. This set is partitioned into M subsets, each subset containing 2^{k1} sequences. The sequences within a subset correspond to paths through the $1-D$ channel trellis that start from a common state and end, after n transitions, in a common state. Sequences within a subset are separated by a minimum Euclidean distance of d_{min}. The subsets are then assigned to the branches of a combined encoder and channel trellis with v states according to Ungerboeck's criterion [3]. There are 2^{k2} branches leaving from and entering into each trellis state. Therefore, in the encoder for the trellis code, k_1 input bits select one of 2^{k1} parallel transitions between two consecutive encoder states, and k_2 input bits determine a state transition in the encoder trellis.

The free Euclidean distance of the innner trellis code is given by

$$d_{free} = \min\{d_{min}, d'_{min}\}, \tag{2}$$

where d'_{min} is the minimum distance of error events of length greater than one trellis step. An algorithm for constructing the set of channel-output sequences is discussed.

For computation of the reliability information relative to the symbols b_n, either maximum-*a-posteriori* symbol estimation or the soft-output Viterbi algorithm (SOVA) can be used [2]. In the case $k_2 = 1$ and $d_{min} < d'_{min}$, a simple algorithm is obtained. At each decoding step, the reliability of the k_1 symbols on the parallel transitions is computed by using the associated branch metrics. The reliability of the remaining symbol is obtained by the SOVA.

Trellis codes with various rates, minimum distances and complexities are presented. Their concatenation with different convolutional codes is investigated. Overall performance is measured in terms of the asymptotic coding gain (ACG) over the baseline system [4], which is expressed in dB as

$$\mathrm{ACG} = 10\log_{10}\frac{d_{free}^{H}\, d_{free}^{2}}{8} - 10\log_{10}\frac{1}{R}\ , \tag{3}$$

where d_{free}^{H} is the free Hamming distance of the outer convolutional code, and $R = (k_c k)/(n_c n)$ is the overall code rate. For example, concatenation of a 32-state, rate-2/3, $d_{free}^{H} = 6$ convolutional code with a 4-state, rate-3/6, $d_{free}^{2} = 24$ trellis code gives R = 1/3 and ACG = 7.8 dB. Concatenation of a rate-3/4 convolutional code with 32 states and $d_{free}^{H} = 5$ with the trellis code of the previous example gives R = 3/8 and ACG = 7.5 dB.

References

[1] G. D. Forney, *Concatenated codes*, M.I.T. Press, Cambridge, Massachusetts, 1966.

[2] J. Hagenauer, P. Hoeher, and J. Huber, "Soft-output Viterbi and symbol-by-symbol MAP decoding: algorithms and applications," submitted to *IEEE Trans. Commun.*, 1991.

[3] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, Vol. IT-28, pp. 55-67, Jan. 1982.

[4] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, Vol. COM-34, No. 8. pp. 765-772, Aug. 1986.

NEW ZERO-RUN LENGTH LIMITED CODES FOR PARTIAL-RESPONSE CHANNELS

Kjell Jørgen Hole and Øyvind Ytrehus

University of Bergen, Department of Informatics, HiB, N-5020 Bergen, Norway. E-mail: Kjell.Hole@ii.uib.no, Oyvind.Ytrehus@ii.uib.no
Supported by the Norwegian Research Council for Science and the Humanities (NAVF).

Abstract

A coset of a convolutional code may be used to generate a zero-run length limited trellis for a 1–D partial-response channel. It is well known that the free Hamming distance d_H of the convolutional code and the free squared euclidean distance of the coset, d^2_{free}, measured at the channel output, are related by $d^2_{\text{free}} \geq d_H$. In this talk we present cosets for which d^2_{free} is larger than the free Hamming distance of any convolutional code with the same rate and constraint length. We also describe how the new bounds, described in [1, 2], on the maximum zero-run length of cosets may be used to ensure a short zero-run length at the channel output. Analytical arguments, supported by results from computer search, suggest that cosets with large free squared euclidean distance also have short maximum zero-run lengths.

Introduction

Several authors (among them Wolf and Ungerboeck [3], and Calderbank, Heegard and Lee [4]) have described recording systems in which the recording channel is regarded as a partial response channel with transfer polynomial of the form $(1-D)^N$, where $N \in \{1,2\}$. In [3], the binary information sequence is encoded by an error-correcting code; sent through a *channel precoder* and subsequently through a $(1-D)$ partial response channel. The precoder essentially inverts the channel transfer function. The overall channel accepts a binary ($\{0,1\}$) input sequence $\mathbf{x} = (\ldots, x_{i-1}, x_i, x_{i+1}, \ldots)$ and produces a ternary ($\{-1,0,1\}$) output sequence $\mathbf{y} = (\ldots, y_{i-1}, y_i, y_{i+1}, \ldots)$, where $|y_i| = x_i$ for all i and the signs alternate.

Codes for such channels should have

(1) *large free squared euclidean distance*, d^2_{free}, where the squared euclidean distance between two output sequences $\mathbf{y}, \widehat{\mathbf{y}}$ is defined as $\sum_i (y_i - \widehat{y}_i)^2$, and
(2) *short maximum zero-run length*, defined as the maximum number of consecutive zeros in a code word.

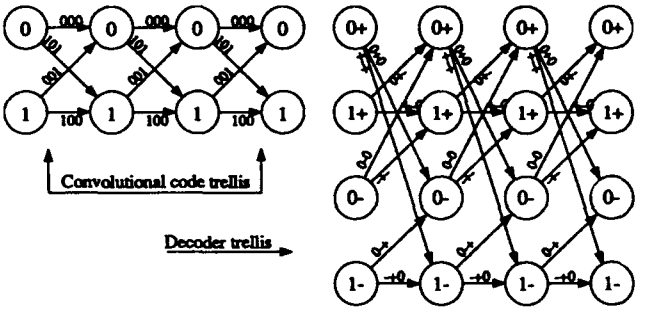
The squared euclidean distance between any pair of noiseless output sequences is at least as large as the corresponding Hamming distance between the original convolutional code words. In [3], a binary convolutional code with large free Hamming distance is used for error correction. In order to satisfy requirement (2), a coset of the code is used rather than the linear code itself.

New Codes

We present new codes for such channels. As in [3], the new codes are cosets of convolutional codes. For example, consider the (3,1) convolutional code with constraint length 1 and parity-check matrix

$$\mathrm{H}(D) = \begin{pmatrix} 0 & 1 & 0 \\ 1+D & 0 & 1 \end{pmatrix}$$

The convolutional code has free Hamming distance 3, as can be seen from the trellis of the convolutional code in the figure. We obtain one coset of the code by adding the vector (010) to each branch label. The decoder trellis for the precoded (1–D) partial-response channel is also shown in the figure.



We exploit the underlying linearity of the decoder trellis to calculate d^2_{free} (14 in the example), or a lower bound on it. The resulting algorithm, which is related to one described by Zehavi and Wolf [5], was used to find cosets with large d^2_{free}. A few examples are provided in the table below.

Rate	log_2 of # decoder states	d^2_{free}	Best comparable Hamming distance	Max zero-run length
1/3	5	≥24	13	2
2/4	3	≥8	6	3
2/4	6	≥14	10	3
3/5	2	6	4	5
3/5	4	≥8	6	5
4/6	3	6	4	7
7/9(Punct.)	3	≥4	3	8

References

[1] K. J. Hole, "Cosets of convolutional codes for symbol synchronization and error control," Department report 64, Department of Informatics, University of Bergen, June 1992.

[2] K. J. Hole, "Runlength limited error control codes of high rates," Department report 58, Department of Informatics, University of Bergen, February 1992.

[3] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. on Communication*, vol. COM-34, pp. 765–773, Aug. 1986.

[4] A. R. Calderbank, C. Heegard, and T. A. Lee, "Binary convolutional codes with application to recording," *IEEE Trans. on Information Theory*, vol. IT-32, pp. 797–815, Nov. 1986.

[5] E. Zehavi and J. K. Wolf, "On the performance evaluation of trellis codes," *IEEE Trans. on Information Theory*, vol. IT-33, pp. 196–202, Mar. 1987.

REED-MULLER CODING FOR PARTIAL RESPONSE CHANNELS

Sedat Ölçer and Gottfried Ungerboeck

IBM Research Division, Zurich Research Laboratory
CH-8803 Rüschlikon, Switzerland

ABSTRACT

This paper deals with Reed-Muller (RM) coding and concatenated soft-decision decoding for binary partial-response class-IV (PRIV) channels. Block interleaved RM codewords are transmitted with precoding over the PRIV channel. In the receiver, soft-decision decoding is accomplished in two stages. An inner decoder accounts for the precoded transmission of binary symbols over the PRIV channel. Approximate log-likelihood ratios for individual code symbols are determined by a new bi-directional symbol estimation algorithm. The obtained soft information values are deinterleaved and passed to the outer decoding stage. With sufficient interleaving, these values represent the appropriate metrics for soft-decision RM decoding. An efficient suboptimal decoding algorithm based on the generalized multiple concatenation (GMC) structure of RM codes is employed. Real coding gains over uncoded transmission with maximum-likelihood sequence decoding were determined by simulation. Results are presented for various RM code parameters and degrees of interleaving. Encoder and decoder realization as well as complexity issues are addressed. A comparison with other coding schemes known for PRIV channels shows significant advantages of the scheme considered here in terms of coding gains versus decoding complexity.

SUMMARY

Coding techniques for binary partial-response class-IV (PRIV) channels with time-discrete transfer function $1-D^2$ have been studied, e.g., in [1,2]. In [1], convolutional codes are used in conjunction with precoding. The free Euclidean distance between sequences of noiseless channel-output signals is then essentially given by the Hamming distance of the convolutional code employed. With the matched spectral-null codes described in [2] gains in free Euclidean distance are achieved by sending constrained sequences with zero spectral energy at the null frequencies of the PRIV channel.

In this paper, we investigate the application of Reed-Muller (RM) block codes. Information bits are encoded by a systematic RM encoder, block interleaved, and then transmitted over the PRIV channel with precoding. In the receiver, concatenated soft-decision decoding [3] is employed. The inner decoding stage accounts for the precoded transmission of binary symbols over the noisy PRIV channel and derives soft information values for these symbols. These values are deinterleaved and passed to the outer decoding stage for soft-decision RM decoding. The systematically encoded information bits are then immediately available from the recovered RM codewords.

We denote the interleaved RM code symbols by b_k, $k = \ldots 0,1,2, \ldots$, and represent these symbols in the bipolar form $b_k \in \{-1,+1\}$ with the mapping *logical 0* $\rightarrow -1$, and *logical 1* $\rightarrow +1$. The precoder generates the binary transmit symbols $a_k = -b_k\, a_{k-2}$. The output signals of the noisy PRIV channel become

$$z_k = a_k - a_{k-2} + w_k \,. \tag{1}$$

where w_k accounts for additive i.i.d. Gaussian noise. If sufficient interleaving is employed, the channel-output signals containing information about one particular code symbol b_k become essentially independent of the other spread-out code symbols belonging to the same codeword. Hence, the log-likelihood ratios

$$\beta_k = \ln \frac{p(z_{-\infty}^{+\infty} | b_k = +1)}{p(z_{-\infty}^{+\infty} | b_k = -1)} \,. \tag{2}$$

where $z_{-\infty}^{+\infty} = \ldots z_0, z_1, z_2, \ldots$ denotes the infinite sequence of observed channel-output signals, represent appropriate metrics for soft-decoding of RM codewords.

In the inner decoding stage, soft information values $y_k \approx \beta_k$ are computed by a new algorithm for precoded-symbol estimation in the presence of intersymbol interference caused by the PRIV channel. As for (approximate) maximum *a posteriori* (MAP) estimation of the transmit symbols a_k [3], Viterbi algorithm path-metric computations are performed both forward and backward in time. The quantities y_k are then obtained by suitably combining forward and backward difference metrics. It is apparent from (2) that signals with even and odd time indices can be processed independently. For practical reasons, two known even- and odd-indexed transmit symbols a_k are inserted between interleaving blocks to provide starting points for the forward and backward recursions. In the outer decoding stage, the transmitted information bits are recovered from the deinterleaved values y_k by an efficient suboptimal soft-decision decoding algorithm for RM codes recently described in [4,5].

A binary RM code R(r,m), for $0 \le r < m$, exhibits the code parameters

$$[\, n = 2^m \,,\, k = \sum_{i=0}^{r} \binom{m}{i} \,,\, d = 2^{m-r} \,] \,. \tag{3}$$

Codewords can be generated either according to a well-structured $k \times n$ generator matrix or as codewords of length $n-1$ of a cyclic code extended by adding one even-parity check bit [6]. The first interpretation implies a generalized multiple concatenation (GMC) structure, defined by the iterative construction

$$R(r+1,m+1) = \left\{ |u|u \oplus v| : u \in R(r+1,m),\ v \in R(r,m) \right\} \,. \tag{4}$$

The GMC structure has been exploited to devise the soft-decision decoding algorithm [4,5] employed in the outer decoding stage. Encoding could be based on the same structure, but a systematic encoder would then not be easily obtained. The extended cyclic-code interpretation of RM codes is preferred as a basis for systematic encoding by simple shift-register circuits. The two interpretations of RM codes lead to different orderings of the code symbols. The reordering required to rearrange the code symbols in the extended cyclic code for decoding by the GMC-based decoding algorithm is explained.

Simulation results are presented which show significant real coding gains obtained over uncoded binary PRIV transmission with optimum maximum-likelihood sequence decoding. We find, for example, that with a $R(5,9) \triangleq [512,382,16]$ code a real coding gain of 4 dB is obtained in terms of E_b/N_0 at a bit-error rate of 10^{-7}. In this case, only 23 elementary arithmetic operations are needed for RM decoding per information bit. Interleaving plays a lesser role for RM codes with large Hamming distance. With convolutional encoding as described in [1], a similar real coding gain could only be achieved with a $R = 3/4$, $\nu = 8$ convolutional code and a 512-state Viterbi decoder. With binary spectral-null codes [2], coding gains of 4 dB cannot be achieved even asymptotically. The approach pursued in this study for RM codes can be used for convolutional coding as well. Further investigation is needed.

References

[1] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, Vol. COM-34, No. 8, pp. 765-772, Aug. 1986.

[2] R. Karabed and P. H. Siegel, "Matched spectral-null codes for partial-response channels," *IEEE Trans. Inform. Theory*, Vol. 37, No. 3, pp. 818-855, May 1991.

[3] J. Hagenauer, P. Hoeher, and J. Huber, "Soft-output Viterbi and symbol-by-symbol MAP decoding: algorithms and applications," submitted to *IEEE Trans. Commun.*, 1991.

[4] G. Schnabl, M. Bossert, and H. Dietrich, "Reed-Muller coded modulation using a new soft-decision decoding algorithm," in *Coded Modulation and Bandwidth-Efficient Transmission*, Proc. of 5th Tirrenia Int'l Workshop on Digital Communications, Sept. 1991, edited by E. Biglieri and M. Luise, (Elsevier, Amsterdam, 1992), pp. 195-200.

[5] G. Schnabl and M. Bossert, "Soft-decision decoding of Reed-Muller codes as generalized multiple concatenated codes," submitted to *IEEE Trans. Inform. Theory*, 1991.

[6] F. J. Mac Williams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, (North-Holland, Amsterdam, 1977).

A Class of Byte Error Control Codes for Memory Systems
— SbEC-(Sb+S)ED Codes —

Mitsuru HAMADA and Eiji FUJIWARA
Dept. of Computer Science, Tokyo Institute of Technology
O-okayama, Meguro-Ku, Tokyo 152, Japan

1 Introduction

Single bit error correcting and double bit error detecting (SEC-DED) codes are widely used in computer semiconductor memory systems organized in a one-bit-per-chip manner. This is because in this organization any failure in one chip can corrupt, at most, one bit per codeword. Recently some systems adopt a b-bit-per-chip organization, where $b \geq 2$ [1]. A chip failure in these systems causes the word read-out to have a b-bit block, called *byte*, in error. Therefore, SbEC-DbED codes, capable of correcting all single b-bit byte errors and detecting double b-bit byte errors, have found applications in this kind of systems. Among the predominant errors, however, are the soft errors induced by α particles, which are said to be apt to manifest themselves as single bit errors still in byte organized systems. Consequently, these systems need protection against a single bit soft error lined up in a codeword with another existing single byte hard error due to a chip failure.

From the standpoint mentioned above, this paper proposes a new class of linear codes, called single b-bit byte error correcting and single b-bit byte and single bit error detecting codes, or SbEC-(Sb+S)ED codes. This class of codes can correct all single byte errors and detect any error that corrupts both a single byte and a single bit in a codeword.

In the later sections, bounds and construction methods of SbEC-(Sb+S)ED codes are given, and it is shown that some codes proposed in this paper meet a lower bound on check bit length.

2 SbEC-(Sb+S)ED Codes

In this paper, codewords of N bits length are divided into n bytes, where all the bytes are b-bit wide except the last one which is allowed to have c-bit width ($0 < c \leq b$). The following notations are used in this paper:

b_i	:	byte length of the ith byte, i.e., $b_i = b$ for $i = 1, 2, \cdots, n-1$ and $b_n = c$.
n	:	code length in byte.
N	:	code length in bit, i.e., $N = b(n-1) + c$.
K	:	information bit length.
R	:	check bit length, i.e., $R = N - K$.
(N, K)code	:	linear code of code length N and information bit length K.
o_d	:	vector composed of d 0's. d is omitted if there is no fear of confusion.
x^t	:	transpose of vector x.
$[XY]$	:	concatenation of vectors/matrices X and Y.

All the matrices and vectors in this paper are over $GF(2)$ and vectors are row vectors unless referred to otherwise. When the parity check matrix of an (N, K) SbEC-(Sb+S)ED code is expressed as $H = [H_1 H_2 \cdots H_n]$, where H_i is an $(N-K) \times b_i$ matrix for $i = 1, 2, \cdots, n$.

SbEC-(Sb+S)ED codes are a class of single b-bit byte error correcting codes, which can detect any double byte error such that, at least, one of the two byte errors has Hamming weight one. Fig. 1 shows examples of a correctable error and a detectable error of S4EC-(S4+S)ED codes.

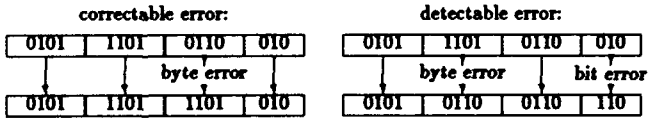


Fig.1. Examples of a correctable error and a detectable one for $b = 4$ and $N = 15$.

In the linear case, the definition of the codes is equivalent to the following theorem.

Theorem 2.1 A linear (N, K) code with parity check matrix $H = [H_1 H_2 \cdots H_n]$ is an SbEC-(Sb+S)ED code iff

$\forall i, j \in \{1, 2, \cdots, n\}$ $(i \neq j)$, $\forall e_1 \in GF(2)^{b_i}, e_2 \in GF(2)^{b_j}$,

$$[e_1 e_2] \neq o \longrightarrow H_i e_1^t + H_j e_2^t \neq o^t,$$

and $\forall i, j, k \in \{1, 2, \cdots, n\}$ $(i \neq j \neq k \neq i)$, $\forall e_1 \in GF(2)^{b_i}, e_2 \in GF(2)^{b_j}, e_3 \in GF(2)^{b_k}$,

$$e_3 \text{ has Hamming weight one} \longrightarrow H_i e_1^t + H_j e_2^t + H_k e_3^t \neq o^t. \quad \square$$

3 Bounds

Theorem 3.1 Linear (N, K) SbEC-(Sb+S)ED codes satisfy

$$R = N - K \geq 2b + 1,$$
$$R = N - K \geq b + \lceil \log_2(N - 2b + 2^b) \rceil,$$
and $$R = N - K \geq \lceil \log_2[(2^b - 1)\{N(b+1) - b^2 - c\}/b + 2^c] \rceil. \quad \square$$

The first inequality in the theorem corresponds to Singleton bound, the second and the last ones to Hamming bound [2]. Roughly speaking, the second bound is tighter than the last when N is relatively small, and vice versa.

4 Code Construction Methods

Two construction methods of SbEC-(Sb+S)ED codes are given in this section. The first one derives codes of arbitrary byte length and code length, while the second one lacks flexibility for code length. The second one, however, provides more efficient codes than the first one.

Construction Method 1

The following procedure derives SbEC-(Sb+S)ED codes from Sb'EC codes [3], where $b' = b - 1$.

1) Let $H' = [H'_1 H'_2 \cdots H'_n]$ denote the $R' \times N'$ parity check matrix of an Sb'EC code. Given the code length N', integers n, c' and b'_i are defined in the same way as in Section 2, so that $N' = b'(n-1) + c'$. If $b' = 1$, an Sb'EC code should be regarded as a simple SEC code.

2) An $R' \times (N' + n)$ matrix $\tilde{H} = [\tilde{H}_1 \tilde{H}_2 \cdots \tilde{H}_n]$ is obtained by $\tilde{H}_i = H'_i \cdot [I_{b'_i} f_i]$, $i = 1, 2, \cdots, n$, where $[I_{b'_i} f_i]$ denotes the $b'_i \times b'_i$ identity matrix $I_{b'_i}$ followed by a b'_i-dimensional even weight column vector f_i.

3) Let an $R'' \times (b_i + 1)$ matrix $U_i = [u_i u_i \cdots u_i]$ denote the collection of the same $b'_i + 1$ odd weight column vectors u_i's, for $i = 1, 2, \cdots, n$. Take $U = [U_1 U_2 \cdots U_n]$ consisting of those U_i's, where $u_i \neq u_j$ for $i \neq j$, $i, j \in \{1, 2, \cdots, n\}$.

4) Finally, the null space of the following matrix is an SbEC-(Sb+S)ED code of byte length $b = b' + 1$, code length $N = N' + n$ and check bit length $R = R' + R''$:

$$\begin{bmatrix} \tilde{H} \\ U \end{bmatrix} = \begin{bmatrix} \tilde{H}_1 & \tilde{H}_2 & \cdots & \tilde{H}_n \\ U_1 & U_2 & \cdots & U_n \end{bmatrix}$$

Theorem 4.1 The codes obtained with the above procedure are SbEC-(Sb+S)ED codes. $\square$

Construction Method 2

Theorem 4.2 The null space of the following matrix is an $(N = b2^b + 3b + 1, K = b2^b)$ SbEC-(Sb+S)ED code:

$$\begin{bmatrix} I & I & I & \cdots & I & I & I & O & O & O & o_b^t \\ I & T & T^2 & \cdots & T^{q-3} & T^{q-2} & O & I & O & I & o_b^t \\ I & T^2 & T^4 & \cdots & T^{2(q-3)} & T^{2(q-2)} & O & O & I & I & o_b^t \\ o_b & o_b & o_b & \cdots & o_b & o_b & o_b & o_b & o_b & 1_b & 1 \end{bmatrix},$$

where $q = 2^b$, I is the $b \times b$ identity matrix, O is the $b \times b$ zero matrix, 1 is the vector composed of d 1's, and the $b \times b$ matrix T is the companion matrix of a primitive polynomial of degree b [1] [3]. $\square$

5 Evaluation

For any byte length $b \geq 2$, the construction method 1 shown in the previous section provides SbEC-(Sb+S)ED codes which meet the first bound in Theorem 3.1. For the practical code parameters of $b = 4$ and $K = 64$, in particular, Theorem 4.2 gives an SbEC-(Sb+S)ED code of check bit length $R = 13$, while the previously known SbEC-DbED codes requires, at least, 14 check bits [4].

6 Conclusion

This paper has proposed a new class of error control codes, SbEC-(Sb+S)ED codes, capable of correcting all single byte errors and detecting any error that corrupts both a single byte and a single bit in a codeword, suitable for semiconductor memory systems organized in a b-bit-per-chip manner.

References

[1] T. N. Rao and E. Fujiwara, *Error Control Coding for Computer Systems*, Prentice-Hall, 1989.

[2] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, 1977.

[3] S. J. Hong and A. M. Patel, "A General Class of Maximal Codes for Computer Applications," *IEEE Trans. Comput.*, Vol.C-21, pp.1322-1331, Dec. 1972.

[4] C. L. Chen, "Symbol Error-Correcting Codes for Computer Memory Systems," *IEEE Trans. Comput.*, Vol.41, pp.252-256, Feb. 1992.

CORRECTING SINGLE PEAK–SHIFTS with PERFECT (d,k)–CODES

V.I. Levenshtein* and A.J. Han Vinck**

*V.I. Levenshtein, Keldysh Institute for Applied Mathematics Miusskaya sq. 4, 125047, Moscow, Russia, Fax: 7–095–9720737.

**A.J. Han Vinck, IEM, Ellernstr. 29, 4300 Essen. Germany, Fax:31–201–3206425, Email MEM100.at.DE0HRZ1A.

ABSTRACT. We consider codes, consisting of sequences

$$0^{\alpha_1}10^{\alpha_2}1\cdots 0^{\alpha_N}1, \text{ where } d\le\alpha_i\le k,$$

and call them (d,k)–codes of reduced length N. We introduce a definition of arbitrary (d,k)– and perfect (d,k)–codes capable of correcting single peak–shifts of given size t. For the construction of perfect codes we use a general combinatorial method connected with finding "good" weight sequences in Abelian groups, and introduce the concept of perfect t–shift N–designs. We give explicit constructions of such designs for t=1, t=2 and t=(p–1)/2, where p is a prime. Our construction is not only effective, but also universal in the sense that it does not depend on the (d,k)–constraints. It also allows to correct automatically those peak–shifts that violate (d,k)–constraints. Furthermore, our construction is naturally extended to (d,k)–codes of fixed binary length and allows the determination of the beginning of the next code word. The question whether the designed codes can be represented as systematic codes with minimal redundancy is considered as well. In particular, for any (d,k)–code with n q–ary (q=k–d+1 ≥ 2) information digits we give a method of finding r q–ary check digits such that the resulting systematic code is capable of correcting single peak–shifts of size 1, where r is determined uniquely by $q^{r-1}(r-1) < 2n+1 \le q^r-2r$. This code is perfect if $2n+1 = q^r-2r$.

INTRODUCTION

In high–density magnetic recording systems Run Length Limited (RLL) sequences are used to increase density and control self clocking[1]. The read–out mechanism detects changes in magnetization and thus from the RLL sequence we can derive a so called (d,k)–sequence, where d+1 and k+1 correspond to the minimum and maximum length of the RLL substrings, respectively. A (d,k)–sequence is represented by consecutive zero symbol runs of length i, d<i≤k, between pairs of one symbols. Read–out circuitry imperfection and clock jittering may cause misdetection of transitions and is supposed to result in peak–shifts left or right in the (d,k)–sequence.

Shamai and Zehavi [2] give bounds on the capacity of the bit–shift magnetic recording channel. Kolesnik and Krachkovski [3] obtained asymptotic bounds on the achievable rates of bit–shift error–correcting codes. Fredrickson and Wolf [4] introduced a class of single bit–shift detecting codes. Codes designed specifically to cope with a single bit shift and multibit shifts of a single position are discussed by Kuznetsov and Vinck and by Ytrehus in [5,6], respectively. Ferreira and Lin [7] give code constructions based on the representation of constrained sequences as integer compositions. Abdel–Ghaffar and Weber [8] extends the results given in [4,6]. We discuss the design of en– and decoding for the multibit peak–shift channel.

We give a definition of a multibit peak–shift and a general definition of a code capable of correcting single peak–shifts of size t. We concentrate on codes C consisting of (d,k)–sequences, and call them (d,k)–codes. For (d,k)–codes with k–d≥2t we introduce the concept of a perfect code capable of correcting single peak–shifts of size t. We remark that the problem of constructing maximum (d,k)–codes is reduced to the same problem for (d,k)–codes with a fixed number N of substrings.

We give a universal and effective construction of (d,k)–codes capable of correcting single peak–shifts of size t. The construction is universal in the sense that it does not depend on the (d,k)–constraints and, in particular, allows to correct single peak–shifts of size t which disturb these constraints. The main idea of the construction consists of using a finite Abelian group G of order m to partition any code C into m subcodes, each having the desired error correcting properties [9]. Since at least one of these subcodes has size at least |C|/m, this construction is efficient if the order of the group G is sufficiently small. In the framework of this construction we reduce the problem of finding perfect (d,k)–codes of reduced length N capable of correcting single peak–shifts of size t to the problem of finding "good" weight sequences in Abelian groups and introduce the concept of perfect t–shift N–designs.

We give explicit constructions of perfect t–shift N–designs for t=1 and any N and for t=(p–1)/2, where p is a prime, and $N=(p^r-1)/(p-1)$. Moreover, we find the necessary and sufficient conditions for the existence of perfect 2–shift N–designs.

We consider the problem of finding the minimum redundancy r of systematic codes which are contained in the constructed perfect (d,k)–codes of reduced length N capable of correcting single peak–shifts of size t. This problem is connected with the existence of a particular ordering of perfect t–shift N–designs. We show that the lower bound $r \ge \lceil \log_q 2tN+1)\rceil$ is attained in some cases, where ⌈x⌉ is the smallest integer at least x and q=k–d+1. Furthermore, for any (d,k)–code with n q–ary information digits we give a method of finding the minimum number of q–ary check digits such that the resulting systematic (d,k)–code is capable of correcting single peak–shifts of size 1.

For an ideal multibit peak–shift channel, decoding errors that do not occur in the Nth substring do not propagate to subsequent blocks, as the length of the code word does not change. However, if a decoding error occurs in the Nth substring, the first symbol of the next block is in error and thus we make a decoding error in this block. Only if again in the Nth substring a decoding error occurs, we may speak of error propagation. By appropriate code selection we may avoid this phenomenon. On the other hand catastrophic error propagation occurs whenever random errors are involved. These errors ruine the structure of code words. They insert new phrases or delete existing phrases in a code word and thus synchronization regarding the beginning of the first symbol of a code word is completely lost. One way to solve this problem is to fix the length of the codeword to a certain value. We construct codes with a fixed binary length L by considering the union of all code words of binary length L belonging to the (d,k)–codes of reduced length N, $L/(k+1) \le N \le L/(d+1)$. The code words of fixed binary length start with d zeros and end with a symbol equal to 1. These code words can be stored without merging digits.

References

[1] K.A. Schouhamer Immink, "Coding Techniques for Digital Recording," Prentice Hall, 1990.

[2] S. Shamai and E. Zehavi,"Bounds on the Capacity of the Bit–shift Magnetic Recording Channel," IEEE Transactions on Information theory, Vol–37, May 1991, pp. 863–872.

[3] V.D. Kolesnik and V.Yu. Krachkovsky, "Generating Functions and Lower Bounds on Rates for Limited Error–correcting Codes," IEEE Trans. on Inf. Theory, May 1991, pp. 778–788.

[4] L.J. Fredrickson and J.K. Wolf, "Error Detecting Multiple Block (d,k) Codes," IEEE Trans. Magn., Vol. MAG–25, pp. 4096–4098, Sept. 1989.

[5] A.Kuznetsov and A.J.Vinck, "Single Peak–shift Correction in (d,k)–sequences," IEEE Int. Symp. Inform. Theory, June 24–28, 1991, Budapest, Hungary, p.256.

[6] O. Ytrehus,"On (d,k) Constrained Error–controlling Block Codes," to be published. See also Abstracts of papers, Int. Symp. Inform. Theory, San Diego, CA, Jan. 1990, p.124.

[7] H.C. Ferreira and S. Lin, "Error and Erasure Control (d,k) Block Codes," IEEE Trans. on Inform. Th., Vol. 37, September 1991, pp. 1399–1408.

[8] K.A.S. Abdel–Ghaffar and J.H.Weber, " Bounds and Constructions for Runlength–limited Error–control Block Codes," IEEE Trans. on Inform. Th., Vol. 37, May 1991, pp. 789–800.

[9] V.I. Levenshtein, "Binary Codes Correcting Spurious Insertions and Deletions of Ones," Probl. Peredachi Inform., Vol 1, pp. 12–25, 1965.

Codes on curves and their geometry

J.W.P. Hirschfeld

University of Sussex, Brighton, U.K.

Linear codes with 'good' parameters can be constructed from algebraic curves over finite fields. Since Goppa's original paper in 1981, there has been a constant flow of research on (1) asymptotic properties of these codes, (2) behaviour of these codes on different types of curves, (3) efficient decoding.

The construction gives linear q-ary $[n, k, d]$-codes satisfying
(a) $|n - (q+1)| \leq 2g\sqrt{q}$; (b) $k = m - g + 1$; (c) $d \geq n - m$.
Here g is the genus of the curve and m is a positive integer satisfying $n > m > 2g - 2$. An important consequence is that d satisfies $n - k + 1 \geq d \geq n - k + 1 - g$.

The length n of the code is at most the number of rational points on the corresponding curve C. So it is of interest to study the codes arising from a curve C with a 'large' number of rational points, that is, points defined over the ground field $\mathbf{F}_q$. The known classes are (1) modular curves; (2) Hermitian curves; (3) Suzuki curves; (4) Ree curves.

The main feature of the modular curves is that they provide a sequence for which $\lim n/g = \sqrt{q} - 1$. This leads to the asymptotic result bettering the Gilbert-Varshamov bound. The number of $\mathbf{F}_q$-rational points on a Hermitian curve is $q\sqrt{q} + 1$; the number of $\mathbf{F}_q$-points on a Suzuki curve is $q^2 + 1$; the number of $\mathbf{F}_q$-rational points on a Ree curve is $q^3 + 1$. In these last three cases as well, interesting codes are obtained. A common feature is the great symmetry that these curves enjoy, in the sense of having a large group of automorphisms.

The geometry of these codes can be explored from two points of view. Their large automorphism group of the curves reflects many interesting geometrical properties. Also a linear q-ary $[n, k, d]$-code can be considered as a set of n points in the projective space $\mathbf{P}^{k-1}$ with at most $n - d$ of these points in any hyperplane. This connects coding theory problems on the maximum value of parameters with combinatorial problems in finite projective spaces on the maximum size of subsets subject to certain intersection conditions.

Algorithms Analogous to Algebraic Geometric Codes

Gilles Lachaud

Laboratoire de Mathématiques Discrètes du CNRS

Luminy Case 930, 13288 - Marseille Cedex 9 - FRANCE

We describe polynomially constructible (PC) algorithms based on curves with many rational points, namely those curves which are used in the theory of algebraic geometric codes, but involved here in some other contexts. We shall develop the following results :

multiplication algorithms

D. and G. Chudnovsky discovered that one can use curves with many points in order to define fast bilinear multiplication algorithms in large extensions of a finite field. If we have a bilinear multiplication algorithm B(n) expressing the product of two elements in $GF(q^n)$, the *relative multiplicative complexity* of B(n) is defined as

$$\mu(B(n)) = \frac{m(B(n))}{n}$$

where m(B(n)) is the number of multiplications by non-constant terms needed in order to perform B(n). Then, following Shparlinsky, Tsfasman and Vladut, there are families of algorithms with asymptotic multiplicative complexity $\mu_q = \liminf \mu(B(n))$ rather small and in any case finite, for instance $\mu_2 < 35/6$.

dense lattices

Let P be the set of centers of a packing of equal non-overlapping spheres in the euclidean space $\mathbf{R}^n$. Denote by $\delta(P)$ the density of P and by

$$\lambda(P) = \log_2 \frac{\delta(P)}{n}$$

the *density exponent* of P. Elkies and Shioda have given a general process of construction of dense lattices based on elliptic curves. The Coxeter-Todd lattice, the Leech lattice can be constructed in this way ; and one gets in high dimensions some lattices densest than those previously known.

asymptotical bounds for sphere packings

If we have a family of lattices $P(n) \subset \mathbf{R}^n$ with $n \to \infty$, then Minkowski showed that there exist families with asymptotic exponent $\limsup \lambda(P(n)) \geq -1$; but in fact it is very difficult to construct asymptotically good families of lattices, i.e. with finite asymptotic density exponent. By the use of algebraic curves with many rational points, several authors (Lytsin, Quebbemann, Rosenbloom, Tsfasman, Vladut) gave PC constructions of families of asymptotically good families of lattices and of sphere packings.

asymptotical bounds for spherical codes

Consider spherical codes X on the unit sphere of $\mathbf{R}^n$ with angular distance φ. The number

$$R = \log_2 \frac{\# X}{n}$$

is called by Shannon the *reliability* of such a code, and he gave a lower bound for the asymptotic reliability of families of such codes. Then there are PC families of spherical codes whose reliability is at least one half of the Shannon lower bound (Lachaud, Stern). There are also families with asymptotic kissing number > 2/15.

ALGEBRAIC GEOMETRY TOOLS IN CODING THEORY

S.G. Vladut
Institute for Problems of Information Transmission
Russian Academy of Science
19 Ermolovy Street
Moscow, 101447, Russia

November 2, 1992

This talk is devoted to some applications of algebraic geometry to coding theory other than algebraic geometry codes. The general scheme of these applications is converse to the Goppa construction, which associates a code to some algebraic geometry data. Here, on the contrary, some problems in coding theory give rise to certain algebraic varieties over finite fields, so that these problems can be formulated as questions about these varieties (usually concerning their ratonal points). One schould mention that usually the algebraic geometry problems arising in this way are rather subtle; nevertheless there are some cases where it is possible to solve them using powerful technique of modern algebraic geometry whih leads to rather interesting results in coding theory.

In this talk we consider the following results:

- complete determination of the covering radius of BCH-codes of large length (both primitive and non-primitive) ; this uses the Lang-Weil bounds for the number of rational points on the variety over a finite field;
- complete determination of weights of codes orthogonal to certain binary and ternary cyclic codes (the Melas code, certain classical Goppa codes, the Zetterberg code), which reduces to counting rational points of certain elliptic and hyperelliptic curves;
- complete determination of the weight enumerator for some of these codes, which depends on certain calculations with the trace formula for Hecke operators;
- direct computation of the weight of certain subcodes of second order Reed-Muller codes (without using the MacWilliams identities), which reduces to the study of a family of supersingular Artin-Schreier curves.

DECODING OF ALGEBRAIC-GEOMETRIC CODES

Michael A. Tsfasman
Institute for Information Transmission Problems
Moscow and Centre National de Recherche Scientifique
Marseille, Russia

November 2, 1992

A decade ago the problem of decoding algebraic-geometric codes looked hardly tangible and rather far from algebraic geometry. Both proved to be wrong.

The break-through, started by Justesen during his visit to Moscow in 1988, last year reached the point of decoding algebraic-geometric codes up to half the designed minimum distance.

This illustrious achievement is due to the work of many mathematicians, including Vladut, Skorobogatov, Larsen, Havemose, Elbrond Jensen, Hoholdt, Porter, Krachkovskii, Pellikaan, and Shen, the final result being obtained by Ehrhard, Feng, Rao, and Duursma.

The algorithms we have now are both of reasonable complexity and rather easy to understand. However they do tangle several specifical difficulties of algebraic geometry nature.

In this talk the principal points of these decoding algorithms will be described for the simplest example of the curve being the line, with the difficulties of the general case being pointed out on the way.

DECODING ALGEBRAIC-GEOMETRIC CODES UP TO (D-1)/2 ERRORS

Dirk Ehrhard
University Duesseldorf
Germany

November 5, 1992

ABSTRACT

We present an equivalent form of the decoding algorithm in [2]. It achieves the designed minimum distance in Decoding Algebraic-Geometric Codes. For a wide class of such Codes the algorithm is described in an elementary way with a minimum of Algebraic Geometry concepts.

1. V.D. Goppa, Codes on algebraic curves, Soviet Math. Dokl., vol. 24, pp. 170-172, 1981.

2. D. Ehrhard, Achieving the designed error capacity in Decoding Algebraic Geometric Codes. To appear in IEEE-Transactions on Information Theory.

CHANNEL CODING STRATEGIES FOR CELLULAR RADIO

Gregory J. Pottie
Electrical Engineering Dept.
University of California, Los Angeles
405 Hilgard Ave.
Los Angeles, CA 90024

A. Robert Calderbank
Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Ave.
Murray Hill, NJ 07974

ABSTRACT

To improve re-use of time/frequency slots in a cellular radio system, it is desirable for the average interference levels seen by all users to be made approximately equal. We provide constructions based on orthogonal latin squares that guarantee different sets of users to interfere in successive slots. We illustrate how this may be combined with convolutional coding to provide large performance improvement with low delay in a slow hopped system.

SUMMARY

In mobile cellular radio, the dominant impairments are multipath fading and interference from other mobiles. In conventional TDMA systems, mobiles are assigned slots which they keep from frame to frame. The interfering mobiles are assigned slots in the same way. As interference levels vary widely between slots, the result is that some mobiles suffer from persistently poor SNR. Systems are generally designed for 90% or 99% worst case conditions. Therefore, the result of this uneven interference distribution is overly conservative restrictions on frequency re-use between cells, and thus reduced capacity.

If instead the slot assignments were arranged such that different interferers were encountered in successive frames or slots, then the worst case error statisitics would improve, particularly in combination with channel coding across the slots or frames. A number of recent papers [1,2,3] have proposed randomizing the interference with beneficial results. We provide specific constructions that lead to good performance with low delay.

The allocation of time/frequency slots to different users in the same cell, and to users in neighbouring cells is a combinatorial problem of some delicacy. We begin by considering a frame to be an n x n array where the rows correspond to frequency slots, the columns to time slots, and the array entries to different users. Each user in each cell has an individual hopping pattern, and the symbol denoting that user occurs exactly once in each row and column of the frame, as for example in frequency-hopped systems. Thus, it is possible to accomodate n different users in each cell. The combinatorial problem is to allocate hopping patterns in neighbouring cells so that two users in these neighbouring cells interfere in at most one time-frequency slot. We show that if n is a prime power then there are n-1 ways of allocating time-frequency slots with the desired interference properties. The construction is based on mutually orthogonal latin squares. Combinatorial designs associated with latin squares lead to allocation strategies for TDMA systems and for TDMA systems with frequency diversity.

Use of these allocation strategies results in independent interference levels across slots, and therefore channel codes may be used to provide diversity protection against the resulting variations in signal to interference ratio. If in addition the slots are at different frequencies, the code will provide frequency diversity protection against fades in the signal level. The diversity protection lowers the signal to interference ratio threshold required for reliable operation, permitting re-use of all slots in neighbouring cells. As in CDMA systems, interference levels will now directly depend on the number of users, with some back-off from the maximum required for acceptable performance. The reduction in the number of users implied by the use of a rate 1/m code, m an integer does not materially affect the capacity provided m is reasonable, since coding can take the form of occupying m slots with reduced power. The interference power is unchanged from the case of transmitting at the nominal power using 1 in m slots. Another benefit of coding is increased resistance to noise, and consequently the average transmitter power requirements are reduced.

Compressed speech presents a challenging channel coding problem. Delay is a critical parameter, with the maximum acceptable delay on the order of 20 to 40 ms. For 8 kb/s speech with a 20 ms delay, there are only 160 information bits. In this time, reasonable frequency and interferer diversity must be achieved, along with coding gain. We analyze a slow frequency hopped TDMA approach involving convolutional codes and differential QPSK. For slots of 8 to 16 information bearing signals, it is very difficult to estimate the signal to interference ratio (C/I) in the presence of rapid multipath fading. Due to the uncertainty in this estimate, soft decision decoding is in some cases outperformed by hard decision decoding combined with an erasure-declaring mechanism. Moreover, for two antenna branches, selection diversity performs quite well relative to combining based on C/I. Our results indicate that a conventional TDMA system with coding is inferior to the CDMA system proposed in [4] at all reasonable outage probabilities, if re-use of all frequencies is attempted in every cell. However, a slow-hopped system using the method of orthogonal latin squares yields as much as a factor of 2 in increased capacity depending on the particular assumptions made about the propagation environment.

Unequal error protection may be of use with compressed speech, since not every bit has an equal effect on the perceived quality of the reconstructed speech. We present an example of how this might be achieved with low delay and minimal extra complexity cost. We also discuss coding in slow fading environments.

[1] B. Gudmundson, J. Skold, and J.K Ugland, "A comparison of CDMA and TDMA Systems," Proc. 1992 IEEE Vehic Tech. Conf., Denver, May 1992, pp. 732-735.

[2] H. Sasaoka, "Block Coded 16-QAM/TDMA cellular radio system using cyclical slow frequency hopping,", Proc. 1992 IEEE Vehic. Tech. Conf., Denver, May 1992, pp. 405-408.

[3] N.R. Livneh et al., "Frequency hopping CDMA for digital radio," Proc. Int'l. Commsphere Symp., Herzliya, Israel, Dec. 1991.

[4] K.S. Gilhousen et al., "On the capacity of a cellular CDMA system," IEEE Trans. Vehic. Tech., May 1992, pp. 303-312.

A Comparison of CDMA and Frequency Hopping in a Cellular Environment

Michael I. Mandell and Robert J. McEliece
California Institute of Technology, 116-81, Pasadena, CA 91125

Abstract

This paper compares the performances of Direct Sequence Code Division Multiple Access (CDMA) and Frequency Hopping (FH) schemes in a cellular multiuser environment. Our multiuser channel model incorporates the effects of propagation, frequency selective fading, and interference among users in the presence of a constrained system bandwidth. The CDMA and FH systems are compared using BPSK modulation. The main point of contrast between these systems is that the orthogonal hopping patterns in a FH system result in a decreased additive interference power, however the frequency spreading nature of CDMA results in the ability to combat fading. An information theoretic analysis is presented, which shows that system capacity is larger for CDMA than for FH. Hence, for this channel, with sufficient coding the CDMA system can achieve a higher level of performance than the FH system. However, it is unclear what level of complexity would be required to achieve such performance, and what effect such complexity would have on the practicality of the system.

Summary

In this paper we compare the performances of direct sequence Code Division Multiple Access (CDMA) and Frequency Hopping (FH) in a cellular multiuser environment. We assume that there is a fixed system bandwidth, B, and a fixed data rate, R, at which each user communicates. The normalized traffic of a system, p, is defined to be the number of users per sector, N_s, divided by the ratio of B to R. We find that the FH system sees less interference power than the CDMA system, however, the FH system is susceptible to frequency selective fades whereas the wide band nature of CDMA offers a level of diversity to such fading. Thus, a tradeoff in performance exists and the FH system performs better at higher levels of traffic with relatively high probability of bit error, and the CDMA system performs better at lower levels of traffic, with relatively low probability of bit error.

At this point, we consider the use of coding and present an information theoretic analysis. Assuming that there is no cooperation among the users in the system on the level of coding, the capacity of the system is defined to be the largest possible value of normalized traffic, p, for which each user in the system can communicate reliably at rate R. The capacity of FH and CDMA are computed and we find that the CDMA system has a larger capacity. This is due to the fact that the the FH system does not allow for the use of very low code rates because using a low rate code a relatively small number of users would occupy the entire system bandwidth and thus result in a small amount of traffic. It turns out that, from an information theoretic viewpoint, that the ability of CDMA to use low rate codes is an advantage over the lower interference power in the FH system.

Since the information theoretic results are obtained over arbitrarily complicated coding schemes, and thus, arbitrarily long delays, we investigate the performance of specific coding schemes and the effects of a finite, controlled, delay using interleaving. These results are obtained primarily through simulation. Making some assumptions about the vehicle speed and transmitter frequency, we evaluate performance using a finite amount of interleaving delay by taking into account the actual amount of correlation among channel samples as seen from one codeword symbol to the next. We have evaluated the performance of several repetition codes as well as an (8,4) bi-orthogonal block code. These coding schemes perform far below information theoretic capacity, and yield performance curves for FH and CDMA that cross with FH performing better higher levels of traffic with relatively high probability of bit error, and the CDMA performing better at lower levels of traffic, with relatively low probability of bit error.

* This work is supported by grants from GTE Laboratories and Pacific Bell, and AFOSR Grant 91-0037.

References

[1] Wallace, Mark S., "High Capacity Digital Cellular Communications Through Slow Frequency Hopping CDMA", Proc. 29th Annual Allerton Conference on Communication, Control, and Computing, pg. 21, 1991.

[2] Verhulst, D., M. Mouley, and J. Szpirglas, "Slow Frequency Hopping Multiple Access for Digital Cellular Radiotelephone", IEEE Journal on Selected Areas in Communications, vol. SAC-2, no. 4 pp. 563–574, July 1984.

[3] Simon, M.K., J. K. Omura, R. A. Scholtz, and B. K. Levitt, "Spread Spectrum Communications", Vol. 2, MD: Computer Science Press, 1985.

[4] Jakes, W. C. Jr., "Microwave Mobile Communications", New York: Wiley, 1974.

[5] Proakis, John G., "Digital Communications", New York: McGraw Hill, 1989.

[6] Gilhousen, K. S., I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, and C. E. Wheatley, "On The Capacity Of A Cellular CDMA System", IEEE Trans. Veh. Tech., vol. 40, no. 2, May 1991, pp. 303–312.

CAPACITY OF COHERENT FREQUENCY-HOP SPREAD-SPECTRUM COMMUNICATIONS

Giovanni Cherubini[1] and Wayne Stark[2]

[1]*IBM Research Division, Zurich Research Laboratory, CH-8803 Rüschlikon, Switzerland*
[2]*University of Michigan, EECS Department, Ann Arbor, MI 48109-2122*

ABSTRACT

The capacity of a coherent frequency-hopped (C-FH) spread-spectrum system is investigated. The channel comprises an M-ary phase-shift keying (M-PSK) modulator, a frequency hopper with phase-continuous carrier, the transmission medium, a nonideal phase-coherent frequency dehopper, an M-PSK demodulator, and a carrier tracking system. Additive white Gaussian noise is considered. The analysis focuses on the effect of imperfect recovery of the carrier phase on the demodulation process. The carrier tracking system includes a maximum likelihood estimator of the phase error and a first-order digital phase-lock loop. The phase error is modeled as a Markov process. An expression for the state transition probabilities of the phase error process is given. Using bounds on the entropy of a function of a Markov process, lower bounds to the capacity of C-FH channels are derived. The input symbols are assumed uniformly distributed and the encoding process independent of the frequency slot selected to send each symbol. Numerical results obtained for various values of M and of the number of frequency slots are presented.

SUMMARY

In this paper, the capacity of a frequency-hopped spread-spectrum communication system with coherent demodulation is investigated. It is known that large gains over systems employing noncoherent demodulation are attainable if trellis coded M-ary phase-shift keying (M-PSK) and coherent demodulation with maximum-likelihood sequence detection are adopted [1]. In frequency-hopped spread-spectrum systems, however, a major obstacle to coherent demodulation is represented by the difficulty for the receiver to maintain phase coherency between the carrier of the incoming signal and a locally generated waveform.

In early work on coherent frequency-hopped (C-FH) spread-spectrum communication systems, ideal carrier tracking was assumed [2,3]. In more recent work, it was proposed to generate the C-FH signal by phase-modulating the harmonics of a reference sinewave [4,5]. The carrier tracking system recovers the phase of the reference sinewave to dehop the received signal coherently, provided the phase relationships between the reference sinewave and each of its harmonics are known. The feasibility of such a method has been demonstrated for low signal-to-noise ratios and binary PSK modulation.

We extend the approach described in [4,5] to a channel that comprises an M-PSK modulator, a frequency hopper with phase-continuous carrier, the transmission medium, a nonideal phase-coherent frequency dehopper, an M-PSK demodulator, and a carrier tracking system. Additive white Gaussian noise is considered. The carrier tracking system includes a maximum-likelihood estimator of the phase error between the transmit and receive reference sinewaves, and a first-order digital phase-lock loop. The analysis focuses on the effect of imperfect phase recovery on the demodulation process. Assuming that the frequency of the reference sinewave is perfectly known, it is shown that the phase error (Φ_n) can be modeled as a Markov process. The state transition probabilities of the process (Φ_n) are evaluated.

We consider two channels: one for which the input to the decoder is just the channel output sequence Y_n and one for which in addition the frequency-hop pattern L_n is known to the decoder. The capacity of the channel which outputs the sequence Y_n only is given by

$$C = \lim_{n\to\infty} \max_{X^{(n)}} \frac{1}{n} I(X^{(n)}; Y^{(n)}) , \tag{1}$$

where $X^{(n)} = (X_1, X_2, \dots, X_n)$ is the vector of channel input symbols, $Y^{(n)} = (Y_1, Y_2, \dots, Y_n)$ is the vector of channel outputs, and $I(X^{(n)}; Y^{(n)})$ is the average mutual information between the random vectors $X^{(n)}$ and $Y^{(n)}$, which can be expressed as the difference between the entropy of $Y^{(n)}$ and the conditional entropy of $Y^{(n)}$ given $X^{(n)}$, i.e.,

$$I(X^{(n)}; Y^{(n)}) = H(Y^{(n)}) - H(Y^{(n)} \mid X^{(n)}) . \tag{2}$$

The maximum in (1) is over all distributions of the random vector $X^{(n)}$. We do not allow this distribution to depend on the random vector $L^{(n)}$, i.e., the encoding process is independent of the frequency slot used to send each symbol. In addition, we observe that each symbol is used with the same probability in most codes of practical interest. Thus we restrict ourselves to the computation of the mutual information for a uniform distribution of the input symbols. Since the channel is not memoryless, capacity is not trivial to calculate. However, bounds on the entropy of a random process which is a function of a Markov process are known [6,7]. Since (Y_n, L_n, Φ_n) is a Markov process, it follows that (Y_n) is a function of a Markov process. If we let

$$H(Y) = \lim_{n\to\infty} \frac{1}{n} H(Y^{(n)}) , \tag{3}$$

$$H(Y \mid X) = \lim_{n\to\infty} \frac{1}{n} H(Y^{(n)} \mid X^{(n)}) , \tag{4}$$

then the following bounds can be applied, for $n = 1, 2, \dots$,

$$H(Y_n \mid Y_{n-1}, \dots, Y_1, Y_0, \Phi_0, L_0) \le H(Y) \le H(Y_n \mid Y_{n-1}, \dots, Y_1) , \tag{5}$$

and

$$\begin{aligned} H(Y_n \mid Y_{n-1}, \dots, Y_1, Y_0, \Phi_0, L_0, X_n, \dots, X_1) &\le H(Y \mid X) \\ &\le H(Y_n \mid Y_{n-1}, \dots, Y_1, X_n, \dots, X_1), \end{aligned} \tag{6}$$

which can be computed by using the known state transition probabilities of the phase error process. If, in addition to Y_n, L_n is also available to the decoder, the capacity becomes

$$C = \lim_{n\to\infty} \max_{X^{(n)}} \frac{1}{n} I(X^{(n)}; Y^{(n)}, L^{(n)}) . \tag{7}$$

Expressions similar to (5)-(6) can be used to bound the capacity given by (7).

Numerical results showing lower bounds to the capacity of a C-FH channel are presented for various values of M and of the number of frequency slots.

References

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, Vol. IT-28, pp. 55-67, Jan. 1982.

[2] M. K. Simon and A. Polydoros, "Coherent detection of frequency-hopped quadrature modulation in the presence of Jamming -- Part I: QPSK and QASK modulations, " *IEEE Trans. Commun.*, Vol. COM-29, pp. 1644-1660, Nov. 1981.

[3] W. E. Stark, "Coding for coherent frequency-hopped spread-spectrum communications in the presence of jamming," Proceedings of the 1982 IEEE Military Communications Conference, Vol. 1, pp. 14.2.1--5, October 18-20 1982.

[4] G. Cherubini and L. B. Milstein, "Performance analysis of both hybrid and frequency hopped phase-coherent spread-spectrum systems -- Part II: An FH system," *IEEE Trans. Commun.*, vol. COM-37, pp. 612--622, June 1989.

[5] C. M. Su and L. B. Milstein, "Analysis of coherent frequency hopped spread-spectrum receiver in the presence of jamming," *IEEE Trans. Commun.*, vol. COM-38, pp. 715--726, May 1990.

[6] J. J. Birch, "Approximations for the entropy for functions of Markov chains," *Annals of Mathematical Statistics*, vol. 33, pp. 930--938, 1962

[7] T. M. Cover and J. A. Thomas, "Elements of Information Theory," *Wiley Interscience*, New York, 1991.

ERLANG CAPACITY OF A POWER CONTROLLED CDMA SYSTEM

Audrey M. Viterbi and Andrew J. Viterbi

QUALCOMM, Incorporated
10555 Sorrento Valley Road
San Diego, CA 92121

Summary

For any multi-user communication system, the measure of its economic usefulness is not the maximum number of users which can be served at one time, but rather the peak load that can be supported with a given quality and with availability of service as measured by the blocking probability. This is the probability that a new user will find all channels busy and hence be denied service, generally accompanied by a busy signal. Adequate service is usually associated with a blocking probability of 2% or less. The average traffic load in terms of average number of users requesting service resulting in this blocking probability is called the Erlang capacity of the system.

In virtually all existing multi-user circuit-switched systems, blocking occurs when all frequency slots or time slots have been assigned to a voice conversation or message. In code division multiple access (CDMA) systems in contrast, users all share a common spectral frequency allocation over the time that they are active. Hence, new users can be accepted as long as there are receiver-processors to service them, independent of time and frequency allocations. We assume that a sufficient number of such processors are provided in the common base station such that the probability of a new arrival finding them all busy is negligible. Rather, blocking in CDMA systems will be defined to occur when the interference level, due primarily to other user activity, reaches a predetermined level above the background noise level of mainly thermal origin. While this interference-to-noise ratio could, in principle, be made arbitrarily large, when the ratio exceeds a given level (about 10 dB nominally), the interference increase per additional user grows very rapidly, yielding diminishing returns and potentially leading to instability. Consequently, blocking in CDMA is defined as the event that the total interference-to-background noise level exceeds a given threshold and we determine the Erlang capacity which results in a given probability of this event (e.g. 1%). We emphasize, however, that this is a "soft blocking" condition, which can be relaxed as will be shown, as contrasted to the "hard blocking" condition wherein channels are all occupied.

Also, in conventional systems a fraction of the time or frequency slots must be set aside for users to transmit requests for initiating service and a protocol must be established for multiple requests when two or more users collide in simultaneously requesting service. In CDMA systems even the users seeking to initiate access can share the common medium. Of course, they add to the total interference and hence lower the Erlang capacity to some degree. We demonstrate that this reduction is very small for initial access requests whose signaling time is on the order of a few percent of the average duration of a call or message.

CODING DECREASES DELAY OF MESSAGES IN NETWORKS

Grigoriĭ A. Kabatianskiĭ
Institute for Problems
of Information Transmission
Ermolovoy 19, Moscow GSP-4
101447 Russia
E-mail: kaba@ippi.msk.su

and

Eugenĭ A. Krouk
Leningrad Aircraft
Instrumentation Institute
Hertzena 67, St. Petersburg
Russia

We consider an application of codes correcting errors and erasures for decreasing delay of messages in networks with datagram service. Let any message consists of k packets and the sender adds r redudancy packets in such a way that all $n = k+r$ packets together form a codeword of some Q-ary code with minimum Hamming distance d, where $Q = 2^m$ and m is the length of any packet in bits. Then the receiver can recover a message immediately after obtaining the first $n - d + 1$ packets, because he considers the rest $d - 1$ packets, which are not yet arrived, as erasures and corrects them. In particulary, the receiver can recover a message after obtaining the first k packets, if Reed-Solomon codes are used [1].

Denote by t_i the delay of the i-th message and assume that the delays $t_1, t_2, ..., t_n$ are independent identically distributed random variables. Denote by $t_{i:n}$ the i-th order statistic of the sample $(t_1, t_2, ..., t_n)$, i.e. $t_{1:n} \leq t_{2:n} \cdots \leq t_{n:n}$. Then the delay of a message equals $T = t_{k:k}$ for ordinary procedure and equals $T^{(R)} = t_{k:n}^{(R)}$ for described above procedure when R-S code of rate $R = k/n$ are used.The superscript R shows also that one had to recalculate the packet delay, because the average customer arrival rate λ increases in $1/R$ times. We suppose (see [2]) that the average packet delay equals

$$\mathbb{E}[t] = a/(1-\rho), \qquad (1)$$

where $\rho = \lambda/\mu$ is the utilisation factor and a is some constant depending on a given structure of network and fixed proportions of input flows. We also assume that the delay of any packet is exponentialy distributed. It is well known that for this distribution the average value of the k-th order statistic equals

$$\mathbb{E}[t_{k:n}] = \mathbb{E}[t] \cdot \sum_{i=n-k+1}^{n} i^{-1}.$$

Using this fact and the assumption (1) and by putting $R = 2\rho/(1+\rho)$, one gets that the procedure of encoding messages certainly gives a gain if

$$2 \ln \frac{1+\rho}{1-\rho} \leq C + \ln k,$$

where C is the Euler constant. We generalize this result for: networks with "impatient" messages; nonreliable networks; networks with "time-out" procedure.

References

[1] N. F. Maxemchuk, "Dispersi··· routing", Proceedings IEEE Conf.Commun. ,San Francisco, 1975, N.Y., 1975, vol.3.

[2] L. Kleinrock, Communication nets, N.Y, Dover, 1964.

The Performance of Frequency Comb Multiple Access(FCMA) In Interference Limited and AWGN Environments

T.J. Stevenson Student Member IEEE, K.W. Yates Senior
Member IEEE, University of Technology, Sydney, Australia

Abstract. The performance of FCMA is analysed in environments where the dominant sources of interference are firstly, the multiple access noise of other users and secondly, AWGN. Ultimately, multiple access noise is the limiting factor in performance. The intended transmission to the selected addressee is always symbol synchronous and phase coherent. The interferers are observed in one instance when they are symbol synchronous and phase coherent and again while symbol synchronous but noncoherent in phase. The symbol synchronous phase coherent interferers are found to represent the worst case performance.

Summary. FCMA was first proposed by Stevenson et al [1], as a new form of multiple access for packet satellite communications.

In the proposed system, users share the resource on a nonorthogonal basis, as is done in variants of Code Division Multiple Access, such as Frequency Hopped Multiple Access (FH/CDMA) and Direct Sequence Multiple Access (DS/CDMA). However instead of user signatures occupying a time varying narrow band (FH/CDMA) or the full bandwidth continuously (DS/CDMA), signatures consist of quasi-orthogonal combs of frequencies. For this reason the scheme is termed ***Frequency Comb Multiple Access (FCMA).*** Signature combs interleave one another over the available bandwidth and provide the basis for both multiple access and information transmission. FCMA belongs to the same family as Random Multiple Access (RMA)[2], with the difference that code signatures are restricted to combinations of discrete frequencies which are *on* for the complete symbol duration. This is in contrast to RMA, where signatures are combinations of energy elements, from the (symbol) time-(available) bandwidth plane.

A unique feature of FCMA, is the manner in which addressing and information conveyance is jointly accomplished.

Each user has a look up table giving the signature sets of all other users in the system. The transmitter can then be programmed to use the signature sets of any other user in the system and hence to communicate with any other user.

The selected addressee has a receiver tuned to its own signature set. There is also considerable advantage in being able to monitor other channel transmissions, as this will determine if the intended addressee's receiver is presently being interrogated and hence reduce the possibility of collision. In FCMA, this is a significant aspect of communication, as opposing schemes, such as Aloha, experience collisions whenever two users share the channel simultaneously, whereas with FCMA, this only occurs when two messages are simultaneously directed to the same user.

Performance is first considered in a multiple access noise only environment, where there are *(X-1)* interferers and it is assumed that signatures are generated by randomly selecting n frequencies from an available pool of M, the results can then be easily checked by simulation. The resulting BER when each user has $M=2^k$ signatures and thus conveys k bits per symbol is:

$$P_b = \frac{2^{k-1}}{2^k-1} \sum_{j=1}^{M-1} \frac{j}{j+1} \binom{M-1}{j} (P_f)^j (1-P_f)^{(M-j-1)} \qquad (1)$$

$$P_f = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} [1-\frac{\nu}{N}]^\mu$$

Equation (1) then enables BER to be plotted versus both active user numbers and channel utilisation(τ), where for FCMA, $\tau = Xk / N$ (bits / hertz).

When signatures with controlled overlap are used, for example where any two signatures have at most one element in common, there is a minimum number of simultaneous users X before errors occur in the absence of AWGN.

Performance in AWGN was also investigated. It was assumed that maximum likelihood detection was used together with signatures having at least one common element Yates [3] and Wu [2]. The system was assumed to be symbol synchronous with all users' frequency combs aligned in phase. The derivation is from first principles but is omitted here due to space limitations.

A comparison of simulation and analytic results for the above case shows close agreement and establishes an upper bound to performance. The claim that this is an upper bound is justified by the fact that inphase interferers give worst erosion of distance between symbols.

Simulation results indicate that noncoherent interferers have a significant performance margin over coherent interferers. This is an intuitively satisfying result, as noncoherent interferers would generally occur in practise. Because of the distance properties of the signature set chosen, the system is power limited(AWGN dominated) when $X < (n+1)$. When $X >= (n+1)$, there is a value of E_b/N_o at which the system becomes dominated by multiple access noise. For $n=3$ a value of $E_b/N_o > 15$ dB was required and for $n=5$ a value of $E_b/N_o > 19$ dB.
Applications for FCMA include: satellite multiple access in which the traffic is bursty and a guaranteed level of performance is required, ie low data rate VSATS; control channel for DAMA; emergency maritime communications; networks in which the active user population is a fraction of a much larger potential user population; mobile communications; indoor wireless; environments requiring frequency diversity; delay intolerant systems eg speech; inquiry response traffic using short packets.

As signatures convey both address and data information, there are no address overheads and very short packets are viable, since rapid acquisition can be achieved using a short preamble provided an FFT receiver is used. System design involves choices for three parameters N, n, and k. Considerable flexibility is therefore available to meet particular traffic requirements and constraints.

[1] T.J. Stevenson and K.W. Yates, "A New Multiple Access Scheme for Packet Satellite Communications", *ISSSE'89*, Erlangen, West Germany, September 1989.

[2] W.W. Wu, Elements of Digital Satellite Communications, vol.2, Computer Science Press, 1985.

[3] K.W. Yates, "Waveform Encoding in Spread Spectrum Systems", presented at the *23rd URSI Conference*, Prague, 1991 (not published).

Analysis of a Hybrid Random-Access System with Multi-User Coding (Throughput)

Rumaih M. Al-Rumaih and Peter Mathys

Department of Electrical and Computer Engineering
University of Colorado
Boulder,Colorado 80309-0425

Abstract

The performance of hybrid random-access systems (HRAS) which use a combination of multi-user codes (MUC) and collision resolution algorithms (CRA) to accomodate the bursty transmissions of many independent users on a single communication channel is analyzed. Besides the computation of the system throughput , another contribution of this paper is the determination of the properties which a MUC must possess such that resulting HRAS performs better than a system which is only based on CRA's.

I. Introduction

We aim at combining the collision resolution(CR) and the Multi-user information theory (MUIT) approaches [1] by using a Hybrid Random-Access System (HRAS) which uses both a collision resolution algorithm (CRA),e.g., as described in [2] , and a multi-user code (MUC), e.g., as described in [3]. The intutive underlying idea is that small collisions among T or less users (e.g., $T = 2$ or $T = 3$) are "resolved" by using MUC's and large collisions (which are assumed to occur very infrequently) are resolved by using CRA's. A notable improvement of the HRAS over one which uses only CRA's occurs if the roundtrip delay is large. Even though the delay performance is of primary interest when adding coding to a CRA, throughput analysis is necessary and yields some initial results about how much gain coding might offer.

II. Blocked and Free-Access HRAS's

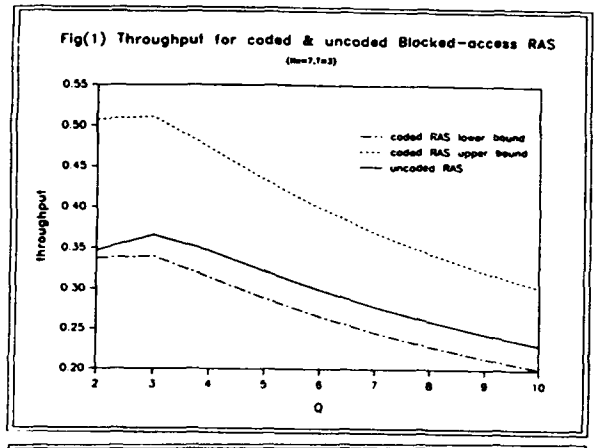
The analysis was done for the Basic Blocked and Free-access channel access protocols(CAP) given in [2] using the capacity of the T user real adder channel [4] as an upper bound for the rate of (N_C, T) codes (a T active out of N_C) and the codes given in [3] as a lower bound. Fig(1) and Fig(2) show the maximum stable throughput for HRAS's(7,3) λ^s (upper and lower bounds) together with the maximum stable throughput for the uncoded RAS's λ_{crit}[2] versus Q (after a collision, each transmitter involved flips a fair "Q-sided coin") for blocked and free-access systems respectively.

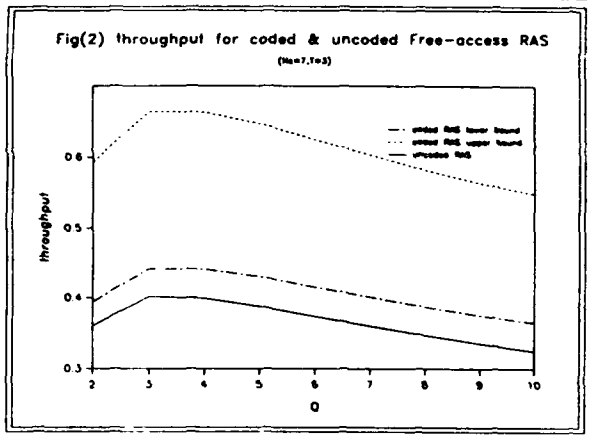
III.Conclusions

1. Maximum stable throughput can be substantialy improved (e.g., with Q =3 for free-access RAS $\lambda_{crit} = 0.4016$, whereas for free-access HRAS(7,3) $\lambda^s_{crit} = 0.664$).

2. $T = 3$ gives the best performance out of all (N_C, T) codes for both Blocked and Free -Access RAS's [2]. Thus, the search for HRAS's can be considerably narrowed, since large T offers no advantage at all.

3.The upper bound of (N_C, T) codes [4]tends to get loose as N_C diverges from T , which might suggest that the capacity of (N_C, T) codes should be a function of N_C. Note that for practical reasons large N_C is not desirable in our case .





References

[1] R.G. Gallager, "A perspective on Multi-Access Channels." IEEE Trans. Info. Theory, vol. IT-31, No. 2, pp. 124-142, Mar. 1985.

[2] P. Mathys and P.Flajolet, "Q-ary Collision Resolution Algorithms in Random-Access Systems with Free or Blocked Channel Access," IEEE Trans. Info. Theory, vol. IT-31, No. 2, pp. 217-243, Mar. 1985.

[3] P. Mathys, "A Class of Codes for a T Active Users Out of N Multiple-Access Communication System," IEEE Trans. Info. Theory, vol. 36, no. 6, pp. 1206-1219, Nov. 1990

[4] S.C.Chang and E.J. Weldon, Jr., "Coding for T-User Multiple-Access Channels," IEEE Trans. Info. Theory, vol. IT-25, No. 6, pp. 684-691, Nov. 1979.

Slow Frequency Hopping Patterns Derived from Polynomial Residue Class Rings

P.Udaya and M.U. Siddiqi
Department of Electrical Engineering
Indian Institute of Technology
Kanpur, 208016 (INDIA)

1.0 Introduction

Frequency hopping is one of the common techniques for spreading the signal spectrum in digital data communication systems. The amount of frequency spread is far more than the minimum bandwidth necessary to transmit the digital data. This fact makes it feasible for many users to share a common channel. This paper is concerned with construction of new families of slow frequency hopping patterns derived from sequences over a given semi–local residue class polynomial ring $P_p^n[w(\xi)]$ (GF(p)[ξ]/w(ξ): Set of polynomials over GF(p) modulo w(ξ)), where w(ξ) is a polynomial of degree n over GF(p).

The frequency library in a slow frequency hopping spread spectrum (SFHSS) system consists of large number of frequency carriers which are chosen to be orthogonal to each other over the transmission time duration T. The carriers are obtained by subdividing the entire bandwidth into contiguous frequency slots. For multiple–access purposes, each user is provided with a distinct hopping pattern of period L. Each symbol in a hopping pattern is drawn from the frequency library and determines the frequency band within which transmission takes place.

Correlation requirements on the patterns:

Normally in frequency hopping systems, it is required that mutual Hamming correlation between sequences should be small. In SFHSS systems, one or more symbols are transmitted within one frequency hop (slot) and a hit would mean total loss of data transmitted in that hop [2,3]. Thus, apart from minimising mutual Hamming correlation between patterns, hits resulting from presence of all the patterns in the system should be *minimised*. This prompt us to define generalised Hamming correlation functions which depend on all the sequences in a family, unlike Hamming correlations which depend on only on two sequences.

Let S^m, m = 1,..,n, be n sequences of length L over certain alphabet $\mathcal{A}$, then the generalised Hamming cross–correlation function concerning m^{th} sequence is given by

$$GHC(\tau_1,\tau_2,\ldots,\tau_{m-1},\tau_{m+1},\ldots,\tau_n) = \sum_{i=0}^{L-1} gh\{S^m_i; S_{i+\tau_j}, \text{for all } j\neq m\},$$

where gh is a function given by

$$gh(a; b_1,b_2,\ldots,b_n) = \begin{cases} 1 \text{ if } a \in \{b_1, b_2, \ldots, b_n\} \\ 0 \text{ otherwise} \end{cases}.$$

The corresponding autocorrelation function is given by.

$$GHA_m(\tau_1,\tau_2,\ldots,\tau_n) = \sum_{i=0}^{L-1} gh\{S^m_i; S^j_{i+\tau_j}, \text{for all } j\}$$

For proper multi–user operations, patterns for SFHSS systems should have ideal GHC properties (Crosscorrelation function is equal to zero for all values of τ_i).

2.0 Main Results

The frequency hopping patterns are obtained by associating with each symbol a in the ring $P_p^n[w(\xi)]$, a distinct frequency f_a belonging to frequency library. Properties of orthogonal ideals of $P_p^n[w(\xi)]$ and the internal direct sum representation of the ring $P_p^n[w(\xi)]$ have been made use of to define frequency hopping patterns. Let $w(\xi) = w_1(\xi)\, w_2(\xi)$, where $w_1(\xi)$ and $w_2(\xi)$ are relatively prime polynomials of degrees n_1 and n_2 respectively; $n = n_1 + n_2$. Then $P_p^n[w(\xi)]$ can be represented as internal direct sum of ideals isomorphic to rings $P_p^{n_1}[w_1(\xi)]$ and $P_p^{n_2}[w_2(\xi))]$. Let $e_1(\xi)$ and $e_2(\xi)$ be orthogonal idempotent polynomials in $P_p^n[w(\xi)]$ corresponding to rings $P_p^{n_1}[w_1(\xi)]$ and $P_p^{n_2}[w_2(\xi))]$. Then elements of the ideals $< e_1(\xi) >$ and $< e_2(\xi) >$ mutually annihilate each other [4]. Thus elements of the cosets of the ideal $< e_1(\xi) >$ in $P_p^n[w(\xi)]$ are all distinct. Sequences are defined in such a way that elements of a sequence belong to a distinct coset. Since these cosets are mutually exclusive (there is no common element among these cosets), ideal generalised Hamming correlation properties follow naturally. Construction of slow hopping sequences makes use of optimal frequency hopping sequences over local rings derived in [1]. Following new families are derived.

1. A family of p^{n_2} sequences of period $L = p^{n_1}-1$ over $P_p^n[w(\xi)]$, where $n=n_1+n_2$, by using a sequence over $P_p^{n_1}[w_1(\xi)]$ of period $p^{n_1}-1$, where $w_1(\xi)$ is an irreducible factor (of degree n_1) of $w(\xi)$. These sequences satisfy ideal generalised Hamming correlation properties.

2. A family of μp^{n_2} sequences of period $L = p^{n_1}-1$ over $P_p^n[w]$, $n=n_1+n_2$, by using μ one–coincidence sequences over $P_p^{n_1}[w_1(\xi)]$ each of period $p^{n_1}-1$. The generalised Hamming cross and auto correlations for any sequence in the family are given by

$$GHAC_m(\tau_1,\tau_2,\ldots,\tau_{m-1},\tau_{m+1},\ldots,\tau_n) \le \begin{cases} p^{n_1}-\mu \text{ for } \tau_j = 0 \\ \mu-1 \text{ otherwise} \end{cases}$$

$$GHCC_m(\tau_1,\tau_2,\ldots,\tau_n) \le \mu-1 \text{ for all } \tau_i \neq \tau_j.$$

A code generation scheme, based on the direct sum decomposition of semi–local rings, for slow hopping multiple access communication systems is given where different users can have different frequency diversity.

REFERENCES:

[1] Udaya P, "Polyphase and Frequency Hopping Sequences obtained from Finite Rings", Ph.D Thesis, Department of Electrical Engineering, I.I.T Kanpur, 1992.

[2] M. K. Simon, J. K. Omura, R. A. Scholtz, B. K. Levit, "Spread Spectrum Communications", Vol. 1, Computer Science Press, 1985.

[3] Bernard Sklar, "Digital Communications", Chapter 10, Prentice–Hall, Englewood Cliffs, 1988.

[4] Hari Bhat, "Linear Sequential Systems over Residue Class Polynomial Rings: Theory and Applications", Ph.D Thesis, Department of Electrical Engineering, I.I.T, Kanpur, 1986.

BOUNDS ON THE CAPACITY OF AN AWGN CHANNEL WITH INTERTRANSITION CONSTRAINED BIPOLAR INPUTS

Shlomo Shamai (Shitz) and Naftali Chayat
Department of Electrical Engineering
Technion - Israel Institute of Technology, Haifa 32000, Israel

We present lower and upper bounds on the capacity of an AWGN channel, the input to which is a bipolar (± 1) waveform with a constraint that the minimum duration between transition is no shorter than $T_{\min}$.

This model is used to characterize certain magnetic recording channels where bipolar signaling is preferred due to the hysteresis phenomenon of the magnetic media and the minimal intertransition duration constraint is imposed as to mitigate the heavy (possibly nonlinear) intersymbol interference effects.

The upper bounds are based on Duncan's formula that interrelates the average mutual information to the average minimum mean-square error (MMSE) of the causal optimal estimator. To this end the MMSE of suboptimal linear and nonlinear estimators is studied [1] and the guard-time random telegraph signal [2] is also considered.

The lower bounds are found by considering bipolar runlength limited (d, ∞) codes where d (integer) is related to the minimal intertransition constraint by $d = T_{\min}/\Delta$ and where Δ stands for the duration of the bipolar channel symbol. The asymptotic ($d \to \infty$) expression for the entropy of the max-entropic (d, ∞) bipolar sequences is invoked along with a recent extension of Mrs. Gerber's Lemma [3] (to account for any binary input-output symmetric channel) to yield the lower bounds, which are optimized with respect to $d \gg 1$ (or equivalently $\Delta = T_{\min}/d$). Pulse amplitude and pulse width modulated signals are also considered in the context of lower bounding the capacity [1].

It is concluded that the capacity behaves as $1/N_o$ (nats/sec) for $SNR \stackrel{\Delta}{=} T_{\min}/N_o \ll 1$ and as $T_{\min}^{-1} \ln\left(\frac{SNR}{\ln SNR}\right)$ (nats/sec) for $SNR \gg 1$, where N_o denotes the power spectral density of the AWGN.

Lower bounds on the capacity with the aforementioned constrained inputs in the presence of a mildly band-limited (in scales of $T_{\min}^{-1}$) channel filter are presented. Explicit expressions are found by generalizing the recently introduced Shamai-Ozarow-Wyner lower bound on the capacity of a dispersive discrete-time Gaussian channel with iid inputs [4], to account for dependen. inputs and incorporating in the generalization a convexity property which is implied by the extended Mrs. Gerber's Lemma.

References

[1] S. Shamai (Shitz) and N. Chayat, "Bounds on the Information Rates of Binary Intertransition Constrained Inputs over the AWGN Channel", EE Report No. 864, Technion, Haifa 32000, Israel.

[2] I. Bar-David and S. Shamai (Shitz), "On Information Transfer by Envelope Constrained Signals over the AWGN Channel", *IEEE Trans. on Inform. Theory*, Vol. **34**, No. 3, pp. 371–379, May 1988.

[3] N. Chayat and S. Shamai (Shitz), "Extension of an Entropy Property for Binary Input Memoryless Symmetric Channels," *IEEE Trans. on Inform. Theory*, Vol. **35**, No. 5, pp. 1077–1079, September 1989.

[4] S. Shamai (Shitz), L.H. Ozarow and A.D. Wyner, "Information Rates for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs," *IEEE Trans. on Inform. Theory*, Vol. **37**, No. 6, pp. 1527–1539, November 1991.

On Capacity of Frequency Non-Selective Slowly Time-Varying Fading Channel

ROGER S. CHENG
Department of Electrical and Computer Engineering
University of Colorado, Boulder, CO 80309
chengr@spot.colorado.edu

Abstract

We consider a frequency non-selective slowly time-varying Rayleigh fading code-division multiaccess (CDMA) additive white Gaussian noise (AWGN) channel. Assuming that the signature waveforms are time-limited to the symbol interval, we find the capacity region of the two-user symbol-synchronous channel. If the signature waveforms span several symbol intervals, we have to further assume that the fading process is constant over the duration of every codeword. In that case, we find the capacity and the optimal input power spectral density (PSD) in a parametric expression similar to that in the classical water-filling argument.

Summary

We consider a frequency non-selective slowly time-varying Rayleigh fading CDMA AWGN channel

$$y(t) = \sum_i X_{1i}a_1(t)s_1(t - iT) + X_{2i}a_2(t)s_2(t - iT) + n(t), \quad (1)$$

where the signature waveforms of the users, $s_1(t)$ and $s_2(t)$, are unit-energy functions strictly time-limited to $[0, LT]$ for some finite L, and $n(t)$ is the zero-mean complex white Gaussian noise with independent real and imaginary parts, each with power spectral density N_0 (i.e., $\mathbf{E}n(t)n^*(t-\tau) = 2N_0\delta(\tau)$.) The power constraints on the users require that every length-n codeword of the k^{th} user has average power at most equal to W_kT. We assume that the channel is a slowly time-varying channel in the sense that $a_k(t) = a_{ki}$ for $t \in [iT, (i+1)T]$, $k = 1, 2$, and $\{a_{1i}\}$ and $\{a_{2i}\}$ are two independent zero-mean stationary m-dependent complex Gaussian fading processes having independent real and imaginary parts. The autocorrelation function of the fading process $\{a_{ki}\}$ is denoted by $\sigma_k^2 R_k(i)$ where σ_k^2 is the power of $\{a_{ki}\}$ (i.e., $R_k(0) = 1$.) Finally, we assume that the receiver has complete knowledge of the fading processes, but the transmitter knows only the statistics of the fading processes.

The corresponding single-user channel with $s(t) \in L_2[0,T]$ (i.e., $L = 1$) is equivalent to the discrete-time frequency non-selective Rayleigh fading AWGN channel $Y_i = a_iX_i + N_i$. We have omitted the subscripts for the users in the single-user case. Since the receiver knows the fading parameters, the channel is equivalent to a stationary channel with output (Y_i, a_i) whose capacity is [1]

$$\frac{1}{T}C(\lambda) = \frac{1}{T}\mathbf{E}\log\left[1 + \frac{WTR}{2N_0}\right] = -\frac{1}{T}\text{Exp}\left(\frac{1}{\lambda}\right)\text{Ei}\left(-\frac{1}{\lambda}\right)$$

where the expectation is taken over $R = |a_i|^2$ which is exponentially distributed with mean σ^2, and $\lambda = WT\sigma^2/(2N_0)$ is the average received signal-to-noise ratio. The function $\text{Ei}(\,)$ is the exponential-integral function and the last expression follows from [2, p. 574].

We extend the above single-user result in two directions: (1) the capacity region of the two-user channel with $s_k(t) \in L_2[0,T]$, and (2) the capacity of the single-user channel with $s(t) \in L_2[0, LT]$ and *very slowly* time-varying fading process.

When the signature waveform are time-limited to the symbol interval, the CDMA channel reduce to a discrete-time frequency non-selective fading multiaccess channel, $\mathbf{Y}_i = \mathbf{H}\mathbf{A}_i\mathbf{X}_i + \mathbf{N}_i$, where $\mathbf{A} = \text{diag}[a_{1i}a_{2i}]$, $\mathbf{E}\mathbf{N}_i\mathbf{N}_j^H = \mathbf{H}\delta_{ij}$, $\mathbf{X}_i = [X_{1i}X_{2i}]^T$, and $\mathbf{H}$ is the crosscorrelation matrix of the signature waveforms. The capacity region of this channel is given by the following theorem.

Theorem 1

The capacity region of the two-user multiaccess channel in (1) with $s_1(t), s_2(t) \in L_2[0,T]$ is the set of all $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \le \frac{1}{T}C(\lambda_1), \qquad R_2 \le \frac{1}{T}C(\lambda_2),$$

$$R_1 + R_2 \le \frac{1}{T}C(\lambda_1) - \frac{1}{T}\mathbf{E}\text{Exp}\left(\frac{1+R_2}{\lambda_1(1+R_2(1-\rho^2))}\right)\text{Ei}\left(-\frac{1+R_2}{\lambda_1(1+R_2(1-\rho^2))}\right),$$

where $\lambda_k = W_kT\sigma_k^2/2N_0$ is the average received signal-to-noise ratio of user k, and ρ is the crosscorrelation of the signature waveforms. The expectation is taken over $R_2 = |a_{2i}|^2$ which is exponentially distributed with mean σ_2^2. □

In the special case when the signature waveforms are identical, the capcaity region reduces to the following expression.

Corollary 1

If the signature waveforms are identical (i.e., $\rho = 1$), the capacity region becomes the set of all $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 \le \frac{1}{T}C(\lambda_1), \qquad R_2 \le \frac{1}{T}C(\lambda_2),$$

$$R_1 + R_2 \le \begin{cases} \frac{1}{T}\frac{\lambda_1C(\lambda_1) - \lambda_2C(\lambda_2)}{\lambda_1 - \lambda_2} & \text{if } \lambda_1 \ne \lambda_2, \\ \frac{1}{T}\left[1 + (1 - \frac{1}{\lambda})C(\lambda)\right] & \text{if } \lambda_1 = \lambda_2 = \lambda. \end{cases}$$

□

When the signature waveforms of the users span over several symbol intervals (i.e., $s(t) \in [0, LT]$), the single-user channel becomes $Y_i = \sum_{j=0}^{L-1} h_j a_{i-j}X_{i-j} + N_i$, where $h_j = R_s(jT)$ and $R_s(t)$ is the autocorrelation function of $s(t)$. Even in the single-user case, the capacity of this channel is known only in a limiting expression. However, if we assume that the channel is very slowly time-varying so that the fading process is a random constant over the duration of any codeword, the capacity of the single-user channel and the optimal input PSD can be obtained in a parametric expression similar to that in the classical water-filling argument.

Theorem 2

The capacity of the frequency non-selective very slowly time-varying Rayleigh fading single-user channel is equal to

$$C = \frac{1}{T}\int_0^1 C(P(f)T(f))df$$

where $P(f)$ is the solution of

$$P(f)T(f) = F^{-1}([cT(f) - 1]^+),$$
$$\lambda = \int_0^1 P(f)df.$$

In the above equations, $T(f) = \sum_{k=-\infty}^{\infty} |S((f-k)/T)|^2$,

$$F(x) = \frac{x^2}{x - C(x)} - 1 = \frac{x^2}{x - \text{Exp}\left(\frac{1}{x}\right)\text{Ei}\left(-\frac{1}{x}\right)} - 1,$$

λ is the average received signal-to-noise ratio, and $S(f)$ is the spectrum of the signature waveform. □

This result can be viewed as a generalization of the water-filling result to the fading channel since if $F(x) = x$ and $C(x) = \log[1+x]$, the above characterization reduces to the classcial water-filling argument.

References

[1] M. G. Robert, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.

[2] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. London: Academic Press, Inc., 1980.

THE CHANNEL CAPACITY IN THE PRESENCE OF IMPULSE NOISE

Kenneth J. Kerpez

Bellcore, Morristown, NJ 07960-1910

Impulse noise is bursty, high amplitude, low probability noise. Impulse noise often occurs from man-made disturbances. Impulse noise is not well understood because it is not Gaussian. However, impulse noise is a significant impairment for digital transmission. This paper analyzes impulse noise by information theoretic calculations. In particular, the channel capacity in the presence of impulse noise is bounded and computed.

The signal and noise are assumed to be band-limited and sampled at the Nyquist rate. Impulses are assumed to have independent Poisson arrivals. Each impulse noise sample, n, is modeled by the probability density function

$$f_S(n) = (1-\lambda)\delta(n) + \lambda f_V(n)$$

where λ is the arrival rate of impulses, $\delta(n)$ is the Dirac delta function, and $f_V(n)$ is the "impulse height" density. Assume that $f_V(n)$ is a continuous function in a neighborhood about $n = 0$. Using the theory of generalized functions it was shown that if the only source of noise is impulse noise, then the channel capacity is unbounded. Thus, later results assume that there is always some additive white noise.

Denote the probability density of the additive white noise as $f_W(n)$. Then the density of the sum of impulse noise and additive white noise is

$$f_N(n) = (1-\lambda)f_W(n) + \lambda f_{V+W}(n)$$

with $f_{V+W}(n) = f_V(n) * f_W(n)$. Upper and lower bounds were derived for the differential entropy, $h(N)$, of the impulse noise plus additive white noise. First, $h(N)$ was upper bounded by the differential entropy of Gaussian noise with the same variance, $h(N) \le (1/2)\ln[2\pi e\sigma_n^2] = UB$, where σ_n^2 is the variance of the sum of impulse noise and white noise. A second upper bound to $h(N)$ was derived by using the non-negativity of the Kullback-Leibler distance, $D(f_N \,||\, f_W)$. A third upper bound was derived by applying the concavity of the logarithm to the definition of $h(N)$. It was proven that the first upper bound is always the tightest.

Assuming that the additive white noise is Gaussian, three lower bounds to $h(N)$ were found. The first lower bound was found with the entropy power inequality, it is: $(1/2)\ln[2\pi e\sigma_w^2] \le h(N)$, where σ_w^2 is the variance of the additive white noise. The second and third lower bounds were found by bounding the integral expression for the differential entropy, they are:

$$1 - \frac{(1-\lambda)^2}{\sqrt{4\pi}\,\sigma_w} - \frac{\lambda^2}{\sqrt{4\pi(\sigma_w^2+\sigma_v^2)}} - \frac{2\lambda(1-\lambda)}{\sqrt{2\pi(2\sigma_w^2+\sigma_v^2)}}\,\beta \le h(N)$$

and

$$1 + \frac{1}{2}\ln[2\pi] - \frac{(1-\lambda)^2}{\sqrt{2}\,\sigma_w} - \frac{\lambda^2}{\sqrt{2(\sigma_w^2+\sigma_v^2)}} - \frac{2\lambda(1-\lambda)}{\sqrt{2\sigma_w^2+\sigma_v^2}}\,\beta \le h(N).$$

Here,

$$\beta = e^{\frac{\mu^2}{2(\sigma_w^2+\sigma_v^2)}\left[\frac{\sigma_v^2}{2\sigma_w^2+\sigma_v^2}-1\right]},$$

and μ and σ_v^2 are the mean and variance of the impulse height density, $f_V(n)$. It was found that any of the three lower bounds may be the tightest, depending on the parameters. The tightest lower bound to the differential entropy, $h(N)$, is denoted as LB.

The bounds on the differential entropy were used to bound the channel capacity. Upper bounds to the capacity were found by using the data processing inequality for Markov chains and the differential entropy of Gaussian noise. The lower bound to capacity was found by using the entropy power inequality. It was shown that

$$\frac{1}{2}\ln[2\pi eP + e^{2LB}] - UB \le Capacity \le \frac{1}{2}\ln[2\pi e(P+\sigma_n^2)] - LB$$

where P is the received signal power. The capacity was also upper bounded by the capacity of a channel with only additive white Gaussian noise and no impulse noise.

The hyperbolic probability density accurately models the heights, $f_V(n)$, of impulses observed on local copper telephone loops. The hyperbolic density is: $f_V(n) = C/|n^3|$ if $VL \le |n| \le VH$, and $f_V(n) = 0$ otherwise; where C is a constant. The capacity bounds were evaluated and compared for both Gaussian and hyperbolic impulse heights. It was found that there is slightly less channel capacity with Gaussian impulse heights than with hyperbolic impulse heights.

The capacity bounds were computed and plotted for a variety of signal and noise powers. It was found that the bounds on channel capacity are tight if the power of the additive white Gaussian noise is large, or if the impulse arrival rate, λ, is small. Most significantly, it was found that the capacity with additive Gaussian white noise and impulse noise is practically the same as it is with just additive white Gaussian noise and no impulse noise, as long as the arrival rate of impulses is less than about 1 impulse every 100 samples. For parameters typical of high speed digital transmission on telephone loops, the difference in capacity with and without impulse noise is less than 4 thousandths of a percent. In general, for most transmission parameters, impulse noise has almost no effect on channel capacity, even if the additive white Gaussian noise power is small.

Formulas were derived for the bit error rate of an uncoded transmission system in the presence of impulse noise and additive white Gaussian noise. The bit error rate was plotted and compared to the channel capacity. It was shown that impulse noise can greatly increase the bit error rate, but have almost no effect on capacity. For parameters typical of telephone loops, the gap between the capacity and the uncoded bit error rate in the presence of impulse noise is a hefty 40.5 dB, at a 10^{-7} bit error rate. This gap can be decreased to 9 dB or less by using interleaved linear block codes to mitigate impulse noise errors, provided that the resulting delay can be tolerated.

WORST-CASE POWER-CONSTRAINED NOISE FOR BINARY-INPUT CHANNELS

Shlomo Shamai (Shitz)
Dept. Electrical Eng.
Technion
Haifa 32000, ISRAEL

Sergio Verdú
Dept. Electrical Eng.
Princeton University
Princeton, NJ 08544, USA

The error probability and the capacity of binary-input additive-noise channels are well-known if the noise is Gaussian. A basic problem in communication theory is to find the worst-case performance achievable by any noise distribution as a function of the signal-to-noise ratio. This paper gives a complete solution to this problem for the two major performance measures: error probability and capacity. Those results are obtained as an application of a general framework developed in [1] which applies to many other performance functionals of information-theoretic interest besides error probability and capacity, such as divergence, cutoff rate, random-coding error exponent, and Chernoff entropy. Those general results show that the worst-case performance functional is given by the convex hull of the functional obtained by minimizing only over power-constrained noise distributions which place all their mass on a lattice whose span is equal to the distance between the two inputs. This implies that the least-favorable distribution is, in general, the mixture of two lattice probability mass functions. This conclusion can actually be generalized to m-ary input constellations on finite-dimensional spaces, as long as the input constellation puts its mass on a lattice. The proof of the results presented in this paper can be found in its journal version [1].

Consider the binary equiprobable hypothesis testing problem:

$$H_1: \quad Y = +1 + N$$
$$H_0: \quad Y = -1 + N$$

where N is a real-valued random variable constrained to satisfy an average-power limitation $E[N^2] \le \sigma^2$. For $k = 1, 2, \cdots$ let $\sigma_k^2 \triangleq (k^2 - 1)/3$. The worst-case probability of error is

$$P_e(\sigma^2) = \frac{1}{2} - \frac{1}{2\sqrt{1 + 3\sigma_k^2}} + \frac{3}{2}\frac{\sigma^2 - \sigma_k^2}{k(k+1)(2k+1)}$$

for $\sigma_k^2 \le \sigma^2 \le \sigma_{k+1}^2$.

A single span-2 lattice achieves the maximum probability of error only when the allowed noise power is equal to σ_k^2, $k = 1, 2, \cdots$ Those worst-case distributions are symmetric and distribute their mass evenly on k atoms. (Those atoms are located at $0, \pm 2, \pm 4, \cdots$ if k is odd and at $\pm 1, \pm 3, \cdots$ if k is even.) When the allowed noise power lies strictly between $\sigma_k^2 < \sigma^2 < \sigma_{k+1}^2$, then a single span-2 lattice is no longer least favorable. Instead, the worst-case distribution is the unique span-1 lattice which is a mixture of the span-2 lattices that are least-favorable for σ_k^2 and σ_{k+1}^2 with respective weights $(\sigma^2 - \sigma_{k+1}^2)/(\sigma_k^2 - \sigma_{k+1}^2)$ and $(\sigma_k^2 - \sigma^2)/(\sigma_k^2 - \sigma_{k+1}^2)$. In particular, if $SNR > 0$ *dB* ($\sigma^2 < 1$), then the worst-case noise is symmetric with nonzero atoms at -1, 0, $+1$, i.e., the channel becomes a symmetric erasure channel. Thus, the noise distribution that maximizes error probability puts all its mass on the integers $\{-M, \cdots M\}$, where M depends on the signal-to-noise ratio and the weight assigned to each of those integers depends (in addition to the signal-to-noise ratio) only on whether the integer is even or odd. Note that for low signal-to-noise ratios, the worst-case noise cdf does not become asymptotically Gaussian, as might have been surmised from capacity considerations.

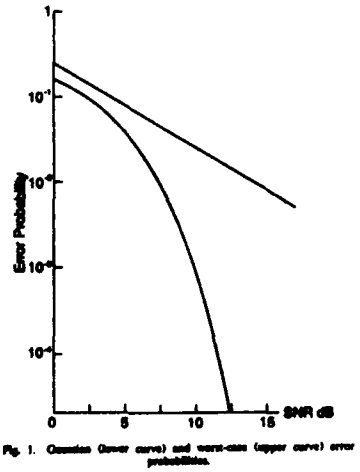


Fig. 1. Gaussian (lower curve) and worst-case (upper curve) error probabilities.

The worst-case capacity problem is

$$C(\sigma^2) = \min_{\substack{N \\ E[N^2] \le \sigma^2}} \max_{X} I(X; X+N) = \max_{X} \min_{\substack{N \\ E[N^2] \le \sigma^2}} I(X; X+N)$$

where the maximum ranges over all distributions on $\{-1,1\}$ and the second equality follows from the concavity-convexity of mutual information in the respective arguments. It is verified in [1] that the least-favorable noise distribution puts its mass in the lattice $\{ \cdots -4, -2, 0, +2, +4 \cdots \}$ with a probability mass function that satisfies

$$\log(1 + \frac{p_{-1}}{p_0}) + \log(1 + \frac{p_1}{p_0}) - \log(1 + \frac{p_{k+1}}{p_k}) - \log(1 + \frac{p_{k-1}}{p_k}) + \lambda k^2 = 0$$

where p_k is the mass at $2k$. For low SNRs the least-favorable cdf approaches a Gaussian shape, whereas in the region $SNR > 0$ *dB* the least-favorable distribution is indistinguishable from a three-mass distribution with weights $(\sigma^2/8, 1 - \sigma^2/4, \sigma^2/8)$ at $(-2, 0, 2)$. The maximum difference between Gaussian capacity and worst-case capacity is 0.118 bit and occurs at 7.2 dB, whereas the maximum relative decrease is 12.5% occuring at 6.7 dB.

References

1. S. Shamai (Shitz) and S. Verdu, "Worst-Case Power Constrained Noise for Binary-Input Channels," *IEEE Transactions on Information Theory*, pp. 1494-1511, Sep. 1992.

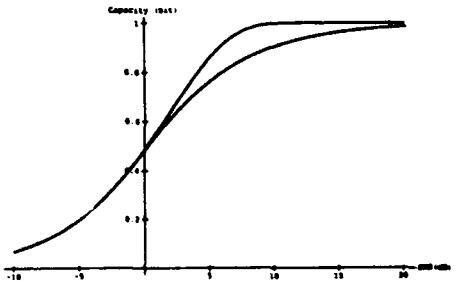


Fig. 2. Gaussian (upper curve) and worst-case (lower curve) channel capacities.

This work was supported in part by the Office of Naval Research under Grant N00014-90-J-1734.

ERROR EXPONENTS FOR THE IDEAL POISSON CHANNEL WITH NOISELESS FEEDBACK

Amos Lapidoth
Technion—Israel Institute of Technology and
Stanford University, Information Systems Laboratory, Stanford, CA 94305-4055

Abstract

The ideal Poisson channel with noiseless feedback models a direct detection photon channel, free of dark current ($\lambda_0 = 0$), in which a causal feedback link informs the transmitter at every time t of the channel output at all times prior to t. The paper discusses the coding for the channel, under peak power and average power constraints on the input. A coding scheme for the channel is presented, and its asymptotic error exponent is shown to coincide, at all rates below capacity, with the Sphere Packing error exponent, which, for the case at hand, is known to be unachievable without feedback for rates below the critical rate. An upper bound on the error exponent achievable with feedback is also derived. It is shown that under a capacity reducing average power constraint, the upper bound coincides with the error exponent achieved by the proposed coding scheme; consequently, in such a case the coding scheme is asymptotically optimal. Thus, the ideal Poisson channel, limited by a capacity-reducing average power constraint, provides a non-trivial example of a channel for which the Reliability Function is known exactly both with and without feedback. While our main concern is fixed transmission time coding schemes, the subject of random transmission times is also briefly discussed; it is shown that a slight modification of the coding scheme to one of random transmission time can achieve zero-error probability for any rate lower than the ordinary average-error channel capacity.

A DIRECT GEOMETRICAL METHOD FOR BOUNDING THE ERROR EXPONENT FOR SPECIFIC FAMILIES OF CHANNEL CODES - PART II: THE CONFINING REGION LOWER BOUND FOR BLOCK CODES

Dejan Lazić, Vojin Šenk

Faculty of Technical Sciences
Computer Science, Control and Measurements Institute
Trg D. Obradovića 6, 21000 Novi Sad, Yugoslavia

The introduction of a confining region that divides the channel output space in two disjoint parts attains the same effect as the Gallager's exponent, but gives much more insight into the behaviour of channel codes. Moreover, the bounds obtained are always tight at low code rates (for optimal codes they are always tight).

Introduction

The channel error exponent (reliability function) is defined as

$$E(R) = \lim_{N\to\infty} \left\{ -\frac{1}{N} \mathrm{ld}\left[P_{eopt}(R,N) \right] \right\}, \quad \mathrm{ld}(x)=\log_2(x), \qquad (1)$$

where R is the code rate, $P_{eopt}(R,N)$ is the smallest possible probability of block decoding error for codes of code rate R and length N used on the given channel.

In Part I of this paper, the random coding argument usually used in lower-bounding the channel error exponent was discarded in favour of the one that uses the known expected Bhattacharyya distance distribution of a code family. If the code family distance distribution is known, the error exponent obtained pertains to this specific code family used on the given transmission channel, and not to the channel itself. The code family that attains the channel error exponent is the optimal one, and its Bhattacharyya distance distribution the optimal distance distribution.

The general expression for E(R)

The distance distribution method in its final form gives the lower bound on the code family error exponent in the form

$$E(R)_{\mathcal{B}} = \min_{d_{B1}\le d_B\le d_{BL}} \left\{ E_{\mathcal{B}}(d_B,R) + \underset{\sim}{E}_e(d_B,R,\mathcal{B}) \right\} - R, \qquad (2)$$

where $\mathcal{B}$ denotes the code family, d_B is the normalized (by N) Bhattacharyya distance, and d_{B1} and d_{BL} are the minimum and maximum normalized Bhattacharyya distances in the code. $E_{\mathcal{B}}(d_B,R)$ is the expected normalized Bhattacharyya distance distribution exponent of $\mathcal{B}$, and $\underset{\sim}{E}(d_B,R,\mathcal{B})$ is the lower bound on the error effect exponent (see [1] for precise definition of these exponents).

The code family Gilbert-Varshamov distance

If a codeword is expurgated from each pair of codewords in the code that is less than the code family Gilbert-Varshamov distance, defined as

$$d_{BGV}(\mathcal{B},R) = \min \{ d_B : E_{\mathcal{B}}(d_B,R) \le R \}, \qquad (3)$$

the cutoff rate lower bound [1] on the expurgated family $\mathcal{B}$ is reduced to $d_{BGV}(\mathcal{B},R)$ at all code rates less than

$$R_{crit1\mathcal{B}}^{*} = R_{o\mathcal{B}} - d^{o}_{B\mathcal{B}eff}. \qquad (4)$$

Here, $R_{o\mathcal{B}}$ is the code family cutoff rate, and $d^{o}_{B\mathcal{B}eff}$ is the normalized Bhattacharyya distance of those codewords that are the only ones that influence the code family cutoff rate bound. At low code rates (not greater than $R_{crit1\mathcal{B}}^{*}$) the cutoff rate bound is tight.

The confining region

The channel output space Y^N may be partitioned in two using the confining region defined for the m'th codeword as

$$CS_m^{int}(r_B) = \{y^{int}\} = \left\{ y\in Y^N : \frac{P[y|x_m]}{\max\limits_{x\in\Omega_N^{ext}(x_m,r_B)} \{P[y|x]\}} \ge 1 \right\}, \quad m = 1,\ldots,M, \qquad (5)$$

where $\Omega_N^{ext}(x_m,r_B) = \{x^{ext}\} = \{x\in\mathcal{X} : d_B(x_m,x) > r_B\}$ is the exterior of the Bhattacharyya confining sphere in the available encoding space $\mathcal{X}$, centered at the actually transmitted codeword x_m and whose radius is r_B (expressed in the normalized Bhattacharyya distance, that is defined in $\mathcal{X}$, but not in Y^N). Upper-bounding the probability of error inside $CS_m^{int}(r_B)$ by the usual union bound, and outside it by the mere probability that $y\in CS_m^{int}(r_B)$, one obtains a significant improvement. For the code families uniformly distributed over $\mathcal{X}$, this procedure yields the Gallager's lower bound on the channel error exponent, implying that these codes are optimal. Moreover, the error exponent of these families is known at all code rates, since it is known below $R_{crit1\mathcal{B}}^{*}$ and also coincides with the space-partitioning upper bound on the channel error exponent at high code rates.

References

[1] D. E. Lazić, V. Šenk, "A Direct Geometrical Method for Bounding the Error Exponent for Any Specific Family of Channel Codes - Part I: Cutoff Rate Lower Bound for Block Codes", *IEEE Trans. Info. Th.*, Vol. 38, pp. 1548-1559, September 1992.

UNIVERSAL DECODING FOR MEMORYLESS GAUSSIAN CHANNELS WITH A DETERMINISTIC INTERFERENCE

Neri Merhav

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL

Abstract

In [1] universal decoding schemes for finite-alphabet, finite-state channels were proposed and shown to be optimal in the sense of attaining the highest possible random coding error exponent, when the channel input vectors are chosen randomly under a uniform probability distribution. In other words, the average error probability over the ensemble of randomly selected codewords, decays at the fastest exponential rate. In the special case of DMC's, the proposed universal decoder selects a codeword that minimizes the empirical conditional entropy of the channel input given the channel output.

We derive an analogous result for memoryless Gaussian channels with an unknown deterministic interference from a fairly wide class. The empirical conditional entropy of the input given the output is induced by an auxiliary backward channel whose parameters are estimated from the given output vector and each one of the codewords. We also allow a more general class of input distributions by slightly modifying the decoding rule. Similarly to [1], it is shown that the proposed universal decoder attains the same error exponent as that of the ML decoder which is fully informed of the channel and the interfering signal.

The proposed decoder is different from an heuristic approach [2], where the channel and message are jointly estimated by the ML method. While the former decoder is based on the backward channel as mentioned earlier, the latter corresponds to the *forward* channel. For the simple special case where there is no interference and the only uncertainty is in the channel fading parameter, it is demonstrated that the error exponent of the proposed rule might be strictly better than that of the joint ML channel-and-message estimation approach.

Summary

Consider a discrete-time, Gaussian memoryless channel characterized by $y_t = ax_t + z_t + w_t$, where x_t is the desired channel input, a is an unknown fading parameter, w_t is zero mean Gaussian white noise with an unknown variance σ^2, z_t is an unknown deterministic interference, and y_t is the channel output. We assume that z_t can be represented by a series of given orthonormal bounded functions, i.e., $z_t = \sum_{i=1}^{\infty} b_i \phi_{i,t}$, where $\sum_{i=1}^{\infty} |b_i| < \infty$ and $|\phi_{i,t}| \le L$ for all i and t, $0 < L < \infty$.

Consider next, a codebook $C = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^M\}$ of $M = 2^{nR}$ equiprobable messages $\mathbf{x}^i = (x_1^i, x_2^i, \ldots, x_t^i, \ldots, x_n^i) \in \mathbb{R}^n$, $i = 1, 2, \ldots, M$, where R is the coding rate in bits per channel use. Clearly, if the parameter a and the interference signal z_t are known, the best is the ML decoder, which in the Gaussian case considered here, selects the message $\mathbf{x}^i$ that minimizes $\sum_{t=1}^{n}(y_t - z_t - ax_t^i)^2$. Similarly to [1], the probability of error associated with the ML decoder will be denoted by $P_{e,o}(C,R,n)$.

Since the design of a codebook C that minimizes $P_{e,o}(C,R,n)$ under an input power constraint is prohibitedly complex for large n, will shall adopt the random coding approach, where each codeword is randomly chosen with respect to some probability density function (PDF) $q(\mathbf{x})$, independently of all other codewords. It is well known [3] that $\bar{P}_{e,o}(q,R,n) \triangleq E\{P_{e,o}(C,R,n)\}$, where the expectation is taken over ensemble of randomly selected codebooks under q, decays exponentially for every $R < R(q)$, where $R(q)$ is a rate depending on q and always less than the channel capacity. The exponential rate of the error probability $E(q,R) \triangleq -\lim_{n\to\infty} n^{-1} \log \bar{P}_{e,o}(q,R,n)$ is called the *random coding error exponent*.

If the fading parameter a and the interfering signal $\{z_t\}$ are unknown, then the ML decoder is obviously inapplicable. We next demonstrate a decoding procedure which is universal in the sense of being independent of a and $\{z_t\}$, and at the same time attaining $E(q,R)$. In other words, let $P_{e,u}(C,R,n)$ denote the error probability associated with the universal rule for a given codebook C, and let $\bar{P}_{e,u}(q,R,n) = E\{P_{e,u}(C,R,n)\}$. Then, $\bar{P}_{e,u}(q,R,n)$ decays exponentially at the same rate $E(q,R)$ as that associated with the ML decoder. This is analogous to an earlier result by Ziv [1] for finite-alphabet, finite-state channels.

We now turn to present the proposed decoding rule. To this end, define an *auxiliary backward channel* of order k by the conditional PDF

$$W(\mathbf{x}|\mathbf{y},\theta,k) = (2\pi\sigma_0^2)^{-n/2} \prod_{t=1}^{n} \exp\left\{-\frac{1}{2\sigma_0^2}(x_t - \alpha y_t - \sum_{i=1}^{k} \beta_i \phi_{i,t})^2\right\}, \quad (1)$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ and $\theta \triangleq (\sigma_0^2, \alpha, \beta_1, \beta_2, \ldots, \beta_k)$ is the parameter vector of the kth order backward channel. Let $\{k_n\}_{n\ge 1}$ be any monotonically nondecreasing integer-valued sequence satisfying $k_n \to \infty$ and $k_n/n^{1/3} \to 0$ as $n \to \infty$. Our decoding rule will select a message $\mathbf{x}^i$ that maximizes the function

$$u(\mathbf{x}^i,\mathbf{y}) \triangleq \frac{\max_{\theta} W(\mathbf{x}^i|\mathbf{y},\theta,k_n)}{q(\mathbf{x}^i)} \quad (2)$$

among all M codebook messages.

Theorem: Assume that $\{z_t\}$ can be expanded to a series of bounded orthonormal functions with an absolutely summable coefficient sequence $\{b_i\}_{i\ge 1}$. Let $q(\mathbf{x})$ be any Gaussian PDF of the form

$$q(\mathbf{x}) = (2\pi\sigma_x^2)^{-n/2} \prod_{t=1}^{n} \exp\left\{-\frac{1}{2\sigma_x^2}(x_t - \sum_{i=1}^{m} \gamma_i \phi_{i,t})^2\right\}, \quad (3)$$

where σ_x^2, m, and $\gamma_1, \ldots, \gamma_m$ are free parameters to be chosen. Then,

$$\lim_{n\to\infty} \frac{1}{n} \log \bar{P}_{e,u}(q,R,n) = \lim_{n\to\infty} \frac{1}{n} \log \bar{P}_{e,o}(q,R,n) = -E(q,R), \quad (4)$$

where $\bar{P}_{e,u}(q,R,n)$ is the average error probability associated with (2).

The proof appears in [4].

The intuitive interpretation of (2) is that $\log u(\mathbf{x},\mathbf{y})$ can be thought of as an empirical version of the mutual information between $\mathbf{x}$ and $\mathbf{y}$. Thus, we select the input $\mathbf{x}^i$ that seems empirically "most dependent" upon the given output vector $\mathbf{y}$. This corresponds to the maximum mutual information (MMI) decoding principle. It should be pointed out that if $\{z_t\}$ is known to be composed from $k < \infty$ basis functions $\{\phi_{i,t}\}$, then the theorem applies with $k_n = k$ in eq. (2) and $m \le k$ in eq. (3).

Ideally, one wishes to choose $q(\cdot)$ so as to maximize $E(q,R)$. However, since the maximizing PDF $q(\cdot)$ depends on the unknown channel, there is no way by which the transmitter can optimally select $q(\cdot)$ unless there is a feedback channel from the receiver to the transmitter. The choice of an input PDF of the form of eq. (3) can be also motivated by the fact that the capacity of the Gaussian channel is attained by a Gaussian PDF.

It turns out that the extension of the above theorem to nonmemoryless channels is not trivial. Consider, for example, a Gaussian channel with a linear intersymbol interference (ISI), characterized by $y_t = \sum_{i=0}^{k} h_i x_{t-i} + w_t$, where $\{h_i\}_{i=0}^{k}$ is the channel impulse response and w_t is a Gaussian white noise. The difficulty appears to be in an appropriate definition of the auxiliary backward channel. We conjecture that an appropriate definition of the backward channel in this case will be

$$W(\mathbf{x}|\mathbf{y},\theta,k) = C_n(\theta,k,\mathbf{y}) \prod_{t=1}^{n} \exp\left\{-\frac{1}{2\sigma_0^2}(x_t - \sum_{i=1}^{k} \alpha_i x_{t-i})^2 + \sum_{i=0}^{k} \beta_i x_{t-i} y_t\right\},$$

where $\theta = (\sigma_0^2, \alpha_1, \ldots, \alpha_k, \beta_0, \ldots, \beta_k)$ and $C_n(\theta,k,\mathbf{y})$ is a normalization factor chosen such that the above PDF will integrate to unity.

References

[1] J. Ziv, "Universal Decoding for Finite-State Channels," *IEEE Trans. Inform. Theory*, Vol. IT-31, No. 4, pp. 453-460, July 1985.

[2] N. Seshadri, "Joint Data and Channel Estimation using Blind Trellis Search Techniques," submitted to *IEEE Trans. Commun.*

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York, Wiley 1968.

[4] N. Merhav, "Universal Decoding for Memoryless Gaussian Channels with a Deterministic Interference," submitted for publication.

ON INFORMATION RATES FOR MISMATCHED DECODERS

Neri Merhav, Gideon Kaplan, Amos Lapidoth and Shlomo Shamai (Shitz)
Department of Electrical Engineering
Technion—Israel Institute of Technology, Haifa 32000, Israel

Consider the reliable transmission of information over a discrete-time memoryless channel with *mismatched* decoding, i.e., where the decoding metric is not necessarily matched to the channel's characteristics. This is a realistic model for time-varying channels or when implementation constraints dictate a given decoder which employs a specific fixed metric.

Hui [1] has derived a lower bound on the capacity of a discrete memoryless channel (DMC) with mismatched decoding, hereafter referred to as Hui's capacity and denoted C_H. Our first result in this work is an extension of this lower bound to an exponential family of channels. This wider class of channels includes, as special cases, DMC's, finite-state channels, Poisson channels and Gaussian channels. Some of the results extend to exponential channels with memory (e.g., finite-state channels), but in this case a single-letter characterization of the achievable rates is not available.

Motivated by the matched decoding case, we prove that in the random coding regime, Hui's capacity is the highest achievable rate under mismatched decoding. This observation, as well as a sphere packing argument for bounding the maximum possible number of disjoint mismatched decoding spheres, support Hui's conjecture (recently proved by Balakirsky [2] for binary-input channels) that C_H is the ultimate reliably transmitted rate. New bounds and interesting properties of C_H are presented [3], and relations among C_H, the generalized average mutual information (defined in terms of Gallager's bound in parallel to the matched case) and the generalized cut-off rate are established.

Some indicative examples of practical interest for continuous and discrete-alphabet memoryless channels with various mismatched metrics are worked out. In particular, a two-dimensional AWGN channel (with Gaussian inputs) subjected to a phase offset of θ is considered. It is found that the deleterious effect of the phase offset θ on C_H manifests itself in attenuating the signal power by a factor of $\cos^2\theta$ and in adding an equivalent noise term with power of $\sin^2\theta$ times the signal power. This expression mimics the behavior of the uncoded complex channel with a phase offset.

We proceed to examine specific examples of encoding/decoding mechanisms motivated by the nature of the mismatch. It is demonstrated that the achievable reliable transmitted rate under mismatched decoding may depend on the performance criterion (bit error vs. message error probability) and on the coding *strategy* (randomized vs. deterministic), in contrast to the well-known behavior of the optimal matched-decoding scenario. As an example, consider a BSC with crossover probability of $p < 0.5$, where the decoder uses the mismatch metric adapted to a BSC with $p' > 0.5$ instead of p. In this case $C_H = 0$ [1], however, by using a variant of differential encoding one can achieve a positive rate with respect to the bit error probability (while the message error probability goes to unity). Moreover, a randomized strategy (e.g. assigning to any possible message, with equal probability, a properly selected binary codeword or its complement) leads to a positive achievable rate with respect to the message error probability. Thus, in several specific cases, with different error criteria and/or randomized coding strategies, reliable rates exceeding C_H are achievable.

References

[1] J.Y.N. Hui, Fundamental Issues of Multiple Accessing. Ph.D. dissertation, Chapter IV, MIT, 1983.

[2] V.B. Balakirsky, "Coding theorem for discrete memoryless channels with given decision rules," Lecture Notes in Computer Science 573, Proceedings of First French-Soviet Workshop on Algebraic Coding, July 1991, pp. 142–150.

[3] N. Merhav, G. Kaplan, A. Lapidoth and S. Shamai (Shitz), "On information rates for mismatched decoders," submitted to IEEE Trans. on Inform. Theory, (EE Report No. 863, Technion, Elect. Eng. Dept., November 1992).

A Markovian Evaluation of the Frame Error Probability for the M Algorithm[1]

Jean Belzile and David Haccoun
Department of Electrical and Computer Engineering
Ecole Polytechnique de Montréal
P.O. Box 6079, station "A"
Montréal, Qc, Canada
H3C 3A7

Abstract

A new Markov chain approach to the evaluation of the frame error probability for the M-Algorithm is presented. Using this model results for values of M=1 to 64 and frame length of L=64 to 512 bits have been evaluated for a convolutional code of memory length v=19 and rate R=1/2. Simulation results are compared to the Markovian model showing that the technique is attractive for the performance evaluation of suboptimal decoding algorithms for convolutional codes.

Summary

Suboptimal decoding algorithms in general and the M-Algorithm in particular have received a significant amount of interest lately [1-5]. These algorithms are used to search large trellises where an exhaustive exploration is impractical. Their suboptimal search is heuristically guided to minimize the number of paths to be explored in order to achieve a reasonable bit error performance. However theoretical analysis of these heuristics is complex and few theoritical mandhods are available for precise evaluation on the error performance of these algorithms.

In order to establish an upper bound on the error performance of the M-Algorithm, the minimum number of path extensions required to include the correct path at each tree depth must be determined. This problem has eluded analysis.

In this paper we present a new approach to the evaluation of the frame error performance of the M-Algorithm over a binary symmetric channel and additive white gaussian noise. It is based on a Markovian description of the decoding dynamics of the M-Algorithm and uses the column weight distribution of the code [6]. The column weight distribution of a particular convolutional code represents the number of paths with a certain Hamming weight at each particular depth in the decoding tree.

The Markov chain consists of an "Initial" state, a "Lost" state and a varying number of intermediate states. The "Initial" state represents the decoder behaviour when the channel is error-free, typically at the beginning of the frame or when the channel has been error-free for a sufficiently long period of time. The "Lost" state is an absorbent state which represents an error propagation event due to the loss of the correct path and its ensuing lack of resynchronization. An intermediate state represents the accumulated channel transitions at a given incorrect subset depth since departure from the "Initial" state. The transitions between the states of the Markov chain are a combination of the correct path loss probability and the channel transition probability.

Using this model a transition matrix A may be constructed. For a frame of length L, the frame error probability is then given by the transition probability from the "Initial" state to the "Lost" state in the matrix A^L.

This technique has been applied for a rate $R = \frac{1}{2}$ convolutional code of memory $v = 19$ with frame lengths varying from L=64 to 512 bits and number of paths varying from M=1 to 64. Results show that as a first order approximation, a sliding window decoder [7] is a good approximation if M, the number of paths to be explored is small. However if the number of explored paths increases, then the number of incorrect subsets to consider must also increase, making the sliding window inadequate.

Using this technique a good upper bound on the frame error performance of the M-Algorithm can be calculated for a given code and some value of M. Once the frame error probability is known, the bit error performance for the M-Algorithm can be approximated by using simple arguments [5]. Assuming that $M \ll 2^v$ then the probability of resynchronization tends towards 0. It has been observed through simulations that on the average the decoder will loose the correct path in the middle of the frame and that one half of the decoded bits in the erroneous portion of the frame will be in error, leading to an error event of $\frac{L}{4}$ bits per erroneously decoded frame. A good approximation to the bit error probability is then given by $P_b = \frac{\frac{L}{4}P_f}{L} = \frac{1}{4}P_f$. This approximation is supported by extensive simulation results.

In summary, the new Markovian based frame error probability analysis for the M-Algorithm and a binary symmetric channel will be presented. A comparison between simulation results and numerical results shows that when the number of explored path increases then the number of incorrect subsets in the Markov model should also increase, increasing with it the complexity of the evaluation of A^L. However the extraction of the frame error probability remains trivial making the technique attractive for evaluating the error performance of suboptimal decoding algorithms for convolutional codes.

References

[1] ANDERSON, J. B. and LIN, C. F., "M-Algorithm Decoding of Channel Convolutional Codes," in *Conf. Proc., 20th Annu. Conf. Inform. Sci. Syst.*, (Princandon, NJ), March 1986.

[2] SIMMONS, S. J., "Breadth-First Trellis Decoding with Adaptive Effort," *IEEE Transactions on Communications*, vol. 38, pp. 3-12, Jan. 1990

[3] ANDERSON, J. B., "Limited Search Trellis Decoding of Convolutional Codes," *IEEE Transactions on Information Theory*, vol. 35, Sept. 1989.

[4] CHAN, F., *Algorithme de décodage adaptatif pour codes convolutionnels*, Master's thesis, Ecole Polytechnique, Montréal QC., Dec. 1989.

[5] BELZILE, J. and HACCOUN, D., "Bidirectional Breadth-first Algorithms for the Decoding of Convolutional Codes," *IEEE Transactions on Communications*, Feb. 1993.

[6] BELZILE, J. and HACCOUN, D., "Column weight distributions of Convolutional Codes,"*technical repport*,EPM/RT-91/10 Ecole Polytechnique de Montréal, Aug. 1991.

[7] GALLAGER, R. G., *Information Theory and Reliable Communication*, Wiley, NY, 1968.

1. This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada, the Fonds pour la formation de Chercheurs and l'Aide à la Recherche of Québec and by a grant from the Canadian Institute for Telecommunication Research under the National Centers of Excellence program of the Government of Canada.

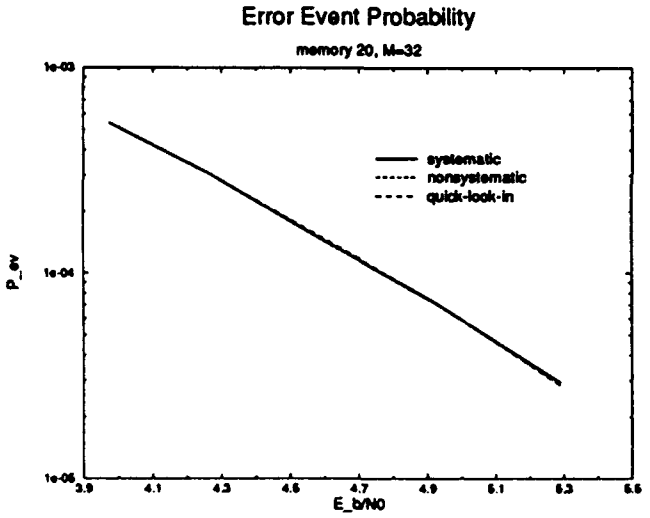
SYSTEMATIC FEED-FORWARD CONVOLUTIONAL ENCODERS ARE AS GOOD AS OTHER ENCODERS WITH AN M-ALGORITHM DECODER

Harro Osthoff*, Rolf Johannesson*, and John Anderson**

* Department of Information Theory
University of Lund
Box 118
S- 221 00 Lund
Sweden

** Electrical, Computer, and
Systems Engineering Department
Rensselaer Polytechnic Institute
Troy, New York 12180-3590
USA

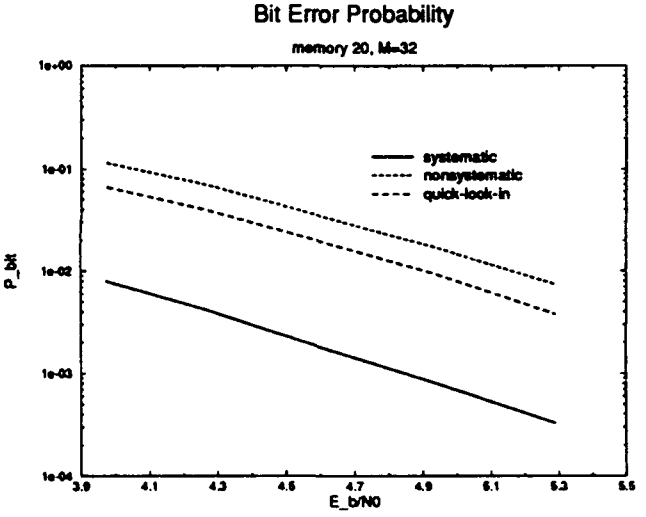
Summary—In this paper we show that systematic convolutional encoders perform as well as nonsystematic ones when they are used together with M-algorithm decoders [1]. We describe the algorithm and give a brief historical review. The following curves show simulation results for the event error probability of the M-algorithm. We compare an optimum distance profile nonsystematic encoder ($d_{free} = 22$) and a quick-look-in encoder ($d_{free} = 18$) with a systematic encoder ($d_{free} = 13$). All encoders have memory $m = 20$ and in the decoder 32 states are extended at every time instant ($M = 32$).



All curves show the same event error probability performance.

Using criteria for the encoder quality like the optimum distance profile and the optimum profile spectrum [2] any encoder is equivalent to a systematic one over the encoder memory. As long as these distance criteria support the decoder performance, the event error probability of the M-algorithm depends only on M. Therefore, in a range of interesting values of M, systematic encoders should behave like nonsystematic encoders in terms of error event probability as our simulations show. The decoder complexity is independent of the memory of the encoder. The free distance does not matter as long it is big enough to correct all paths within the set of extended decoder states. Hence, for M-algorithm decoders, in contrast to the Viterbi decoder, code quality cannot be expressed in terms of the free distance and the distance spectrum. A rapid growth of the column distances is more important than a large free distance.

As a bonus when used together with the M-algorithm, systematic encoders outperform nonsystematic encoders in terms of bit error probability as shown in the next picture (framelength = 1024). The reason is that systematic encoders are superior from a correct path loss point of view.



References

[1] J. B. Anderson, "Limited Search Trellis Decoding of Convolutional Codes", IEEE Trans. Information Theory, IT-35, pp. 944-956, Sep. 1989.

[2] H. Osthoff, R. Johannesson, B. Smeets and H. Vinck, "On the Linear M-Algorithm", Proceedings IEEE Information Theory Symposium, San Diego, Jan. 1990, pp.85.

This work was supported in part by the Swedish Research Council for Engineering Sciences under Grant 91-91.

ANALYSIS OF LIST DECODING FOR CONVOLUTIONAL CODES

Kamil Zigangirov and Harro Osthoff
Department of Information Theory
University of Lund
Box 118
S- 221 00 Lund
Sweden

Summary—We analyse a list decoding algorithm [1] (*M*-algorithm [2]) for binary rate $R = b/c$ convolutional codes. In every decoding step, starting from $\lfloor \frac{\log_2 L}{b} \rfloor + 1$ where $\lfloor . \rfloor$ denotes the integer part, the decoder selects the L most likely code sequences and calculates their successors. This procedure is continued until the decoder reaches the end of the tree or the trellis.

We study the distance properties and the probabilistic performances of the algorithm. We introduce a natural extension to the free distance of the convolutional code, viz., the L-list minimal distance d_L. The L-list decoder corrects all combinations of $\lfloor \frac{d_L - 1}{2} \rfloor$ or less errors. Using computer search we found convolutional encoders having maximal L-list minimal distance.

Analogously to the Costello bound we derive the following lower bound on d_L for rate $R = b/c$ binary convolutional codes: *There exists a time-invariant convolutional code such that* (bound 1)

$$d_L \geq \frac{-\log_2 L}{\log_2(2^{1-R}-1)} + \varphi(R),$$

where $\varphi(R)$ does not dependent on L.

For rate $R = 1/2$ this can be strengthend further (bound 2):

$$d_L \geq \frac{\log_2(L+\frac{1}{2})}{-\log_2(\sqrt{2}-1)} - \frac{2}{\log_2(\sqrt{2}-1)} - 2.$$

The bound for rate $R = 1/2$ can be tightened if we choose and fix the first $i+1$ matrices G_k , $k = 0, 1 \ldots i$ (bound 3):

$$d_L \geq \frac{\log_2(L - 2^i + 1) + 1 - \log_2 T_{[0,i]}(\sqrt{2}-1)}{-\log_2(\sqrt{2}-1)} - 2,$$

where

$$L \geq 2^i$$

and

$$T_{[0,i]}(W) = \sum_{\mathbf{v} \in V_{[0,i]}} W^{w_H(\mathbf{v})},$$

where $w_H(\mathbf{v})$ is the Hamming weight of $\mathbf{v}$, is the *weight enumerator* of the i-th truncation of the code.

Based on a Hamming type upperbound for L-list minimal distance of block codes we prove the following upperbound for d_L:

$$d_L < \frac{-2\log_2 L}{\log_2(2^{1-R}-1)} + o(\log_2 L),$$

where

$$\frac{o(\log_2 L)}{\log_2 L} \longrightarrow 0, \text{ for } L \longrightarrow \infty$$

Finally, we derive a random coding type upper bound on the decoding error probability $P(\mathcal{E}_t)$ of the L-list decoding algorithm on the t-th step and an expurgation type and sphere packing type lower bounds. The random coding upper bound is:

$$E[P(\mathcal{E}_t)] \leq \begin{cases} L^{-1}O_1(1), & R < R_{comp}, \\ L^{-1}O_2(1)\log_2 L, & R = R_{comp}, \\ L^{-\varrho}O_3(1), & R > R_{comp}, \end{cases}$$

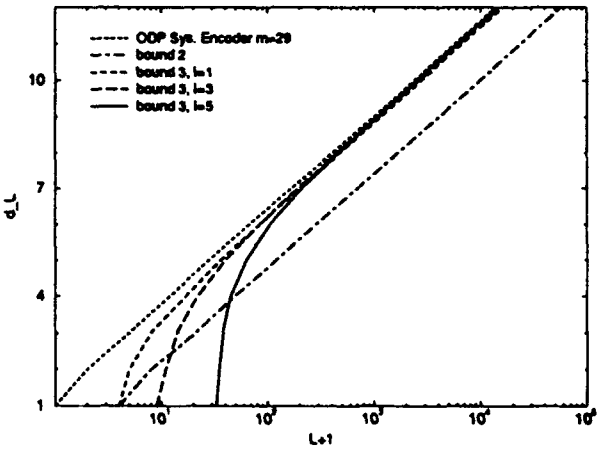
where ϱ is the solution of the equation $\varrho R = G(\varrho)$, $G(\varrho)$ is the Gallager function for the BSC, and $O_i(1)$, $i = 1, 2, 3$, are values depending on R and ε but not on L. Using the expurgation bound we get:

$$P(\mathcal{E}_t) \leq L^{-\frac{\log_2 \sqrt{4\varepsilon(1-\varepsilon)}}{\log_2(2^{1-R}-1)}} O(1), \; R < R_{comp},$$

where $O(1)$ is a value depending on R and ε, but not on L.

The derivation of the lower bound for the L-list decoding error probability is based on the corresponding lower bound for block codes. For a given L and R there exists an integer t such that the L-list decoding error probability on the t-th step satisfies the inequality

$$P(\mathcal{E}_t) > L^{-\varrho_2[-O(\sqrt{\log_2 L})]}.$$



Lower bounds for d_L compared with d_L for the systematic ODP code with $m = 29$.

References

[1] K. Sh. Zigangirov, V. D. Kolesnik "List Decoding of Trellis Codes", Problems of Control and Information Theory, No. 6, 1980.

[2] J. B. Anderson, "Limited Search Trellis Decoding of Convolutional Codes", IEEE Trans. Information Theory, IT-35, pp. 944-956, Sep. 1989.

This work was supported in part by the Swedish Research Council for Engineering Sciences under Grant 91-91 and in part by the Royal Swedish Academy of Sciences in liasion with the Russian Academy of Sciences.

TABLE-DRIVEN DECODING OF BINARY ONE-HALF RATE NONSYSTEMATIC CONVOLUTIONAL CODES

Ajay Dholakia
Dept. of Elec. & Comp. Eng.

Mladen A. Vouk
Donald L. Bitzer
Department of Computer Science

North Carolina State University, Raleigh, NC 27695-7911

SUMMARY

Table look-up based decoding schemes for convolutionally encoded data have been designed for both block coding [5], and convolutional coding [6, 2]. However, in the latter case most of the work was done for systematic codes. Nonsystematic codes offer better error correcting capability than systematic codes if more than one constraint length of received blocks are considered [7]. Our approach to table look-up based decoding of nonsystematic convolutional codes was introduced in [1, 3].

A 1/2-rate convolutional encoder is characterised by two generator sequences $\mathbf{g}^{(j)} = (g_0^{(j)}, g_1^{(j)}, ..., g_\nu^{(j)})$, $j = 1, 2$, where ν is the *constraint length* of the code, i.e., the number of memory elements in a minimal realisation of the convolutional code [4]. The *output constraint length* is defined as $n_A = 2(\nu + 1)$ and is equal to the number of encoded bits affected by a single input information bit [7]. An input information sequence $\mathbf{u}$ is encoded into two encoded output sequences $\mathbf{v}^{(j)}$, $j = 1, 2$, using $\mathbf{v} = \mathbf{u}\mathbf{G}$, where $\mathbf{v} = (v_0^{(1)}, v_0^{(2)}, v_1^{(1)}, v_1^{(2)}, ...)$ is the composite encoded sequence, also called a codeword, obtained by multiplexing the two encoded sequences, and $\mathbf{G}$ is the semi-infinite code generator matrix [7].

The encoded bits in 1/2-rate coding are generated at twice the input information rate. However, it is possible to find an encoding operation that relates blocks of input information bits to equal length blocks of encoded bits [1, 3].

Proposition: For 1/2-rate convolutional coding with constraint length ν, there exists a correspondence between equal length blocks of input information bits and the encoded bits. The length of these corresponding blocks is 2ν bits. (For proof, see [1, 3].)

We can formalise this relationship as follows. Let $[\mathbf{u}]_{t,t+2\nu-1} = (u_t, u_{t+1}, ..., u_{t+2\nu-1})$ be a 2ν-bit block of the input information sequence, and $[\mathbf{v}]_{2t,2t+2\nu-1} = (v_{2t}, v_{2t+1}, ..., v_{2t+2\nu-1})$ be the 2ν-bit block of the corresponding encoded sequence. Given the generator sequences $\mathbf{g}^{(j)}$, $j = 1, 2$, for a 1/2-rate convolutional code with constraint length ν, we define the *reduced encoding matrix* as

$$[\mathbf{G}]_{2\nu} = \begin{bmatrix} g_\nu^{(1)} & g_\nu^{(2)} & & & & \\ g_{\nu-1}^{(1)} & g_{\nu-1}^{(2)} & g_\nu^{(1)} & g_\nu^{(2)} & & \\ \vdots & \vdots & \ddots & \ddots & & \\ g_1^{(1)} & g_1^{(2)} & \cdots & \cdots & g_\nu^{(1)} & g_\nu^{(2)} \\ g_0^{(1)} & g_0^{(2)} & \cdots & \cdots & g_{\nu-1}^{(1)} & g_{\nu-1}^{(2)} \\ & & & & \vdots & \vdots \\ & & & & g_0^{(1)} & g_0^{(2)} \end{bmatrix}, \quad (1)$$

where blanks denote zeros. A 2ν-bit input information block is encoded to obtain the corresponding encoded block of length 2ν bits as follows:

$$[\mathbf{v}]_{2t,2t+2\nu-1} = [\mathbf{u}]_{t,t+2\nu-1}[\mathbf{G}]_{2\nu}. \quad (2)$$

For 1/2-rate convolutional coding with constraint length ν, an encoded sequence obtained by equation (2) can be uniquely decoded if and only if the inverse $[\mathbf{G}]_{2\nu}^{-1}$ of the reduced encoding matrix exists, i.e., $[\mathbf{G}]_{2\nu}[\mathbf{G}]_{2\nu}^{-1} = [I]_{2\nu}$, where $[I]_{2\nu}$ is the $2\nu \times 2\nu$ *identity matrix*. A 2ν-bit input information block is recovered by decoding the corresponding 2ν-bit encoded block as follows:

$$[\mathbf{u}]_{t,2t+t-1} = [\mathbf{v}]_{2t,2t+2t-1}[\mathbf{G}]_{2\nu}^{-1}. \quad (3)$$

It is desirable to find nonsystematic codes that can be decoded quickly (e.g. Quick-Look-In codes [7]). The decoding operation specified by equation (3) can be implemented as a look-up table (or can be built into the decoder hardware), and leads to the definition of a new class of convolutional codes that allow fast decoding. We define 1/2-rate convolutional codes with constraint length ν that have an invertible reduced encoding matrix $[\mathbf{G}]_{2\nu}$ as *invertible codes*. Note that both systematic and nonsystematic codes can belong to the class of invertible codes.

The received sequence $\mathbf{r}$ can be expressed as $\mathbf{r} = \mathbf{v} + \mathbf{e}$, where $\mathbf{e}$ is the error sequence. The syndrome sequence $\mathbf{s}$ is data independent and can be calculated one bit at a time as follows:

$$s_t = < [\mathbf{e}]_{2t,2t+2\nu+1}, \mathbf{h} >, \quad (4)$$

where $< \cdot, \cdot >$ indicates the vector dot product, $[\mathbf{e}]_{2t,2t+2\nu+1}$ is a $(2\nu + 2)$-bit block of the error sequence, and $\mathbf{h}$ is $(\nu + 1)$th column of the syndrome former $\mathbf{H}^T$ described in [4]. Since each received bit affects $(\nu+1)$ syndrome bits, it is sufficient to consider a finite length syndrome sequence denoted by $[\mathbf{s}]_{t,t+\nu}$ and called the *s-address*. In order to fully compute $[\mathbf{s}]_{t,t+\nu}$, $4\nu+2$ encoded bits are needed [1]. From equation (4),

$$[\mathbf{s}]_{t,t+\nu} = [\mathbf{e}]_{2t,2t+4\nu+1}[\mathbf{H}^T]_{(4\nu+2)\times(\nu+1)}, \quad (5)$$

where $[\mathbf{H}^T]_{(4\nu+2)\times(\nu+1)}$ is a submatrix of the syndrome former $\mathbf{H}^T$, starting with the $(\nu + 1)$th column, and $[\mathbf{e}]_{2t,2t+4\nu+1}$ is a $(4\nu + 2)$-bit block of the error sequence. Each $[\mathbf{e}]_{2t,2t+4\nu+1}$ is an error pattern of finite length corresponding to the s-address $[\mathbf{s}]_{t,t+\nu}$.

Equation (5) provides the basis for table-driven correction of errors. During the correction table generation phase, error patterns of length at least $(4\nu + 2)$ bits are considered. The s-addresses are calculated for the most likely error patterns. Typically, we first consider single bit errors. Next, two-bit errors are considered, etc. The bits in error can be located in any of the $(4\nu+2)$ positions. The bit pair $(e_{2t+2\nu}^{(1)}, e_{2t+2\nu+1}^{(2)})$ participates in the calculation of every bit of the s-address $[\mathbf{s}]_{t,t+\nu}$. This pair is called the *correction pair*, and is the location where correction will be applied. The correction information for the correction pair is denoted by $(c^{(1)}, c^{(2)})$, and is stored in the table at address $[\mathbf{s}]_{t,t+\nu}$. The table generation continues for higher order errors until all the entries in the table are filled.

Since there are more error bits than syndrome bits in equation (5), multiple error patterns of length $(4\nu + 2)$ can generate the same $(\nu + 1)$-bit s-address. All such patterns are syndrome-indistinguishable [4] (i.e. there is a conflict). The size of the correction table can be increased so that conflicting error patterns can map to different (longer) s-addresses. This is known as the syndrome extension method [1, 3]. Let the s-address be extended by s bits to $[\mathbf{s}]_{t,t+\nu+s}$. Then the corresponding correction window becomes $(4\nu+2)+2s$ bits long. The extension can be one-sided or two-sided. A two-sided extension results in a correction window $[\mathbf{r}]_{2t-s,2t+4\nu+s+1}$. This extension of the correction window allows our approach to perform better than minimum distance feedback decoding.

During the error correction phase, an s-address is calculated based on a block of the received bits. It is then used to access the correction table to retrieve the correction information. The correction is applied to the appropriate pair of received bits and the s-address is recalculated. This is the feedback step. If the new s-address is zero, the correction is assumed to be perfect. A new received bit pair is shifted into the correction window and the process is repeated. If the new s-address is not zero, the correction is reversed and the error is allowed to pass. Feedback decoders suffer from error propagation and appropriate measures need to be taken to avoid or control it.

For illustration, consider the performance of a 1/2-rate, constraint length $\nu = 7$ convolutional code with table-driven decoder shown in Figure 1. A binary symmetric channel (BSC) is assumed and a simple correction algorithms is used. More sophisticated algorithms that promise better performance are being studied. This corresponds to a table set up to correct all errors up to 4 bits and all 5,6 and 7-bit errors that do not have conflicting addresses. The model is seen to be better than the simulation results. This is because the model does not account for error propagation.

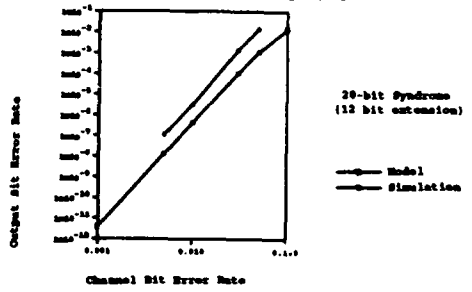


Figure 1: Table-Driven Error Correction Performance

References

[1] D.L. Bitzer and M.A. Vouk. A table-driven (feedback) decoder. In *Proc. of IPCCC '91*, Phoenix, AZ, March 1991.

[2] G.C. Clark, Jr. and J.B. Cain. *Error-Correction Coding for Digital Communications*. Plenum Press, New York, 1981.

[3] A. Dholakia, et. al. An efficient table-driven decoder for one-half rate convolutional codes. In *Proc. of 30th Annual ACM Southeast Conference*, Raleigh, NC, Apr. 1992.

[4] G.D. Forney, Jr. Convolutional codes I: Algebraic structure. *IEEE Trans. on Info. Theory*, IT-16(6):720–738, November 1970.

[5] R.G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[6] J.A. Heller. Feedback decoding of convolutional codes. In A.J. Viterbi, ed., *Adv. in Comm. Sys.*, vol. 4:261–278. Academic Press, 1975.

[7] S. Lin and D.J. Costello, Jr. *Error Control Coding - Fundamentals and Applications*. Prentice-Hall, Englewood Cliffs, NJ 07632, 1983.

SEQUENTIAL DECODING ON MEMORYLESS SOFT DECISION CHANNELS UNDER THE Pe-CRITERION

Ivonete Markman John B. Anderson

Electrical, Computer and Systems Engineering Department
Rensselaer Polytechnic Institute, Troy, NY 12180-3590

Abstract The Pe-criterion is a recent analysis [1] of sequential BSC channel decoding based on the design condition that decoders may fail with a probability $Pe > 0$. The result is a definite boundary for the tree search region and a well defined estimate of path numbers searched.

This work extends the criterion to the memoryless binary input soft decision channels resulting from the quantization of the AWGN channel output at the receiver, the binary input Q-ary output ($Q > 2$) and the binary input continuous output (semicontinuous) channels [2]. We show that the use of soft decisions implies a reduced search region compared to the BSC case and that large savings in the number of paths searched may be achieved when soft decision information is available.

Summary

We initially derive the shape of the search region in a generalized distance versus depth diagram. The generalized distance is a function of the metric for the standard sequential decoding analysis [3]. In the generalized distance versus depth diagram, the search region is bounded by the drop line, the set of points outside of which the correct path in the tree wanders with probability Pe or less. It is unique for each channel and each Pe, and independent of the code rate, R. Comparisons with the corresponding drop lines for the BSC under the same value of Pe show the decrease in the search region resulting from the use of soft decision (figure 1) .

We can then estimate the expected number of paths that a non-backtracking algorithm views within its search region, by applying an analytical method based on difference equations [1,2]. A second derivation of it is obtained for the semicontinuous channel, using an integral equation [2].

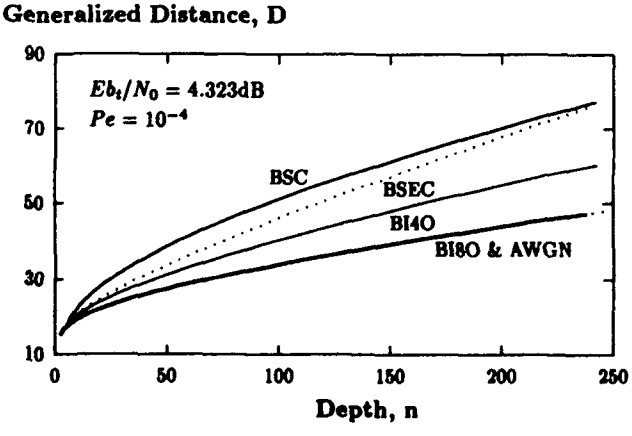


Figure 1: Drop Lines for the Binary Input Binary (BSC), Ternary (BSEC), 4-ary (BI4O), 8-ary (BI8O) and Continuous (AWGN) Output Channels.

A numerical analysis is then performed for the BSC and some equivalent soft decision channels. For a fixed Eb_t/N_0, the BSC and the semicontinuous channel are unique channels. This is not so for the Q-ary output channels with $Q > 2$, where a family of channels exists for a given Eb_t/N_0. Therefore, an additional optimization is necessary, in order to find out the best Q-ary output channel, in terms of minimum number of paths searched. Table 1 shows the results for $Eb_t/N_0 = 4.323$dB (which corresponds to the crossover probability $p = 0.01$ in the BSC case) and different values of R and Pe.

The results emphasize once again the importance of the use of soft decision in the decoding process, this time from the point of view of the number of paths searched. It is shown that large savings in path searching can be achieved, even when the most simple form of soft decision is applied ($Q = 3$). In addition, it is shown that, as opposed to common belief [4], soft decision savings are not necessarily related to an increase in the channel capacity, as compared to the hard decision channel (BSC). For many cases, the best soft decision channel has even smaller capacity than the corresponding BSC, even though it represents substantial savings in paths searched. Furthermore, the best Q-ary output quantization depends critically on R and Eb_t/N_0. The latter in turn shows that signal level and noise variance estimation, or equivalently automatic gain control (AGC), is important in the design of limited search decoders.

Acknowledgment The first author was supported by CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

References

[1] J.B. Anderson "Sequential Decoding Based on an Error Criterion", *IEEE Transaction on Information Theory*, vol.COM-32, no.5, May 1992.

[2] I. Markman and J.B. Anderson *The Pe-Criterion Sequential Decoder with Soft Decisions*, Report TR92-2, ECSE Dept., R.P.I., Troy, NY, March 1992.

[3] R.G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons Inc., New York, 1968.

[4] A.J. Viterbi and J.K. Omura, *Principles of Digital Communication and Coding*, McGraw Hill, New York, 1979.

Pe	10^{-6}		10^{-8}	
R	1/2	2/3	1/2	2/3
BSC	77.8	1.22E+4	417.8	3.20E+5
BSEC	6.4	92.8	15.1	462.3
BI4O	4.6	42.2	9.5	161.7
BI8O	2.9	16.3	5.1	45.4
AWGN	2.4	12.1	4.1	30.5

Table 1: Number of paths searched for the Binary Input Binary (BSC), Ternary (BSEC), 4-ary (BI4O), 8-ary (BI8O) and Continuous (AWGN) Output Channels ($Eb_t/N_0 = 4.323$dB).

OPTIMAL TRELLIS DECODING AT GIVEN COMPLEXITY

Tor Aulin
Department of Computer Engineering, Telecommunication Theory
Chalmers University of Technology, S-412 96 Göteborg, Sweden.

Abstract

A class of algorithms performing Maximum Likelihood Sequence Detection under various structural and complexity constraints is derived (BSC or AWGN). Complexity is measured by the number of paths used. By partitioning the S states into C classes and selecting B paths into each class, the signals closest to the received one shall be selected and hence N_P=BC paths are used. This class of algorithms has the name SA(B,C) (SA=Search Algorithm) and the Viterbi Algorithm (VA) is the unconstrained solution denoted SA(1,S).
An analysis method concerning the probability of the first error event at large SNR is developed for the whole SA(B,C) family and results in an analysis tool named the Vector Euclidean Distance (VED) of which the traditional Euclidean Distance (ED) is a scalar special case. The smallest number of paths resulting in the same asymptotic detection performance as the VA is calculated for several classes of trellis codes.

Summary

What limits the use of the VA is the number of states, S, which becomes very large if e.g. joint MLSD is applied to a whole system. By instead setting initially the number of paths to be traced in the trellis and the requiring that MLSD is to be performed, a family of MLSD procedures is the result with the complexity as a parameter. Structural constraints can also be imposed but this will be at the price of an increased number of paths. A structure can be given by partitioning the S states into C classes [2] and then in each iteration keeping B paths into each class. Assuming M-ary transmission, the BC paths will be extended to MBC , from which BC are selected again, in each recursion. This selection procedure is important and if the paths with the smallest ED (Hamming distance) are selected, MLSD will be performed for the AWGN channel (BSC).

The probability of a first error event for the AWGN channel is [1]

$$P(\varepsilon) \sim K_1 Q\left(\sqrt{d^2_{min}\frac{E_b}{N_0}}\right) + K_2 Q\left(\sqrt{d^2_{l,min}\frac{E_b}{N_0}}\right)$$

where the SNR, E_b/N_0 is large and d^2_{min} is the (normalized and squared) ED between any two different paths in the trellis. The first term is the traditional asymptotic error event probability for the VA whereas the second is an increment due to the non-exhaustive search of the trellis and is associated with the probability that the correct path is lost after the selection procedure. The quantity $d^2_{l,min}$ is determining the asymptotic probability of this CPL (correct path loss) and for reasons given below called the VED.

The VED d_l^2 for the most efficient member of the SA(B,C) family, SA(B,1), is associated with a B × B matrix Σ and a B × 1 vector μ whose entries are given by

$$\begin{cases} \sigma_{ij} = \frac{1}{2}(d_i^2 + d_j^2 - d_{ij}^2) & i, j = 1, 2, \ldots, B \\ \mu_i = \sigma_{ii} & i = 1, 2, \ldots, B \end{cases}$$

where d_i^2 is the ED between the correct path and contender #*i*. Also, d_{ij}^2 is the ED between contenders #*i* and *j*. When Σ has full rank,

$$d_l^2 = \min_{y \le 0} (\mu - y)^t \Sigma^{-1} (\mu - y)$$

and the minimization is performed over the constellation of B paths. Should there be C>1 classes, the minimization is first performed for each class and then over the classes. When Σ has rank R<B, a $R \times R$ sub-matrix is taken out from Σ and the constraints in the minimization are also changed [1]. It is always possible, however, to find a set of A paths called **active** (or outer) such that

$$d_l^2 = \mu_A^t \Sigma_{A,A}^{-1} \mu_A$$

where the dimensions are $A \times 1$ and $A \times A$ respectively [1].

To achieve the same asymptotic detection performance as the VA, it is necessary to have $d^2_{l,min} \ge d^2_{min}$. By considering SA(B,1), the most efficient member of the SA(B,C) family (and also any other algorithm), it is now possible to find the least B for which this inequality is satisfied, having the notation B*. For the class of trellis codes built up from convolutional codes and antipodal modulation the result is $B^* \sim \sqrt{S}$ which is asymptotic in the sense that the number of states is very large. By considering codes having from say 64 to 1024 states the asymptotic result is good aslo for these. Yet another class of trellis codes, namely r=2/3 convolutionally encoded 8PSK (often referred to as Trellis Coded Modulation, TCM) the same result still applies. For Continuous Phase Modulation (CPM) no simple rule in complexity reduction, S/B*, to establish asymptotic detection optimality is applicable but substantial savings are demonstrated, see example below.

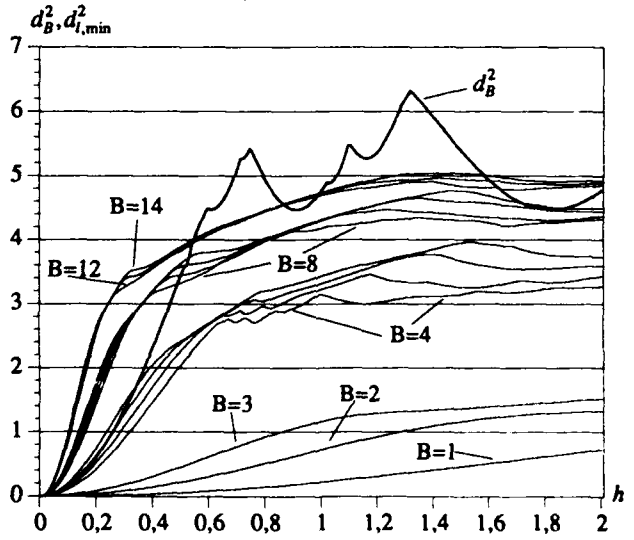


Figure. $d^2_{l,min}$ and d^2_{min} (actually d_B^2) for M=4, 3RC CPM. SA(B,1).

The developed analysis tools are general so that coded systems and systems where the channel memory has been included in the system model (trellis) now can be exactly analyzed.

References

[1] T. Aulin "Breadth First Maximum Likelihood Sequence Detection", submitted for publication in *IEEE Trans. on Information Theory*, October 1992.

[2] T. Larsson "A State-Space Partitioning Approach to Trellis Decoding", Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, December 5, 1991.

BIDIRECTIONAL SEQUENTIAL DECODING ALGORITHMS *

Kaiping Li and Samir Kallel
Department of Electrical Engineering
The University of British Columbia
Vancouver, B.C., Canada, V6T 1Z4

Abstract

In this paper, we present efficient bidirectional sequential decoding (BSD) techniques. With BSD, the code tree is searched from the root and end states of the encoded tree simultaneously. It is shown by analysis as well as computer simulations that with BSD, the computational variability per decoded block can be substantially decreased. In fact it is shown that the Pareto exponent of the distribution of the block computational effort with BSD is twice that with unidirectional sequential decoding (USD). Good codes suitable for BSD are found. Also, an efficient bidirectional multiple stack algorithm (BMSA) is proposed and analyzed. This BMSA offers a good trade-off between computational effort and error performance.

Summary

Sequential decoding is a very powerful decoding technique for convolutional codes. The main drawback of sequential decoding is the variability of its decoding effort. As a consequence of this variability, the decoding effort for a given data block may exceed, in certain situations, the physical limitations of the decoder, leading inevitably to buffer overflows and information erasures. In the past, several modifications have been developed to reduce the computational variability of sequential decoding [1–4]. In this paper, new decoding techniques which further alleviate this drawback of sequential decoding are proposed and analyzed.

In a system using convolutional coding and sequential decoding, information is usually transmitted in blocks, and each block is terminated by a tail of some known bits. Starting from the root node of the encoded tree, a conventional sequential decoder moves into the tree in the forward direction, one branch at a time, along the most likely transmitted path. Decoding of a block is terminated whenever the decoder reaches the end node of the tree. Since the final encoder state is known by the decoder, decoding can also be performed in the backward direction.

We propose in this paper an efficient sequential decoder that explores the tree simultaneously in both forward and backward directions. The bidirectional search idea has been used for computing the free distance of convolutional codes [5, 6]. Recently, Rouanne and Costello have applied the bidirectional stack algorithm for computing the distance spectrum of trellis codes [7]. This idea of bidirectional decoding has also been applied to the *M-algorithm* [8], and to the decoding of block codes [9]. The BSD algorithm proposed in this paper is based on the well known stack algorithm [1], and hence it is called ***bidirectional stack algorithm*** (BSA). In the BSA, two separate stacks are used. One is used for the forward search of the tree and is called the *forward stack* (FS). The other stack is used for the backward search and is called the *backward stack* (BS). Starting from the root and final states of the encoder, forward and backward search operations are performed simultaneously according to the regular stack algorithm. The tree search is terminated whenever the best path on the forward or the backward direction merges with a path on the opposite direction.

With BSD, it is desirable to use codes that possess the same distance profile on both forward and backward directions. Using computer search techniques, we have found good non-systematic rate-1/2 codes suitable for BSD.

It is shown by analysis and computer simulations that with BSD, the computational variability per decoded block can be substantially decreased. In fact it is shown that the Pareto exponent of the distribution of the block computational effort with BSD is twice that with USD.

The idea of bidirectional sequential search can be incorporated to the *multiple stack algorithm* (MSA) [4]. An efficient *bidirectional multiple stack algorithm* (BMSA) is described and analyzed. It is shown that this BMSA offers very good performances in terms of both error probability and computational efforts.

References

[1] F. Jelinek, "Fast Sequential Decoding Using a Stack," *IBM J. Res. Develop.*, vol. 13, pp. 675–685, Nov. 1969.

[2] G. D. Forney, Jr. and E. K. Bower, "A High-Speed Sequential Decoder: Prototype Design and Test," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 821–835, Oct. 1971.

[3] D. Haccoun and M. J. Ferguson, "Generalized Stack Algorithms for Decoding Convolutional Codes," *IEEE Trans. on Inform. Theory*, vol. IT-21, pp. 638–651, Nov. 1975.

[4] P. R. Chevillat and D. J. Costello, Jr., "A Multiple Stack Algorithm for Erasurefree Decoding of Convolutional Codes," *IEEE Trans. on Comm.*, vol. COM-25, pp. 1460–1470, Dec. 1977.

[5] L. R. Bahl, C. D. Cullum, W. Frazer, and F. Jelinek, "An Efficient Algorithm for Computing Free Distance," *IEEE Trans. on Inform. Theory*, vol. IT-18, pp. 437–439, May 1972.

[6] K. J. Larsen, "Comments on 'An Efficient Algorithm for Computing Free Distance'," *IEEE Trans. on Inform. Theory*, vol. IT-19, pp. 577–579, July 1973.

[7] M. Rouanne and D. J. Costello, Jr., "An Algorithm for Computing the Distance Spectrum of Trellis Codes," *IEEE J. on Selected Areas in Commun.*, vol. 7, pp. 929–940, Aug. 1989.

[8] D. Haccoun and J. Belzile, "Bidirectional Algorithms for the Decoding of Convolutional Codes," *1990 IEEE Book of Abstracts of Information Theory Symposium, San Diego*, p. 177, Jan. 1990.

[9] F. Hemmati, "Bidirectional Trellis Decoding," *1990 IEEE Book of Abstracts of Information Theory Symposium, San Diego*, p. 107, Jan. 1990.

* This research was supported by the National Sciences and Engineering Research Council of Canada.

ON THE COMPUTATION PROBLEM OF THE STACK AND FANO DECODERS FOR SPECIFIC TIME-INVARIANT CONVOLUTIONAL CODES

K. Muhammad and K. Ben Letaief

Electrical & Electronic Engineering Department
The University of Melbourne
Parkville, Victoria 3052, Australia.

It is well known that the behavior of sequential decoding is limited by its computational effort [1]. Let C denote the number of tree nodes examined in order to make a correct decision. Then the *distribution of computation* [1] is the distribution of C. Traditionally, the performance of sequential decoders has been analyzed using *random coding arguments* which obtain results in the form of averages over the ensemble of random tree codes [1]. Based on this analysis, it is well known that the distribution of C is essentially Pareto distributed, a function of code rate, but independent of the code's *constraint length*, K [1]. However, by their very nature, ensemble averages are not tied directly to the properties of a particular code and therefore, techniques for the performance analysis of specific codes are highly desirable.

In this paper, we investigate the performance of the *Fano* and *stack* decoders [1] using *exact* analysis methods based on *importance sampling* [2]. In contrast to the classical analysis, the simulation-based analysis presented in this paper uses no random-coding arguments and is applicable to specific time-invariant convolutional codes. Hence, it serves as a useful complement to the ensemble average analysis as one can study the characteristics of sequential decoders for any given code and operating condition.

Fig. 1 shows the computational effort of various convolutional codes operating over a *binary symmetric channel* (BSC) [1] and employing the stack decoder. The simulation results show an interesting effect; the computational effort improves as K decreases. This shows the effect of the "remerging phenomenon" [1] on the computational effort of a sequential decoder. In brief, for a code with small K, the incorrect paths traced by the decoder tend to merge more often with the correct path, thereby resulting in an undetectable error. Hence, for a given SNR ratio, the distribution of C actually depends[1] on K. However, as K increases, this dependency becomes less significant and the distribution tends to converge with the Pareto tail. Note that the classical analysis [1] shows no such dependence and/or characteristics.

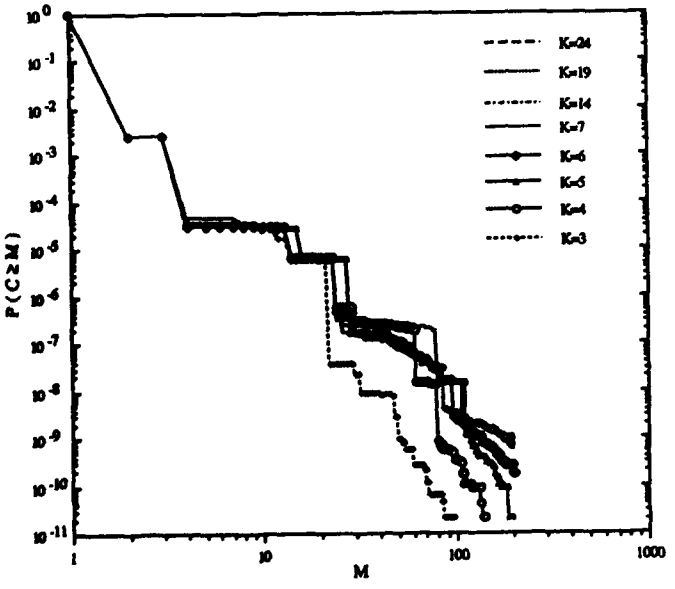


Fig. 1: Sequential Decoding Computational Effort Performance Over a BSC with SNR = 9 dB.

For the hard-quantized memoryless channels, the distribution of C presents another interesting feature; the curves exhibit a stair-case appearance before assuming the usual asymptotic Pareto behavior. This effect becomes more obvious as SNR increases. For low-to-moderate values of SNR, however, the curves present smooth appearance without any abrupt changes. It is noted that this phenomenon is not apparent in case of the unquantized *additive white Gaussian noise* (AWGN) channel.

It is well known that *optimal distance profile* (ODP) and *optimal free distance* (OFD) codes make excellent choices for sequential and Viterbi decoding, respectively [1]. To compare the relative performance of the two types of codes, several simulations were conducted. It was verified that the ODP codes perform better than OFD ones for all SNR conditions. The results of the study also indicated that systematic ODP codes performed much better than the non-systematic ones. However, this improvement in the distribution of C performance was again, at the expense of increased error probability.

Throughout this study, we used the *Fano metric* [1] with the bias term, B, equal to the code rate, R. Several simulations were conducted in order to investigate a good value of B. Using some recently developed techniques by the authors [2] for estimating the bit error rate (BER), the effect of B on BER was also considered. A summary of the results of this investigation will be presented.

The performance of the Fano decoder depends on the value of threshold increment, Δ [1]. Our simulations indicate that when Δ is increased initially, the computational effort (determined by considering the forward looks) of the Fano decoder improves. As Δ is increased further, the computational effort severely degrades. In contrast, the distribution of the number of (distinct) nodes searched in order to make a specific correct decision always degrades as Δ is increased. This dependence, however, becomes less significant for higher SNR ratios.

Finally, the relative performance of hard-decision decoding versus soft-decisions was investigated. An interesting result was obtained; the computational effort over the unquantized and hard-quantized AWGN channel is almost identical when the loss associated with hard decisions is about 2.3 dB. It was also found that the BER's for the two operating conditions were identical. Hence, it seems that if a sequential decoder operates over an unquantized AWGN channel, then to achieve the same performance over a hard-quantized AWGN channel, the decoder must generate an additional signal power of the order of 2.3 dB.

References

[1] S. Lin and D. J. Costello, "Error Control Coding: Fundamentals and Applications," *Prentice-Hall, New-Jersey*, 1983.

[2] K. Muhammad, "New performance analysis techniques for coded communication systems employing sequential and Viterbi decoders," *M.Eng.Sc. dissertation, University of Melbourne, Melbourne, Australia*, Oct. 1992.

[1] It is noted that in the ensemble average analysis, the path merging phenomenon is carefully avoided by the use of long constraint length codes.

Sequential Decoding of Linear Block Codes

D. J. Tempel and E. Shwedyk
Dept. of Electrical and Computer Engineering
University of Manitoba
Winnipeg, MB
R3T 2N2

Abstract

This paper describes the use of the sequential stack algorithm to decode cyclic (or extended cyclic) block codes. Once a block code is endowed with a trellis structure decoding with any of the convolutional decoding algorithms is viable. Since trellises for block codes are very wide a sequential algorithm, working at moderate signal-to-noise ratios, is an effective decoding alternative to the Viterbi algorithm. Using Wolf's trellis, Chang and Yao's sequential stack algorithm, and the Fano metric the (24,12) Golay code can be efficiently decoded. Computer simulations show that by 6 dB the sequential algorithm is the most efficient (using Be'ery and Snyders' definition of complexity) soft decoding algorithm for the (24,12) Golay code.

Summary

Owing to their algebraic properties linear block codes are typically decoded using algebraic techniques. Generally, these algebraic techniques make hard decisions on the received bits causing an inherent loss of 2 dB in error performance. On the other hand, convolutional codes are decoded using the Viterbi algorithm or a sequential algorithm which use soft decisions and hence have a 2 dB advantage over block codes. Therefore, the ability to extend the convolutional decoding techniques to block codes would clearly be advantageous. By applying the Viterbi algorithm to a trellis [Wolf78] block codes can be decoded using soft decisions. Unfortunately, the width of this trellis grows exponentially with the number of parity symbols, thereby, making the Viterbi algorithm inefficient. A solution to this problem is the use of a sequential decoding algorithm.

Given a trellis a convolutional decoding algorithm such as the sequential stack algorithm can be applied. A trellis for a cyclic or extended cyclic code can be constructed by using the code's shift register encoder [Wolf78]. The advantage of using the encoder is that the sequential algorithm can generate trellis states as needed rather than having to store the complete trellis beforehand. This investigation uses an improved stack algorithm which stores partial paths in a priority queue [ChYa86]. The priority queue is highly parallel and hence most comparisons are done simultaneously making the algorithm all the more efficient. The Fano metric, used when decoding convolutional codes, is used as a measure to determine the best path through the trellis.

Following Snyders and Be'ery [SnBe89] complexity is measured in terms of equivalent real number additions. The sequential algorithm manipulates four pieces of information (viz. state, metric, path, and depth). Of these, the metric is the only real number and thus the two metric operations solely comprise the complexity. The first operation is metric addition which is performed when the branch metric is added to the path metric. The second operation is metric comparison which is performed in the priority queue after every deletion or insertion.

As a comparison, the complexity of several soft decision techniques used to decode the (24,12) Golay code are listed below. Simulations (AWGN channel with BPSK modulation) were performed to measure the complexity and error performance of the stack algorithm. Based on these simulations the sequential algorithm is the better algorithm for decoding the Golay code when the signal-to-noise ratio is at least:

Technique	Maximum Complexity		snr
Correlation Decoder	98303		2 dB
Viterbi Algorithm	20473		3 dB
Conway-Sloane (86)	1614	*	5 dB
Be'ery-Snyders (86)	1551	*	5 dB
Forney (88)	1351	*	5 dB
Snyders-Be'ery (89)	827	*	6 dB
Vardy-Be'ery (91)	651	**	6 dB

* [SnBe89] ** [VaBe91]

As signal-to-noise ratios increase the sequential algorithm quickly becomes the most efficient algorithm for decoding block codes. For high signal-to-noise ratios the algorithm approaches its minimum decoding complexity of 292 addition equivalent operations. The simulations also confirm that the sequential algorithm performs maximum likelihood soft decision decoding.

References

[ChYa86] Chang, C.Y., and Yao, K., "Systolic Array Architecture for the Sequential Stack Decoding Algorithm," SPIE vol. 696 Advanced Algorithms and Architectures for Signal Processing, 1986, pp. 196-203.

[SnBe89] Snyders, J., and Be'ery, Y., "Maximum Likelihood Soft Decoding of Binary Block Codes and Decoders for the Golay Codes," IEEE Trans. IT, Sept. 1989, pp. 963-975.

[VaBe91] Vardy, A., and Be'ery, Y., "Even More Efficient Soft Decoding of the Golay Codes," Proc. IEEE ISIT, Budapest, Hungary, June 24-28, 1991, p. 190.

[Wolf78] Wolf, J.K., "Efficient Maximum Likelihood Decoding of Linear Block Codes Using a Trellis," IEEE Trans. IT, Jan. 1978, pp. 76-80.

A TREE-STRUCTURED POLYTOPAL VECTOR QUANTIZER FOR REAL-TIME IMAGE CODING

Shih-Chi Huang[1]
CAERE Corporation
100 Cooper Court
Los Gatos, CA 95030

Yih-Fang Huang
Laboratory for Image and Signal Analysis
Department of Electrical Engineering
University of Notre Dame.

Abstract

Recently, a tree-structured polytopal vector quantization scheme referred to here as the Principal Component Vector Quantization algorithm (PCVQA), has been developed and has been shown to have the design complexity only linearly proportional to the codebook size. This paper proposes an efficient technique to implement PCVRA so that it can be made viable for the real-time environment.

I. An Overview

Vector quantization (VQ) has been considered as a viable technique for still image data compression [1-2]. One of its advantages is that although its encoding is a complex operation, its decoding is a simple table look-up [3]. However, many VQ techniques, especially those clustering-based ones, are seriously hampered by complexity, particularly design complexity. It is well known that the design complexity of unconstrained VQ algorithms are exponentially proportional to both the codebook size and the dimension of input vectors. An effective approach to reducing the complexity is to consider constrained VQ techniques. A recently developed VQ technique, referred to here as the principal component vector quantization algorithm (PCVQA), is one such approach.

It was shown that the design complexity of PCVQA is only linearly proportional to the codebook size [5,6]. The fundamental concept of PCVQA rests on the use of principal component as the normal direction of the partitioning hyperplanes in designing a tree-structured polytopal VQ. Thus, the design complexity of PCVQA is critically related to the complexity of estimating the principal components.

The objective of this paper is to develop techniques for further reducing the implementational complexity of PCVQA. Of particular interests here are still image compression for real-time codebook retransmission in applications including HDTV broadcasting. The issues addressed here are those of complexity, quality of coded images measured by peak signal-to-noise ratio, and transmission bit rates.

The proposed approach is to implement PCVQA in combination with the method of self-organizing codebook [7] and JPEG (Joint Photographic Experts Group) [8] for still image compression. Simulation results show that this approach can achieve better performance than JPEG does alone. The price for achieving such good performance is only a slight increase in the design complexity. The amount of such complexity increase is governed by the complexity of calculating the principal components.

Four numerical methods are examined here for calculating the principal components, namely, the gradient descent method, the power method, the eigenvalue shift acceleration, and the modified Aitken's δ^2 acceleration [9].

This paper also considers a local search encoding scheme to further improve PCVQA's performance especially when the input vectors tend to be clustered as in the case of transform VQ. It known that in coding high definition images, it is desirable that the input vectors to the vector quantizer be high dimensional to reduce the transmission bit rate. Thus transform techniques such as DCT (discrete cosine transform) are needed to reduce the input vector dimension.

II. The Proposed Scheme

The proposed image coding scheme is summarized here. It first divides an image into 8×8 pixels of subimage blocks. The DC coefficient of DCT is subtracted from each block and is coded separately. A codebook is then designed by the PCVQA for the subtracted blocks which contain only AC signals. The self-organizing encoding method [7] is employed to interleave transmission (or storing) of the codewords and labels. In this way, the codewords will be coded in a more compact way by JPEG. The basic idea for the self-organizing encoding is that if a codeword is selected the first time, it is transmitted (or stored). Otherwise, only its label is transmitted (or stored). Therefore, the codebook is self-organized in the sense that, the first selected codeword is shifted to the first position in the codebook buffer and all the codewords above this codeword are moved down. The decoder will reconstruct a codebook in the same order. Note that only the codewords that are transmitted (or stored) will be encoded by JPEG. Thus, without considering the codebook design complexity, this approach involves even less encoding complexity than JPEG does.

We then examine various numerical techniques, namely, the gradient descent method, the power method, the eigenvalue shift acceleration, and the modified Aitken's δ^2 acceleration, for efficient estimation of the principal components which are essential to the implementation of PCVQA. We show that the gradient descent method which uses the Rayleigh quotient as an estimate for the largest eigenvalue has the same performance as the power method. The power method with the eigenvalue shifted by one-third of the predicted largest eigenvalue has more rapid convergence. And the eigenvalue shift by the Aitken's δ^2 acceleration has the most rapid convergence. Consequently, we employ Aitken's δ^2 acceleration to implement the proposed compression scheme. This amounts to a complexity of only three to five times more than that of JPEG alone, according to our simulation experience. This implies that the proposed image coding scheme can, in general, code a color image of 256×256 pixels in less than one minute.

III. Concluding Remarks

We introduce here an image coding technique which combines PCVQA, the self-organizing codebook and JPEG. Simulation results demonstrate that this technique can achieve better performance for still image compression than JPEG alone. Numerical methods for efficiently implementing PCVQA are also studied. The proposed scheme, thanks to much reduced complexity, can be employed to construct codebooks that are to be retransmitted regularly in HDTV broadcasting.

References

[1] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Publishers, 1992.

[2] N. M. Nasradadi and R. A. King, "Image Coding Using Vector Quantization: A Review", *IEEE Trans. Commun.*, Vol. COM-36, pp. 957-971, August 1988.

[3] K. L. Hehler, E. A. Riskin and R. M. Gray, "Unbalanced Tree-Growing Algorithms for Practical Image Compression", *Proc. 1991 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Vol. 4 (Toronto, Canada), pp. 2293-2296, May 14-17, 1991.

[4] S. C. Huang and Y. F. Huang, "A Neural Network Structure for Vector Quantizers," *Proc. IEEE 1991 Int. Symp. Circuits Syst.*, pp. 2506-2509, Singapore, June 1991.

[5] S. C. Huang, *Multilayer Perceptrons for Image Data Compression and Speech Recognition*. Ph.D. Dissertation, University of Notre Dame, Notre Dame, IN, Dec. 1991.

[6] S. C. Huang and Y. F. Huang, "Principal Component Vector Quantization," *J. Visual Commun. Image Representation*, Vol. 3, No. 1, March 1993 (to appear).

[7] L. Wang, M. Goldberg and S. Shlien, "Interleaved Image Adaptive Vector Quantization," *Proc. 1991 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Vol. 4 (Toronto, Canada), pp. 2305-2308, May 14-17, 1991.

[8] G. K. Wallace, "The JPEG Still Picture Compression Standard," *Communications of the ACM*, vol. 34, pp. 30-45, April 1991.

[9] A. Jennings, *Matrix Computation for Engineers and Scientists*. San Francisco: John Wiley & Sons, 1977.

[1] Formerly with the Laboratory for Image and Signal Analysis, Department of Electrical Engineering, University of Notre Dame.

INTRODUCTION TO TEMPLATE CODING: AN ALTERNATIVE TO SUBPICTURE CODING IN BLACK-WHITE IMAGE COMPRESSION

John C. Kieffer and Greg Nelson*
Dept. of Electrical Engineering
University of Minnesota
Minneapolis, MN 55455

For each positive integer n, let $\mathcal{M}_n$ be the set of all $n \times n$ matrices of zeroes and ones containing at least one "1". We suppose that a black-white image which is not all white is represented as an element $M \in \mathcal{M}_N$ for a large enough N. For lossless encoding of M, we contrast two possible image compression methods. One method, called template coding, is a multiresolution technique which finds a string of templates (shapes) from a finite dictionary that can be used to successively reconstruct M starting from the matrix [1] in $\mathcal{M}_1$; the string of templates is then encoded template-by-template. The other method is the classical image compression technique known as subpicture coding, in which M is partitioned into square sub-blocks of a given size which are then encoded sub-block by sub-block.

We first discuss template coding of M. An integer parameter k is fixed, $2 \le k < N$. If A is a square zero-one matrix of order $\ge k$, we define $C(A)$ (the "core" of A) to be the largest square submatrix of A lying in the upper left corner of A whose order is divisible by k, and we define $P(A)$ (the "projection" of A) to be the matrix we obtain from $C(A)$ by replacing each of the submatrices in the partitioning of $C(A)$ into $k \times k$ submatrices with a one or zero depending upon whether the submatrix does or does not contain a "1". Template coding of M is performed in four steps: (1) form the matrices $\{M^i : i = 1,2,...,t\}$ where $M^1 = C(M)$, $M^2 = C(P(M^1))$, ..., $M^t = C(P(M^{t-1}))$, $P(M^t) = [1]$; (2) take the template dictionary D to be the union of the set $\{0,1\}$ and the set of matrices in $\mathcal{M}_k$ that appear in the partitions of the M^i into $k \times k$ submatrices; (3) form the string of templates from D that are seen as one horizontally scans each of the following in the order described: the elements of the partition of M^t into $k \times k$ submatrices, the elements of $P(M^{t-1})$ not in M^t, the elements of the partition of M^{t-1} into $k \times k$ submatrices, the elements of $P(M^{t-2})$ not in M^{t-1},..., the elements of the partition of M^1 into $k \times k$ submatrices, the elements of M not in M^1; (4) encode the template string template-by-template.

In subpicture coding of M, we also fix an integer parameter k, $1 \le k < N$. We form the string consisting of the elements of the partition of $C(M)$ into $k \times k$ submatrices (scanned horizontally) followed by the elements of M not in $C(M)$ (scanned horizontally). We then encode this string entry-by-entry.

We state our results, which indicate that template coding is preferable to subpicture coding in a certain asymptotic sense. If $M \in \mathcal{M}_n$ and $n > k$, let $B_k[M,t]$ ($B_k[M,s]$) be the minimum total number of bits that are achievable in template coding (subpicture coding) of M with integer parameter k. Then statements (i) (ii) hold:

(i) Given any k and $\varepsilon > 0$, there exists $k^* = k^*(k,\varepsilon)$ and $N^* = N^*(k,\varepsilon)$ such that $B_{k^*}[M,t] \le (1+\varepsilon)B_k[M,s]$ for any $n \ge N^*$ and any $M \in \mathcal{M}_n$.

(ii) Let S be any nonempty subset of the unit square whose box-counting dimension is less than two, and for any n let $M_n \in \mathcal{M}_n$ be the matrix in which an element equals zero if and only if the corresponding sub-square in the partition of the unit square into $n^{-1} \times n^{-1}$ sub-squares contains no point of S; then, for any k, the ratio $B_2[M_n,t]/B_k[M_n,s]$ converges to zero as $n \to \infty$.

Statement (i) tells us that template coding always yields compression performance at least as good as subpicture coding. Statement (ii) gives us a wide class of black-white images for which template coding outperforms subpicture coding.

We illustrate with an example. Template coding with k=2 applied to the matrix

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

yields $M^1 = M$, $M^2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$, and $M^3 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$.

Letting T_1, T_2 be the templates $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ respectively, we obtain the string of templates $\mathbf{t} = (T_1, T_2, T_1, T_1, T_2, T_2, T_1, T_2, T_1, T_2, T_2, T_1, T_1)$. (The first entry of $\mathbf{t}$ allows one to determine M^3, entries 2-4 allow one to reconstruct M^2 from M^3 and entries 5-13 allow the recovery of M from M^2.) We have $B_2[M,t] = 13$ bits. On the other hand, subpicture coding with k=2 applied to M gives us the string $(\underline{0}, \underline{0}, T_2, T_2, \underline{0}, \underline{0}, T_1, \underline{0}, \underline{0}, T_2, \underline{0}, T_1, T_2, T_2, T_1, T_1)$ where $\underline{0}$ is the 2×2 zero matrix, from which one concludes that $B_2[M,s] = 25$ bits.

*Authors supported by NSF Grant NCR-9003106

SUBPIXEL ACCURACY FOR DIGITIZED STRAIGHT LINES

Jack Koplowitz
Department of Electrical & Computer Engineering
Clarkson University
Potsdam, NY 13699-5720

Abstract

A straight line $y = mx + b$ has as its digital representation on an integer grid the set of points $\{(x_i, y_i): x_i = i, y_i = \lfloor mx_i + b \rfloor, i = 0, 1, \ldots n\}$. It is shown that for a uniform distribution on the set of lines, the error in estimating the line from its digital representation is O(log n/n).

Summary

For a line $y = mx + b$ the corresponding digital straight line is defined as those points on or immediately below the line $y = mx + b$, i.e. the set of points

$$\{(x_i, y_i): x_i = i, y_i = \lfloor mx_i + b \rfloor, i = 0, 1, \ldots n\} \quad (1)$$

Without loss of generality assume $0 \le m \le 1$, $0 \le b \le 1$. The question addressed here is if it is known that the edge is a straight line how well can we estimate the line from its digital representation. For simplicity we define the error at x as $e(x) = |mx + b - m'x - b'|$, where $m'x + b'$ is the estimated edge, and the error ε as

$$\varepsilon = \max_x \{e(x)\}, \; 0 \le x \le n \quad (2)$$

We assume a uniform measure on ρ and θ, the length and angle respectively of the normal to a straight line, producing a non-uniform measure on the set of digital straight lines. Generally this gives greater measure to digital straight lines with larger estimation error, affecting the order of the error. The expected value of the error ε will be shown to be upper and lower bounded by O(logn/n).

The order of the error defined in (2) is the same as the order of d_{min}, the minimum distance to the line of points (x_i, y_i), i = 0, 1, ...n, in (1). We first give a proof of the result that if an arbitrary small set of lines can be neglected then the error in estimating a line from its digitized representation has error O(1/n). More precisely we prove the following.

Theorem 1. *For any function f(n), increasing arbitrarily slowly with n,* $P\{d_{min} < f(n)/n\} > 1 - 8/f(n)$.

Proof: For a line $y = mx + b$, let $d_i = mx_i + b - y_i$, i = 0, 1, ...n, where y_i is defined in (1). Assume for the moment b = 0. For $m = p/q$, $0 < p < q \le n$, p and q relatively prime, then $d_i \in \{k/q, k = 0, 1, \ldots q-1\}$. Equivalently, distances to the line for points on the array immediately above the line are in the set $\{kq, k = 1, 2, \ldots q\}$. Thus b cannot be increased by more than 1/q, while keeping the digital straight the same. Similarly, for any b, its value can range only in an interval of width 1/q, for a fixed digital straight line. Thus,

$$d_{min} = \min_i \{d_i\} < 1/q \quad (3)$$

Consider a line with $0 < m < 1$. From number theory, m can be approximated by p/q, $0 < p < q \le n$, p and q relatively prime, such that

$$|m - p/q| < 1/qn \quad (4)$$

For any m its corresponding q will be the maximum satisfying (4). For $y = mx + b$ and $y = (p/q)x + b$, the maximum distance between them is $\max_x \{|y - y|\} < 1/q$, since $0 \le x \le n - 1$. Thus for any line $d_{min} \le 2/q$.

The set of slopes with approximation p/q in (4) is a subset of the interval $(p/q - 1/qn, pq + 1/qn)$. It's measure (assuming a uniform distribution on θ) is bounded by 4/qn. Let m_q denote the set of slopes with fixed q in the approximation in (4). Its measure is

$$u(m_q) < \phi(q)\,(4/qn) < 4/n \quad (5)$$

where $\phi(q)$ is the number of values p can take on (and be relatively prime to q).

The measure u of all m having $\le 2n/f(n)$ is

$$u < [2n/f(n)]\,[4/n] = 8/f(n) \quad (6)$$

For $q \le 2n/f(n)$ we have that $d_{min} < f(n)/n$, from which the theorem follows.

The expected error is of the same order as the expectation of d_{min}. Hence we show the following.

Theorem 2. $O(\log n/n) \le E\{d_{min}\} \le O(\log n/n)$

Proof. From the proof of Theorem 1 we have that for any line $d_{min} \le 2/q$, with q defined in (4). Thus from (5)

$$E[d_{min}] \le \sum_{q=1}^{n} (2/q)\,(4/n) = 8/n \sum_{q=1}^{n} 1/q = O(\log n/n) \quad (7)$$

To lower bound $E[d_{min}]$ consider $m \in m_\alpha = \{p/q, 0 < p < q \le n\}$ with p and q relatively prime. The intercept b can range over an interval of width 1/q without changing the digital straight line. The measure of all lines bounded from above by $mx + a + 1/q$ and below by $mx + a$ is

$$u_q \ge 1/2q^2n \quad (8)$$

which lower bounds the measure of all lines with that digital representation.

For a line $y = mx + b$, $m \in m_\alpha$ and $b \in b_\alpha = \{k/q, k=0, 1, \ldots q-1\}$ the ordered sequence $\{d_i = mx_i + b - y_i, i=0, 1, \ldots n\}$ is unique to the triplet (p, q, b). Taking the set of digital straight lines determined by these triplets we get

$$E[d_{min}] \ge \sum_q \sum_p \sum_b E[d_{min} \mid p,q,b]\, u_q \ge \sum_q \sum_p \sum_b (1/q)(1/2q^2n)$$

$$E[d_{min}] \ge \sum_q \varphi(q)/2q^2n = (1/2n) \sum_{q=1}^{n} \varphi(q)/q^2 \quad (9)$$

where $\varphi(q)$ is the number of integers $p < q$ and relatively prime to q. From (9) it can be shown $E[d_{min}] \ge O(\log n/n)$.

References

[1] J. Koplowitz, "Maximum likelihood slope estimation for reconstruction of digitized line segments," Proc. Conf. Inform. Sci. and Systems, Princeton, NJ, Mar. 1984, p. 43.

[2] L. Dorst and A.W.M. Smeulders, "Discrete representation of straight lines," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, pp. 450-463, 1984.

[3] C. Berenstein, L.N. Kanal, D. Lavine, and E.C. Olson, "A geometric approach to subpixel accuracy," Comput. Vision, Graphics, Image Process., vol. 40, pp. 334-360, 1987.

[4] C. Berenstein and D. Lavine, "On the number of digital straight line segments," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-10, pp. 880-887, 1988.

ENTROPY-CONSTRAINED SUBBAND CODING OF IMAGES USING A PERCEPTUAL DISTORTION CRITERIA

C.F. Harris and J.W. Modestino
Electrical, Computer and Systems Engineering Department
and
Center for Image Processing Research
Rensselaer Polytechnic Institute
Troy, New York 12180

Abstract

A new perceptually relevant entropy-constrained coding scheme based on the just-noticable-distortion (JND) level of the human observer is described and its properties demonstrated. The JND at each pixel location is defined as the threshold of detectability of the human visual system (HVS) to errors in reproducing that pixel. Because of the masking effect of the HVS, errors below the JND are rendered imperceptible. The JND is determined empirically as a function of spatial frequency, local texture and local contrast. A distortion measure is developed, making essential use of the JND, for a subband coding environment which attempts to mimic the subjective evaluation effects of the HVS. This distortion measure employs a weighted squared-error metric, where the weighting depends upon the JND value at each pixel position. It essentially assigns near-zero distortion to subthreshold errors and approximately squared-error distortion to superthreshold errors. This perceptual distortion measure was incorporated into a previously developed design procedure for entropy-constrained subband coding (ECSBC) schemes based upon training data. We demonstrate that, compared to use of the conventional squared-error distortion, significant improvements in subjective image reconstruction quality can be achieved at low average bit rates using this perceptual distortion measure.

Summary

Image compression is a very important area of research today, especially for use in bandwidth intensive applications such as high-definition television (HDTV) and multimedia systems. The aim of image compression, or coding, is to minimize the average distortion, as indicated by a specified fidelity or distortion measure, for a fixed transmission rate. This can be accomplished by exploiting any redundancy present in the image, together with use of an appropriate quantization strategy.

While there has been extensive research directed toward characterizing the rate-distortion performance of various image compression schemes, almost all of these studies have been based on use of the mean-squared error fidelity criterion. For example, previously reported results for entropy-constrained subband coding (ECSBC) have shown good quality image reconstructions, as well as excellent rate-distortion performance, using a minimum mean-squared error distortion criterion [1]. Minimum mean-square error (MMSE), however, is not the best measure of human psychophysical evaluation because it does not take into consideration the relative *visibility* of coding artifacts. It is becoming increasingly necessary to define distortion measures based on subjective evaluation criteria [2], which will allow minimization of the perceived distortion, rather than mean-square error, for a desired transmission rate. Determination of a perceptually based distortion metric has therefore been a subject of renewed interest in the image coding literature with early work described in [3].

In this paper we focus on the development and use of a perceptually relevant distortion measure for use in a subband coding environment which better mimics the subjective evaluation properties of the HVS than the squared-error metric. This distortion measure is then incorporated in a straightforward manner into previously developed design procedures for ECSBC schemes based upon training data, as described in [1] for the specific case of mean-square distortion.

The perceptual distortion measure is based upon the concept of a just-noticable-distortion (JND) level at a given pixel location in the reconstructed fullband image when errors occur in different subbands. This results in a spatially varying perceptual threshold $T_i(\mathbf{x})$ indicating the JND due to errors at pixel site $\mathbf{x}$ in the i'th subband. The evaluation of $T_i(\mathbf{x})$ is determined empirically similar to the procedure described in [4] and depends upon spatial frequency (subband), local texture and local contrast. However, unlike the coding approach in [4], where the perceptual threshold was used to simply set the stepsize of a uniform threshold scalar quantizer, in this work the perceptual threshold is used to describe a distortion measure which is then used in the design of ECSBC schemes. By making use of the appropriately adapted ECSBC design procedure reported in [1], a variety of scalar and vector quantization schemes can be investigated for encoding the subband components. This includes entropy-constrained vector quantization (ECVQ) [5] as well as entropy-constrained predictive vector quantization (ECPVQ) schemes [6]. Optimum bit allocation is provided as an integral part of this design approach.

A number of results are presented illustrating the superior subjective performance associated with the use of this perceptual distortion measure compared to the conventional squared-error distortion criterion. Suggestions for further extension of this approach are provided.

References

[1] Y.H. Kim and J.W. Modestino, "Adaptive Entropy Coded Subband Coding of Images," *IEEE Transactions on Image Processing*, Vol. 1, pp. 31-48, January 1992.

[2] N. Jayant, "Signal Compression: Technology Targets and Research Directions," *IEEE J. Select. Areas in Commun.*, Vol. JSAC-10, pp. 796-818, June 1992.

[3] D.J. Sakrison, "Image Coding Applications of Vision Models," in *Advances in Electronics and Electron Physics*, W. Pratt, Ed., Academic Press, New York, 1979.

[4] R.J. Safranek and J.D. Johnston, "A Perceptually Tuned Subband Image Coder With Image Dependent Quantization and Post Quantization Data Compression," *Proc. ICASSP'89*, pp. 1945-1948, 1989.

[5] P.A. Chou, T. Lookabough, R.M. Gray, "Entropy-Constrained Vector Quantiztion," *IEEE Trans. on Acoust., Speech and Signal Proc.*, Vol. ASSP-37, pp. 31-42., January 1989.

[6] Y.H. Kim and J.W. Modestino, "Adaptive Entropy Coded Predictive Vector Quantization," *IEEE Trans. on Sig. Proc.*, Vol. 40, pp. 633-644, March 1992.

Multivariate modeling of subband image statistics using spherically symmetric distributions

Frank Müller and Christoph Stiller

Institute for Communication Engineering, Aachen University of Technology (RWTH)
5100 Aachen, Germany, Phone: +49-241-807681, Fax: +49-241-807669

Introduction

Subband coding (SBC) is an attractive image coding scheme. For compression the subband signals must be quantized. Vector quantization (VQ) of the subband signals exploits the statistical bindings between the samples.

The performance of a particular vector quantizer is highly dependent on how well its codebook matches to the source statistics. Investigations concerning VQ performance require multivariate source models.

In this respect the SIRP models, which will be recalled in the following section, have many interesting properties. SIRP model sources can be efficiently quantized using lattice VQ which reduces implementation complexity drastically compared to VQ with unstructured codebooks. The known designs employ contour–gain separated VQ (similar to [1]) and a lattice structured codebook for quantization of the contour vector.

Besides, traditional VQ can benefit from SIRP models by training with pseudo random data generated according to the model [2].

SIRP models

A random process is called a *spherically invariant random process* (SIRP), iff every joint probability density function (pdf) in n variables $p_n(\mathbf{x})$ is a function of the quadratic form $\mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}$ only:

$$p_n(\mathbf{x}) = f_n(\mathbf{x}^T\mathbf{M}^{-1}\mathbf{x}), \qquad (1)$$

where M denotes the corresponding $n \times n$ covariance matrix. The function f_n describes the shape of the distribution.

This means that all contours of equal probability density are multidimensional ellipsoids. In particular, all contour lines of equal density of any bivariate distribution taken from two samples of a SIRP are ellipses.

Du [3] gave a comprehensive SIRP model for image signals which uses generalized gaussian functions to describe the univariate marginal distributions. He showed further that image blocks (after subtraction of the sample mean) can be modeled as realizations of a SIRP. It will be shown here that image blocks in the subband domain can be modeled as SIRPs as well.

Subband image statistics

Numerous still pictures (512 × 512 pel, 8 bit quantization) were filtered with separable quadrature mirror filters (QMF). The resulting images were divided into blocks of size 4 × 4. In the baseband the sample mean of the blocks was subtracted. These blocks were then transformed by a principal axes transformation H into uncorrelated vectors with unit variances:

$$\mathbf{y} = \mathbf{H}\mathbf{x} \quad \text{with} \quad \mathbf{H}^T\mathbf{H} = \hat{\mathbf{M}}^{-1}, \qquad (2)$$

where $\hat{\mathbf{M}}$ denotes an estimate of the covariance matrix.

If the hypothesis of spherical invariance (1) is true, the vectors in the principal domain obey the pdf:

$$p_n(\mathbf{y}) = \frac{1}{\det \mathbf{H}} f_n(\mathbf{y}^T\mathbf{y}). \qquad (3)$$

Thus the pdf of y depends only on the radius $r = \sum y_i^2$ of the vectors. This hypothesis has been tested with a χ^2-test of goodness of fit. The results have shown that the spherical symmetry is much stronger in the subbands than in the original domain. It turned even out that spherically symmetric distributions are better suited (by an order of magnitude) than those distributions related to the common model of statistical independence in the principal axes domain.

This gives rise to use lattice VQ in the subband domain, thus combining the advantages of subband coding with the performance of VQ, without the need for the storage of many different codebooks.

References

[1] K.T. Malone and T.R. Fischer, Contour–Gain Vector Quantization, *IEEE Trans. Acoust., Speech and Signal Proc.* ASSP-36, pp. 862-870, 1988

[2] Y. Du. SIRP-Model Based Generation of Image VQ Training Sequences, *13-th GRETSI Symposium on Signal and Image Processing, pp. 889–892*, Juan-les-pins, France, 1991.

[3] Y. Du. *Ein sphärisch invariantes Verbunddichtemodell zur Vektorquantisierung von Bildsignalen.* Ph. D. Dissertation, Aachen University of Technology, 1991.

Optimal Predictive Coding of 2 D Fields*

José M. F. Moura

Dep. Electr and Comp. Eng.
Carnegie Mellon University
Pittsburgh PA 15213-3890

Nikhil Balram

IBM Corporation
1000 N. W. 51st Str.
Boca Raton FL 33432

We discuss coding of 2D data using a recursive framework for *noncausal* Gauss Markov random fields (GMRF) defined on finite lattices. This framework exploits to advantage the structure of GMRFs providing the means to achieve recursive optimal processing, while preserving the *noncausality* of the field.

The compression scheme uses noncausal prediction coupled to vector quantization (VQ). The noncausal prediction fits first a noncausal GMRF to the data, then whitens the data by an inverse filtering type operation, and finally vector quantizes the prediction error field. In this paper, we explain the details of the noncausal prediction. Lack of space prevents us to discuss the parameter estimation algorithm that is needed to fit a 2D model to the data, see [1].

GMRF Recursive Structure

Important in the coding of GMRFs is the issue of parameterization. This leads to the question of when is a positive definite matrix the covariance of a GMRF? Partial answers are available only in very special cases. In general, for GMRFs on finite lattices, it is not possible to answer the question directly. It turns out that the right way to pose it is in terms of the inverse of the covariance matrix which we refer to as the potential matrix, see [2] for details.

Let $\{x_{i,j}\}$, $1 \leq i,j \leq N$, represent the 2D field on a finite lattice (taken as a square, for simplicity.) Woods [3]'s minimum mean square error representation of a homogeneous first order GMRF (nearest neighbors) is

$$x_{i,j} = \beta_h(x_{i,j-1} + x_{i,j+1}) + \beta_v(x_{i-1,j} + x_{i+1,j}) + e_{i,j}, \quad (1)$$

where β_h and β_v are the strengths of the neighbor horizontal and vertical field interactions, respectively. We call these the field potentials. Collecting all N^2 equations, taking care of boundary conditions (b.c.) (which here we assume Dirichlet zero boundary conditions, see [2] for general b.c.,) we get

$$A\mathbf{x} = \mathbf{e} \quad (2)$$

where the potentials are collected in the matrix $A = I \otimes B + H \otimes C$, and $\otimes$ is the Kronecker product. The N^2 vector $\mathbf{x} = \text{vec}[\mathbf{x}_i]$, where the N vectors $\mathbf{x}$ collect the intensities of the pixels of the ith - row. I is the N^2 identity matrix, $B = I_N - \beta_h H_N$ and $C = -\beta_v I_N$, H is an N^2 matrix of zero entries, except the upper and lower diagonal (all ones,) and I_N and H_N are like I and H but of dimension N.

The noise $\mathbf{e}$ has correlation $\Sigma_e = \sigma^2 A$. Apart the normalizing factor of σ^2, the covariance $\Sigma_{\mathbf{x}}$ of $\mathbf{x}$ is then the potential matrix A.

By Cholesky factorization, $A = U^T U$. Equation2) gives

$$U\mathbf{x} = \mathbf{w} \quad (3)$$

where the covariance of $\mathbf{w}$ is $\sigma^2 I$. The Cholesky factor U is not a full matrix. It is block diagonal with band $N+1$. The diagonal and the upper diagonal blocks of U are obtained from the iterates of a Riccati type equation. In [2], the convergence behavior of this iterative scheme is studied. For practical purposes, one may stop it after less than 10 iterations, considerably reducing the associated computational effort.

2D Coding

To code 2D data, we need the field parameter values $\beta_h, \beta_v, \sigma^2$. In [1], we analyze the parameter space of GMRFs and study their maximum likelihood (ML) estimation.

We have used this to code two dimensional data. The basic structure of the (lossy) codec is the following: (i) The global mean is subtracted from the 2D data, which is then input to an ML - estimator; (ii) a Cholesky factorization of A leads to the unilateral representation of the field; (iii) the field is whitened leading to the error field; (iv) the error field is vector quantized; (v) lossless entropy type coding can be used to achieve further compression. When applied to image data, we have verified that we can get over a factor of 3 - 10 of more compression ratio than DCT based techniques. This procedure and modifications to it are presently under study.

References

[1] Nikhil Balram and José M. F. Moura. Noncausal Gauss Markov random fields: Parameter structure and estimation. Technical report, LASIP, Department of Electrical and Computer Engineering, Carnegie Mellon University, April 1991. Accepted for publication after minor revisions, 45 pages, revised February 1992.

[2] José M. F. Moura and Nikhil Balram. Recursive structure of noncausal Gauss Markov random fields. **IEEE Transactions on Information Theory**, IT-38(2):334-354, March 1992.

[3] J. W. Woods. Two-dimensional discrete Markovian fields. *IEEE Trans. Inform. Theory*, IT-18:232-240, 1972.

*Work partially supported by ONR grant # N00014-91-J-1001

An Optimally Bit Allocated Wavelet Pyramid Image Coding System

Jie Chen and Shuichi Itoh

University of Electro-Communications, Chofu, Tokyo 182, Japan

Abstract

Reconstruction error properties for a wavelet pyramid image coding system are described. It is shown that when optimal bit allocation scheme is adopted, the reconstruction noises and the quantization noises of the wavelet pyramid coding system become regular, and the reconstruction noises can be approximated to stationary white noises. Based on the error property analysis, an optimal bit allocation scheme with respect to the minimum reconstruction-mean-square-error (RMSE) criterion is given. The system reconstruction distortion at a given bit rate $\overline{R}$ is proved to be directly proportional to $2^{-2\overline{R}}$. Experimental results are given.

Summary

For a J-stage discrete orthonormal wavelet pyramid image coding system [2], let $\{P_J, (D_j^1)_{1\le j\le J}, (D_j^2)_{1\le j\le J}, (D_j^3)_{1\le j\le J}\}$ be the wavelet decompositions of an input image P_0 and let $\{\epsilon_J, (\epsilon_j^1)_{1\le j\le J}, (\epsilon_j^2)_{1\le j\le J}, (\epsilon_j^3)_{1\le j\le J}\}$ be their quantization MSE's, respectively. Furthermore, let ϵ_j denote the reconstruction MSE at j-th layer of the pyramid. Based on the orthonormality of the discrete orthonormal wavelets and signal processing theory, the reconstruction MSE at a layer is given as

$$\epsilon_{j-1} = \epsilon_J + \sum_{i=j}^{J}\sum_{k=1}^{3}\epsilon_i^k, \quad 1 \le j \le J. \tag{1}$$

Theorem 1 *In the wavelet pyramid image coding system, if the quantizers which minimize the system reconstruction MSE at a given bit rate are employed, then the quantization MSE's and the reconstruction MSE's at every layer of the wavelet pyramid satisfy the following equations:*

$$\epsilon_j^k = \epsilon_j \tag{2}$$

$$\epsilon_{j-1}^k = 4\epsilon_j^k \tag{3}$$

$$\epsilon_{j-1} = 4\epsilon_j = 4^{J-j+1}\epsilon_J, \tag{4}$$

where $1 \le k \le 3$, $1 \le j \le J$. Furthermore, if the quantization noises are cross-uncorrelated and white, then the reconstruction noise at each layer is also white.

Theorem 1 indicates a kind of regularity about the quantization noises and the reconstruction noises. The regularity is reflected at least in three aspects: 1) The quantization MSE's at the same layer of the wavelet pyramid are equal to each other, and also equal to the reconstruction MSE at the same layer. 2) The quantization MSE's or the reconstruction MSE's at two successive layers are related by a factor 4 in quantity. 3) If the quantization noises are cross-uncorrelated and white, then the reconstruction noises at all layers will be white. This regularity should be useful for the practical applications such as post-processing of the reconstructed image, the progressive transmission and so on. Theorem 1 has also simplified the estimation of the reconstruction MSE's or the quantization MSE's, since only one quantization MSE is needed to know.

For $1 \le k \le 3$, $1 \le j \le J$, if we assign R_j^k bits to each component of sub-images D_j^k and R_J to P_J, then the optimal bit allocation problem can be formulated as

$$\textbf{minimize } \epsilon_0 = K_J 2^{-2R_J} + \sum_{j=1}^{J}\sum_{k=1}^{3} K_j^k 2^{-2R_j^k} \tag{5}$$

$$\textbf{subject to } \overline{R} = \frac{1}{4^J}R_J + \sum_{j=1}^{J}\frac{1}{4^j}\sum_{k=1}^{3}R_j^k, \tag{6}$$

where K_j^k (or K_J) denote quantization factors. The optimal solutions are obtained by using Lagrange multipliers:

$$R_j^k = \overline{R} + j - \frac{4}{3}(1-\frac{1}{4^J}) + \frac{1}{2}\log_2\frac{K_j^k}{K} \tag{7}$$

$$R_J = \overline{R} + J - \frac{4}{3}(1-\frac{1}{4^J}) + \frac{1}{2}\log_2\frac{K_J}{K}, \tag{8}$$

where K is given by

$$K = (K_J)^{\frac{1}{4^J}}\prod_{j=1}^{J}\prod_{k=1}^{3}(K_j^k)^{\frac{1}{4^j}}. \tag{9}$$

Furthermore, the minimum MSE of the system reconstruction is given as

$$\epsilon(\overline{R}) = \min\epsilon_0 = 2^{\frac{4}{3}(1-\frac{1}{4^J})}K 2^{-2\overline{R}}, \tag{10}$$

which is directly proportional to $2^{-2\overline{R}}$.

A wavelet pyramid image coding system composed of 10-tap W-QMF's and an optimally bit allocated uniform quantizer was implemented and the experimental results are shown in Fig. 1.

References

[1] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets" Communications on Pure and Applied Mathematics, Vol.41, No.7, pp. 909–996, 1988.

[2] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. on PAMI, Vol.11, No.7, pp. 674–693, July 1989.

[3] P. H. Westerink, Boekee D. E., Biemond JJ. and J. W. Woods, "Subband Coding of Image using Vector Quantization," IEEE Trans. on Commun., Vol.36, No.6, pp. 713–719, 1988.

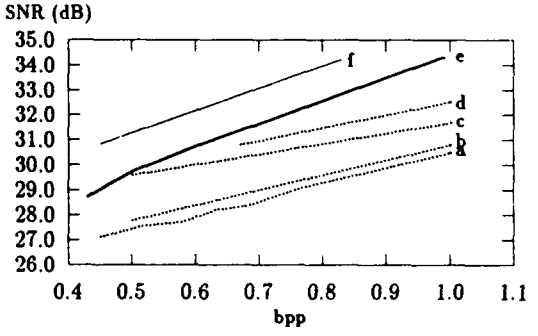


Fig. 1 SNR versus bit-rates for 256 by 256 pixel image "Lena": (a) SBC+SQ; (b) adaptive DCT; (c) SBC+spatial differential VQ; (d) SBC+adaptive DPCM; (e) Wavelet+SQ+Huffman; (f) entropy of Wavelet+SQ, where dashed lines are taken from [3].

ROTATIONALLY INVARIANT TRELLIS CODES FOR QAM

by

Eric J. Rossin and Chris Heegard
School of Electrical Engineering, Cornell University, Ithaca, NY 14853

SUMMARY

Rotationally invariant (RI) trellis codes are important whenever the modulation signal set has a rotational symmetry and the transmission system can introduce a phase rotation. Rotational invariance means that a trellis code is closed under rotation of the individual elements of the signal set onto which it is mapped. In this paper we look at this "rotation" as an "isometry sequence" under which the code is invariant. We concentrate on trellis codes that can be described as the orbit of a group of isometry sequences acting on the cosets of a lattice partition in Euclidean space.

An *isometry*, T, of a trellis code, $\mathcal{C}$, is a map such that if c_1 and c_2 are codewords in $\mathcal{C}$, $\|c_1 - c_2\| = \|T(c_1) - T(c_2)\|$, $\forall c_1, c_2 \in \mathcal{C}$. The distance function is the Euclidean distance on the individual components of each codeword, given by the one-to-one map between codeword labels and the modulation signal set. A *coordinate isometry* is an isometry such that $\forall\, i$, $\|[T(c_1) - T(c_2)]_i\| = \|[c_1 - c_2]_i\|$. Note that the shift operator is an isometry that is not a coordinate isometry. A *symbolic dynamic group*, S, is a subshift [1] over a finite group, that itself forms a group in sequence space by applying the group operation coordinate-wise. The group S is guaranteed to be a subshift of finite type, conjugate to a full shift, which in part implies that it is generated by a deterministic, labeled directed graph that admits a sliding window inverse [2].

To describe a geometrically uniform trellis code $\mathcal{C}$ [4], begin with the set of isometries, U, that map cosets of a particular constellation partition onto themself. This set forms a finite (non-abelian) group under composition. Describe a symbolic dynamic group, S, over U; S is generated by a graph that is referred to as an isometry graph of the code. The trellis code, $\mathcal{C}$, is then the orbit of an initial sequence c_0 (a sequence of points in Euclidean space), under the action of S. The code $\mathcal{C}$ can then be viewed as a generalization of a Slepian group code [3, 4]. Note that if c_0 is a constant sequence, the encoder can be obtained by taking the action of each edge label of the graph of S on the "point" c_0. However, the resulting graph will in general not be minimal, (i.e., the graph that generates $\mathcal{C}$ may be smaller than the graph that generates S). Rotational invariance now corresponds to the orbit of a symbolic group S that includes the "all rotations" sequence.

We demonstrate these ideas by concentrating on maps of the form $Ac + b$, where $c \in \mathcal{C}$, $A : \mathcal{C} \to \mathcal{C}$ is an invertible matrix and $b \in \mathcal{Z}_2 \times \mathcal{Z}_4$. These maps, which operate on the labels of the QAM constellation, form a group of 32 elements, U. Each map in U induces an isometry on the QAM signal set (the standard 8-way partition of $\mathcal{Z}^2$ under an isometric labeling [4, 5]). The code $\mathcal{C}$ is the orbit of an isometry graph over U. This implies that the graph of the encoder is embedded in the isometry graph, since all codewords can be generated as the action of the isometries on the "all zero's" sequence. In other words, if the isometry graph is only labeled with the b's (i.e., let A be the identity map), the resulting graph must be reducible to the graph of the encoder for $\mathcal{C}$. For example, consider the following class of rotationally invariant trellis codes [5]: let the generator G be the rate 1/2 convolutional code over $\mathcal{Z}_4$ (the integers modulo 4) given by $[1 - D, G_p(D)]$. The input sequence is also over $\mathcal{Z}_4$, but the output is taken as $(m,p) \in \mathcal{Z}_2 \times \mathcal{Z}_4$, where m is the most significant bit generated by the $1 - D$ term, and p is the output from $G_p(D)$. The outputs of the code are then mapped to the 8 cosets. Two specific examples will be presented: a 4-state code with generator $[1 - D, 2 - D]$, that has a free Euclidean distance of 3 and an 8-state isometry graph, and an 8-state code with generator $[1 - D, 2 + D + 2D^2]$, that has a free Euclidean distance of 5 and also an 8-state isometry graph. The later code is equivalent to the 8-state RI trellis code used in the CCITT V.32 standard [6], while the first demonstrates an example where the isometry graph is larger than the graph of the encoder.

REFERENCES

[1] B. Marcus, "Sofic systems and encoding data," *IEEE Transactions on Information Theory*, vol. IT-31, no. 3, pp. 366–377, May 1985.

[2] B. Kitchens, "Expansive dynamics on zero-dimensional groups," *Ergodic Theory and Dynamical Systems*, vol. 7, pp. 249–261, 1987.

[3] D. Slepian, "Group codes for the Gaussian channel," *Bell System Technical Journal*, vol. 47, pp. 575–602, Apr. 1968.

[4] G. D. Forney, jr., "Geometrically uniform codes," *IEEE Transactions on Information Theory*, vol. 37, pp. 1241–1260, Sep. 1991.

[5] E. Rossin and C. Heegard, "Rotationally invariant codes with a linear structure", in *Proceedings 26th Conference on Information Sciences and Systems*, Princeton University, Mar. 18–20, 1992

[6] L. F. Wei, " Rotationally Invariant Convolutional Channel Coding with Expanded Signal Space - Parts I and II," *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, pp. 659–686, Sep. 1984.

[This work was supported in part by NSF grant NCR-8903931 and NCR-9207331.]

On 90° Rotationally Invariant Lattice Codes

J. A. Sheppard and A. G. Burr[1]

Introduction

A key step in the design of a lattice code is the design of an n-dimensional constellation — having selected a suitable lattice, the designer must choose a finite subset of points, perhaps with some translation or rotation, to define a set of code words with the required properties.

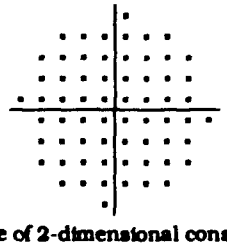
It is well known that the SNR performance of and n-dimensional code is partly determined by the shape of the region of n-space occupied by the code. The maximum *shape gain* is given by an n-sphere, but this is difficult to implement and leads to a constituent 2-dimensional constellation with a high peak to average power ratio [5]. Various other techniques have been suggested, including Voronoi Constellations [1, 2] and Shell Constructions [3], but none of these works address the problem of rotational invariance.

In a practical system, the receiver must recover the correct phase of the constellation. If phase symmetries exist then the receiver may lock to the wrong phase. Thus the code must either have no phase symmetries, or be immune to any rotations resulting from such symmetries. The latter type, known as Rotationally Invariant codes, are attractive since phase symmetries can make phase recovery easier and more reliable.

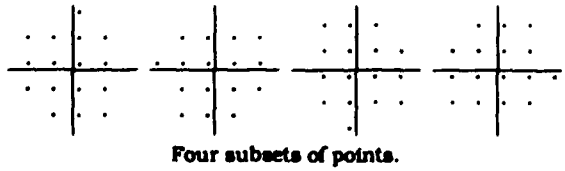
This paper outlines techniques for designing rotationally invariant codes using Leech and Sloane's constructions A and B [4]. These lattices are either optimally dense or offer a good performance/complexity trade-off in up to 32 dimensions, and their simplicity relative to other lattice constructions makes them worthy of considerations in higher dimensions.

Construction A Lattices

Construction A forms a lattice from the union of cosets of the lattice of even integers, in which the coset leaders are the words of a linear binary code. If this is offset by the vector $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$ then the points lie on the half integer grid, $\mathbb{Z}^n + (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$. Thus a two dimensional constituent constellation consists of four subsets of points: $2\mathbb{Z}^n + (\frac{1}{2}, \frac{1}{2})$, $2\mathbb{Z}^n + (\frac{1}{2}, \frac{1}{2}) + (0, 1)$, $2\mathbb{Z}^n + (\frac{1}{2}, \frac{1}{2}) + (1, 0)$, and $2\mathbb{Z}^n + (\frac{1}{2}, \frac{1}{2}) + (1, 1)$.This is illustrated in figure 1.



Example of 2-dimensional constellation



Four subsets of points.

Figure 1 — Decomposition of 2-D Constellation.

In the encoder, some of the data bits generate a code word from the constituent linear binary code, which is used to define the sequence of subsets to be transmitted. The remaining bits are mapped directly onto points within the required subsets.

Rotational invariance is achieved by regarding the subsets as rotations of $2\mathbb{Z}^n + (\frac{1}{2}, \frac{1}{2})$, rather than translations. Data bits mapped onto point numbers will clearly be unaffected by 90° rotations, and only the bits used to determine the subset sequence need special attention.

It can be shown that if the constituent binary code contains the words (1, 1,...) and (0, 1, 0, 1,...) then a 90° rotation of all the symbols will result in another valid code word. It is also possible to list the 2^k words of the binary code so that a 90° rotation corresponds to a cyclic shift of $2^{(k-2)}$ places down the list. Thus the data bits can be used to specify the position in the list of new binary code word relative to the previous word, and the lattice code becomes rotationally invariant.

Construction B Lattices

In a Construction B lattice code, the two dimensional constellation is divided into sixteen subset, and two binary codes are used to define the sequence of subsets. Rotational invariance can be achieved with two sets of word-level differential encoding, using similar techniques to those outlined above. However, space does not allow a detailed explanation here.

Conclusions

Techniques for designing 90° rotationally invariant construction A and B lattice codes have been outlined above. These have potential applications both in the design of lattice codecs, and in multidimensional trellis codes.

References

[1] J. Conway and N. J. A. Sloane. A fast encoding method for lattice codes and quantizers. *IEEE Transactions on Information Theory*, 28:820–824, 1983.

[2] G. D. Forney, Jr. Multidimensional constellations—partII: Voronoi constellations. *IEEE Journal on Selected Areas in Communications*, 7(6):942–958, Aug. 1989.

[3] P. Fortier, A. Ruiz, and J. M. Cioffi. Multidimensional signal sets through the shell construction for parallel channels. *IEEE Transactions on Communications*, 40(3):500–512, Mar. 1992.

[4] J. Leech and N. J. A. Sloane. Sphere packings and error correcting codes. *Canadian Journal of Mathematics*, 23:718–745, 1971.

[5] L.-F. Wei and G. D. Forney, Jr. Multidimensional constellations—partI. Introduction, figures of merit, and generalised cross constellations. *IEEE Journal on Selected Areas in Communications*, 7(6):877–892, Aug. 1989.

[1]Communications Research Group, Department of Electronics, University of York, Heslington, York, YO1 5DD, England. Telephone: +44 904 432396. Fax: +44 904 432335. e-mail: jas@ohm.york.ac.uk

A Semi-algebraic Construction to Achieve Rotationally Invariant Coded QAM on the Basis of Multilevel Convolutional Codes

Werner Henkel and Michael Koch

Deutsche Bundespost Telekom, Research Centre
PO Box 10 00 03, D-6100 Darmstadt, Germany
M. Koch is now with Siemens AG (ÖV), D-8000 München

1 Introduction

In order to account for carrier phase instabilities especially on satellite or mobile links, several proposals have been made to define rotationally invariant coded modulation. They were based on multidimensional or nonlinear convolutional codes, on separate encoding of the I- and Q-coordinates, or on multilevel block codes, especially with Reed-Muller codes as component codes.

This contribution describes a semi-algebraic approach with multilevel convolutional codes that leads to schemes with considerably low complexity. The construction guarantees 90°-invariance of the code, not yet of the information symbols itself. Hereto, a special differential en/decoder structure has been developed.

2 Conditions for the binary convolutional component codes

Assuming a binary set partitioning of the 2^m-QAM, with a labelling that is chosen to be 90°-invariant from the third partition label on, one obtains the following conditions:

I The all-ones sequence must be a valid code sequence of code (1).
$(\ldots,1,1,\ldots,1,\ldots) \in \mathcal{A}^{(1)}$
II All valid code sequences of code (1) must be valid code sequences of code (2), too. $\mathcal{A}^{(1)} \subset \mathcal{A}^{(2)}$
III No conditions for $\mathcal{A}^{(j)}$, $j = 3, \ldots$

3 Differential en- and decoding

The modulo-4 differential decoder is located *after* the multistage convolutional decoder. Otherwise the noise power would be doubled at the input of the differential decoder, significantly reducing the achievable coding gain.

The modulo-4 differential encoder is located *between* the encoding stages one and two (see Fig.). It can be shown that this demands for a systematic second-level code.

4 The semi-algebraic construction

As outlined in section 2, the all-ones sequence has to be a valid code sequence of code (1). For $k^{(1)} = 1$ a code with all generators having an odd weight obviously fulfills this condition.[1] For $k^{(1)} > 1$, the all-ones code sequence can be obtained, if there is the possibility of creating odd weighted generators by combining some rows (by means of the information sequence) of the Forney matrix of code (1). This, e.g., is fulfilled, if one row consists only of odd weighted generator polynomials or if the whole code is only composed of odd weighted generators.

To ensure rotational invariance for code (2), as a necessary and sufficient condition, one has to ensure that every valid code sequence $A^{(1)}$ of code (1) is also belonging to the set of code sequences $A^{(2)}$ of code (2).

$$\forall_{I^{(1)}} \exists_{I^{(2)}} : A^{(2)} = I^{(2)} \cdot G^{(2)} = I^{(1)} \cdot G^{(1)} = A^{(1)} .$$

($I^{(j)}$: Info series, $G^{(j)}$: Forney generator matrix)
As this equation has to be fulfilled for arbitrary $I^{(1)}$, $I^{(2)}$ appears as a function of $I^{(1)}$. A possible approach for the construction of code (2) is to define the components of $I^{(2)} = (I_1^{(2)}, I_2^{(2)}, \ldots, I_{k^{(2)}}^{(2)})$ as shifted versions of $I^{(1)}$ (assuming $k^{(1)} = 1$):

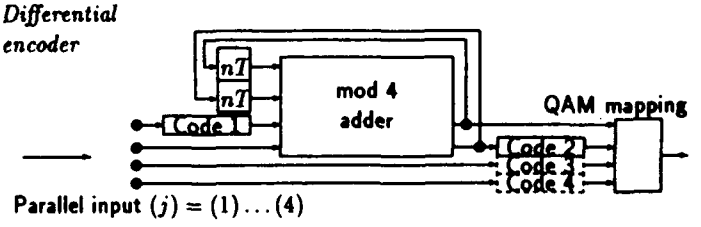
$$I_h^{(2)} = I^{(1)} \cdot D^{j_h} \quad (k^{(1)} = 1, h = 1, \ldots, k^{(2)}, j_h \in \{0, \ldots, L^{(1)} - 1\}) .$$

$L^{(j)}$ is the constraint length of code (j) (not multiplied with $k^{(j)}$). D is a time delay factor (z^{-1} of the Z-transform).

There has to be at least one $I_h^{(2)} = I^{(1)}$, i.e. $j_h = 0$, in order to express the low-order term $D^0 = 1$, appearing in $G^{(1)}$, by means of $G^{(2)}$. Furthermore, one j_h has to equal $j_h = L^{(1)} - L^{(2)}$. This is necessary as the term $D^{L^{(1)}-1}$ appearing in $G^{(1)}$ has to be expressed by $G^{(2)}$ with the maximum exponent $L^{(2)} - 1$. For reasons of decoding complexity it is useful to have $L^{(2)} \leq L^{(1)}$, because $k^{(2)}$ is usually greater than $k^{(1)}$. This can be achieved by the proposed construction leading to a considerably low decoding complexity.

Some results are given subsequently. A coding scheme with an asymptotic coding gain of 6 dB, e.g., has a complexity of 4 states for the first stage and 8 states for the second (and, maybe, additionally the Wagner decoding of a parity-check code as a third stage).

	Code (1)	Code (2)
Gen. non-rec.	(4,7,7)	(10,13,15), (16,13,15)
$R^{(j)}, L^{(j)}, d_f^{(j)}$	$\frac{1}{3}$, 3, 6	$\frac{2}{3}$, 2, 3
Gain / dB	4.7	
Gen. non-rec.	(15,15,13)	(51,61,73)
$R^{(j)}, L^{(j)}, d_f^{(j)}$	$\frac{1}{3}$, 4, 9	$\frac{2}{3}$, 3, 5
Gain / dB	6.5	
Gen. non-rec.	(1,2,7,7)	(46,52,61,73)
$R^{(j)}, L^{(j)}, d_f^{(j)}$	$\frac{1}{4}$, 3, 8	$\frac{3}{4}$, 2, 4
Gain / dB	6	



[1] $k^{(j)}$: number of info bits per frame, coderate $R^{(j)} = \frac{k^{(j)}}{n}$

Rotationally Invariant Multilevel Codes[1]

J. N. Livingston
Texas A&M University, College Station, TX 77843-3128

I. Introduction

The idea behind rotationally invariant codes is to find an encoder that ensures the following: given any coded sequence, if we rotate each symbol through a fixed rotational symmetry, then the new sequence of rotated symbols is also a valid code sequence. If this is the case, we may use differential encoding and decoding to overcome the effects of phase rotation.

In this work, we describe an approach to the design of rotationally invariant codes using multilevel coding. This technique allows the designer to achieve, *a priori*, a given performance level, as well as being invariant to rotations through constellation symmetries.

II. Rotational Invariance and Multilevel Codes

It has been argued (see e.g., [1]) that a "natural labeling" is best for achieving rotational invariance. In Figure 1, we illustrate natural labeling on a 16-QAM constellation. Below each point is a binary label. Note that for the 16-QAM constellation, the two least significant bits are not rotationally invariant, while the two most significant bits are already invariant to rotations through multiples of $\pi/2$ radians. This leads us to the following observation:

Observation 1: *Only the two least significant bits (for QAM) need to be encoded in such a way as to make them rotationally invariant*

The theory behind multilevel codes involves partitioning the signal space into subsets. The multilevel code employs an L-level code, $C = [C_1, \cdots, C_L]$, where the C_i are component codes. Each component code is responsible for selection of its corresponding subset, a_i, i.e., C is the set of all sequences $\{(a_1^k, \cdots, a_L^k)\}$ of subsets in the constellation that satisfy $(a_i^k) \in C_i$ for all $i = 1, 2, \ldots, L$ [2]

We note that in Figure 1, that the two LSB's are affected by a rotation. This again yields an observation:

Observation 2: *Both code C_1 and C_2 must provide rotational invariance for a 16-QAM constellation.*

It can be shown that for multilevel codes, the condition of rotational invariance is given by the following.

90 Degree Rotational Invariance: 90 degree rotational invariance is guaranteed if the codes C_1 and C_2 (assumed linear) meet the following criteria:

1. Code C_1 must contain the all one's sequence.
2. Code C_1 must be a *subcode* of code C_2.

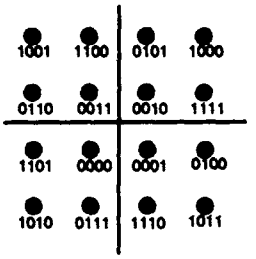


Fig. 1. 16 QAM constellation.

These conditions are the same as those found in [3], but couched in different terms. In [3], this method is extended to M-PSK signaling. However, the conditions for rotational invariance become harder to meet for M greater than 4. The importance of these observations is that now we have a constructive method of finding rotationally invariant codes. It is a simple matter to find codes that achieve coding gains of 2 to 5 dB with little complexity. For example, binary BCH codes of the same length meet these conditions. The design of convolutional codes that meet these conditions can be difficult. However, through the use of generator matrix descriptions of the codes, we have been able to find a systematic method for designing them. One approach is to extend the convolutional code over two time epochs and then prune paths. A second approach is to delete generator polynomials from a high rate code to form the subcode. And a third approach is to form new generator polynomials from a high rate code by multiplying them with what we call sequence limiting polynomials. All three approaches involve some trial and error in finding the codes with the best distance, and ensuring the all one's sequence remains in the subcode.

References

[1] S. S. Pietrobon, G. Ungerboeck, and D. J. Costello, Jr., "Rotationally Invariant Nonlinear Trellis Codes for Two-Dimensional Modulation," submitted to *IEEE Trans. on Inform. Theory*, 1991.

[2] A. R. Calderbank, "Multilevel Codes and Multistage Decoding," *IEEE Transactions on Communications*, vol. 37, no. 5, pp. 222-229, March 1989.

[3] Kasami, T., *et. al.*, "On linear structure and phase rotation invariant properties of block M-PSK modulation codes," *IEEE Trans. on Inform. Theory*, vol. 37, no. 1, pp. 164-167, Jan. 1991.

[1]This work was supported in part by NSF grant number NCR-9016354

EIGHT-DIMENSIONAL MODULATION FOR BANDLIMITED CHANNELS

Spase L. Drakul (*) and Ezio Biglieri (†)

(*) University of Ljubljana • Faculty of Electrical and Computer Engineering • 61000 Ljubljana (Slovenia)
(†) Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy).

N-dimensional (N-D) signals are generated by selecting N an orthogonal basis $\psi_1(t), \psi_2(t), ..., \psi_N(t)$. The data vector $\mathbf{x} = (x_1, x_2, ..., x_N)$ is carried over the channel by the signal $s(t) = \sum_{j=1}^{N} x_j \psi_j(t)$. Here we assume that the symbols x_i take on values ± 1. At the receiver, the data vector is recovered by exploiting the orthogonality of the basis functions: a bank of N filters, each matched to one $\psi_j(t)$, gives $\mathbf{x}$ at its output.

In this paper we describe a technique to generate an 8-D basis for transmission over bandlimited channels. The idea here is the following. Assume we have a set of N orthogonal signals $\Phi = \{\phi_1(t), \ldots, \phi_N(t)\}$. We generate a $2N$-dimensional basis by taking the products $\xi_1(t)\Phi$ and $\xi_2(t)\Phi$, with $\xi_1(t)$ and $\xi_2(t)$ chosen properly.

As an application of this procedure, we get Q^{2}PSK [1, 2] by choosing $\Phi = \{p(t), q(t)\}$ and $\xi_1(t) = \sin\omega_c t$, $\xi_2(t) = \cos\omega_c t$. Similarly, a four-dimensional set of signals defined over the interval $(-2T_b, 6T_b)$, T_b the bit duration, can be generated by choosing

$$\phi_1(t) = \frac{1}{\sqrt{2T_b}} \cos\left(\frac{\pi t}{4T_b}\right) \Pi\left(\frac{t}{4T_b}\right) \qquad \phi_2(t) = \phi_1(t - 4T_b) \tag{1}$$

$$\xi_1(t) = \frac{1}{2\sqrt{T_b}} \sin\left(\frac{\pi t}{8T_b}\right) \qquad \xi_2(t) = \frac{1}{2\sqrt{T_b}} \cos\left(\frac{\pi t}{8T_b}\right) \tag{2}$$

where $\Pi(t/4T_b) = 1$ if $|t| < 2T_b$, and $= 0$ otherwise. Inspection of (1) shows that $\phi_1(t)$ and $\phi_2(t)$ are orthonormal. An eight-dimensional signal basis can now be obtained by taking the products $\{a_{ij}(t)\cos\omega_c t, a_{ij}(t)\sin\omega_c t\}$, where

$$a_{ij}(t) = A_{ij}\, \phi_i(t)\xi_j(t), \qquad i, j = 1, 2,$$

and the constants A_{ij} may take on any non-zero value. Here A_{ij} are chosen so as to have $\| a_{11}(t) \| = \| a_{22}(t) \| = 0.5817$ and $\| a_{12}(t) \| = \| a_{21}(t) \| = 0.8134$. The cumulative power spectral density is shown in Fig. 1. The spectrum of 8D-4P2C is more compact than that of QPSK and of Q^{2}PSK.

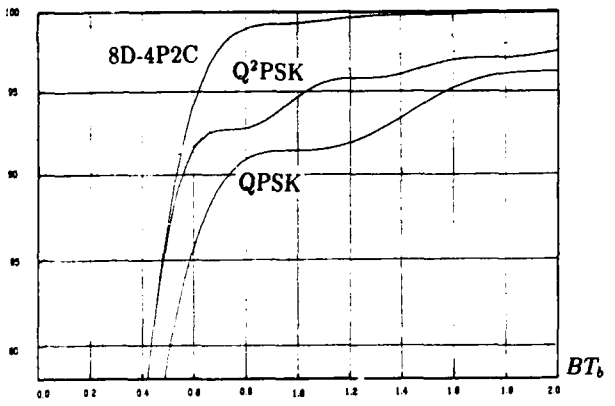


Figure 1: Cumulative power spectral density of QPSK, Q^{2}PSK, and 8D-4P2C: Percentage of signal power in the bandwidth BT_b.

We can transmit the 8-D vector $\mathbf{x} = (x_1, \ldots, x_8)$ through the signal

$$x(t) = s_c(t)\cos\omega_c t + s_s(t)\sin\omega_c t \tag{3}$$

where

$$s_c(t) = x_1 a_{11}(t) + x_2 a_{12}(t) + x_5 a_{21}(t) + x_6 a_{22}(t)$$

and

$$s_s(t) = x_3 a_{11}(t) + x_4 a_{12}(t) + x_7 a_{21}(t) + x_8 a_{22}(t).$$

This modulation scheme can be interpreted from a different point of view by defining the two waveforms $\beta_1(t) = a_{11}(t) + a_{12}(t)$ and $\beta_2(t) = -a_{11}(t) + a_{12}(t)$, where $-2T_b \le t < 2T_b$. Moreover, observe that we have $a_{22}(t) = -a_{11}(t - 4T_b)$ and $a_{21}(t) = a_{12}(t - 4T_b)$. Consequently, we can write for example:

$$a_{11}(t) + a_{22}(t) + a_{21}(t) - a_{22}(t) = \beta_1(t) + \beta_1(t - 4T_b),$$

and similar relations hold for all the possible values of the 4-tuple (x_1, x_2, x_3, x_4). Thus, we can represent our modulation scheme by writing the transmitted signal in the form

$$s_c(t) = \pm\beta_i(t) \pm \beta_j(t - 4T_b), \tag{4}$$

$$s_s(t) = \pm\beta_k(t) \pm \beta_\ell(t - 4T_b), \tag{5}$$

where $i, j, k, \ell = 1, 2$. $\beta_1(t)$ and $\beta_2(t)$ are *orthonormal* pulses.

Optimum demodulation of signal (3) transmitted over the AWGN channel can be done in the standard way by exploiting the orthogonality of signals. Here we examine a suboptimum demodulator, which exploits the special structure of our basis signals. To simplify our presentation, consider only the in-phase branch of the demodulator, and assume that noise is not present. The demodulator outputs the signal

$$\hat{s}_c(t) = \pm\beta_i(t) \pm \beta_j(t - 4T_b). \tag{6}$$

We observe that the first term in the right-hand side of (6) is non-zero in the first half of the $8T_b$ symbol interval, while the other term is non-zero in the second half.

The demodulator structure can be considerably simplified by avoiding multiplication of the received signal by $a_{ij}(t)$. Upon observation of signal (6) we form the new signal

$$\pm\beta_i(t) \mp \beta_j(t - 4T_b) \tag{7}$$

obtained by inverting the signal polarity in the second half of the observation interval. By taking the sum $\sigma(t)$ and the difference $\delta(t)$ of the observed signal and the signal with the polarity reversed, we obtain $\sigma(t) = \pm 2\beta_i(t)$ and $\delta(t) = \pm 2\beta_j(t - 4T_b)$. Thus, the detection problem is reduced to discriminating between the two pulses $\beta_1(t)$, $\beta_2(t)$, and the polarities of these pulses. This can be done by sampling twice in each subinterval of duration $4T_b$, each time comparing the sample value with a suitable threshold.

Error probability was simulated for transmission over a channel affected by additive white Gaussian noise and intersymbol interference and with a suboptimum detection strategy. Intersymbol interference was modeled by introducing a transmitting filter $H_T(f)$ and an equalizing filter $H_E(f)$ followed by a receiving filter $H_R(f)$ at the receiver's front end. $H_T(f)$ was selected so as to achieve the requirements of FCC standards. The equivalent noise bandwidth of the overall filter is $B_{eq} = 0.333/T_b$. The model also includes the effect of a carrier recovery circuit and of a symbol timing recovery unit. Fig. 2 shows that the degradation due to intersymbol interference is on the order of 1 dB for a bit error probability 10^{-6}. The loss in performance with respect to orthogonal eight-dimensional modulation over the AWGN channel (and with perfect carrier and timing recovery) is around 2 dB for the same error probability, but the spectrum of 8D-4P2C is more compact.

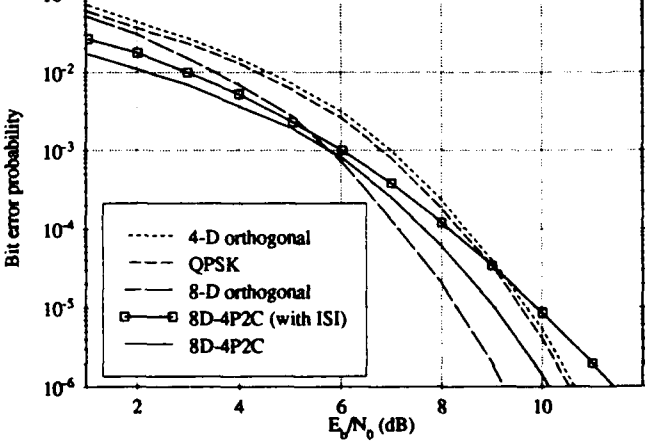


Figure 2: Bit error probability of 8D-4P2C for additive white Gaussian noise and intersymbol-interference channel ($B_{eq} = 0.33/T_b$) vs. E_b/N_0. Performance of orthogonal 4- and 8-dimensional schemes and of QPSK is also shown for comparison.

References

[1] D. Saha, and T.G. Birdsall, "Quadrature-Quadrature Phase Shift Keying," *IEEE Trans. on Commun.*, Vol. 37, No. 5, pp.437-448, May 1989.

[2] M. Visintin, E. Biglieri, and V. Castellani, "Four-dimensional signaling over bandlimited channels," *Proc. of the 1991 IEEE International Symposium on Information Theory*, Budapest, Hungary, p. 3, June 24-28, 1991.

Efficient splitting of multidimensional alphabets for modulation codes

Rolf Johannesson, Joakim Persson, and Kamil Sh. Zigangirov

Department of Information Theory
University of Lund
Box 118
S-221 00 Lund, Sweden

Summary—We propose a new combined coded modulation construction which gives a reduced decoding complexity. It is a generalization of the constructions of Ginzburg [1] and Ungerboeck [2] and is based on splitting a multidimensional alphabet with $2^k, k \geq 2$, symbols into k binary alphabets. The encoder consists of a set of k binary convolutional encoders, *elementary encoders*, operating at code rates $R_1, R_2, \ldots, R_k$, where $R_1 < R_2 < \ldots < R_k$. The data bits are split into k streams, each encoded by one of the elementary encoders. The set of k elementary encoder outputs is mapped onto the set of 2^k-ary modulator symbols. The decoder consists of k *elementary decoders*. The decoding is performed step by step beginning with the first elementary decoder, then the second etc. Each elementary decoder uses information from the outputs of the previous decoders. The code rate and memory of each elementary encoder is chosen such that the elementary decoders have approximately the same complexity and reliability.

We describe this method in more detail for the Gaussian channel and 4-PSK with soft decisions. Our rate R encoder consists of two (elementary) parallel binary rates R_1 and R_2 convolutional encoders, where $R = R_1 + R_2$. Each encoder generates one binary code symbol per time unit; $v_t^{(1)}$ for the first encoder and $v_t^{(2)}$ for the second encoder. The pairs of code symbols are represented as numbers $j = (v_t^{(2)}, v_t^{(1)})$ written in binary representation.

These numbers are mapped into modulation signals

$$s_j(t) = \cos(wt + \varphi_j),$$

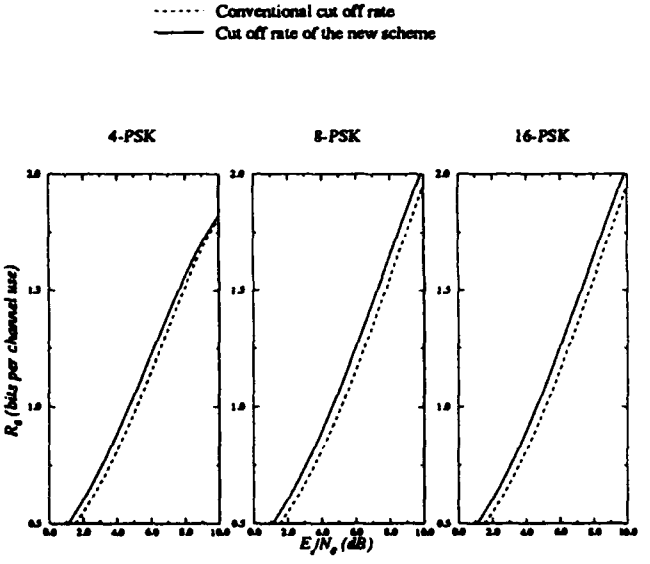
where $\varphi_j = j\pi/2$.

The decoder consists of two Viterbi decoders. The first one, corresponding to the first encoder, operates without taking the output sequence of the second encoder into account, i.e., it makes its estimates based on the conditional probability distribution for the received signal given that the code sequence $\underline{v}^{(1)} = (v_1^{(1)}, v_2^{(1)} \ldots)$ was transmitted.

The second Viterbi decoder estimates the second code sequence based not only on the second received sequence but also on the estimated code sequence from the first Viterbi decoder.

If the first decoder output is error free, then the second decoder knows exactly which of the signal pairs $(s_0(t), s_2(t))$ and $(s_1(t), s_3(t))$ that corresponds to each code symbol. Hence, the decoding process is reduced to decoding of BPSK signals.

In this paper we prove that we can choose the design parameters of the encoders and decoders such that for given reliability and complexity the transmission rate R for our scheme is greater than that for the conventional coded modulation scheme. Estimations of the systems performance for 4-PSK, 8-PSK, and 16-PSK (see fig.) with soft decisions show coding gains of about 0.5 dB compared to the conventional constructions.



References

[1] V.V. Ginzburg, "Multidimensional signals for a continuous channel", *Probl. of Inform. Transm.*, Vol. 23, No. 4, pp. 20–34, 1984.

[2] G. Ungerboeck, "Channel coding with multilevel/phase signals", *IEEE Trans. on Inform. Theory*, Vol. IT-28, pp. 55–67, 1982.

This work was supported in part by the Swedish Research Council for Engineering Sciences under Grant 92-661 and in part by the Royal Swedish Academy of Sciences in liasion with the Russian Academy of Sciences.

High Performance and Low Complexity Coded Modulation Schemes for Reliable Data Communications [1]

Sandeep Rajpal, Do Jun Rhee and Shu Lin
Department of Electrical Engineering
University of Hawaii at Manoa
Honolulu, Hawaii 96822, U.S.A.

This paper presents two coded modulation schemes for achieving reliable data transmission over the AWGN and the Rayleigh fading channels with large coding gains, high spectral efficiency, and reduced decoding complexity. In the first scheme, coded modulation [1] is used in conjunction with concatenation [2]. This combination of coded modulation and concatenation is known as concatenated coded modulation. In concatenated coded modulation schemes, the concatenation can be carried out either in single or multiple levels. The inner codes are bandwidth efficient modulation codes, and the outer codes are Reed-Solomon (RS) codes. If the inner codes, outer codes, and the level of concatenation are properly chosen, good error performance can be achieved with reduced decoding complexity, high spectral efficiency, and large coding gain.

In a single-level concatenated coded modulation scheme, a single RS code is concatenated with a single modulation code. In this paper, several single-level concatenated coded modulation schemes for the AWGN and the Rayleigh fading channels are proposed. In these schemes, both block and trellis modulation codes are being used as the inner codes and they are designed for either the AWGN channel or the Rayleigh fading channel and to have simple decoding complexity. In a q-level concatenated coded modulation scheme, q pairs of outer and inner codes are used. RS codes with different levels of error correcting capabilities are used as outer codes, and coset codes constructed from a block modulation code and its subcodes are used as the inner codes. The encoding and decoding are accomplished in q levels respectively. The decoding at each level consists of inner and outer code decodings. Closest coset decoding is performed at the first level inner code decoding based on the received sequence to obtain a sequence of estimated coset representatives for the first level inner code. This sequence of estimated coset representatives is converted to RS code symbols and decoded based on the first level RS outer code. From the decoded RS symbols, an estimated sequence of coset representatives is formed and the estimates are passed to the second level inner code decoder where the decoding process is repeated. Successive applications of closest coset decoding at each of the individual levels give estimates of the coset representatives at all the q levels. In this paper, several multilevel concatenated coded modulation schemes are proposed for the AWGN and the Rayleigh fading channels, and they achieve very good performance and large coding gains over uncoded reference systems of the same spectral efficiencies.

In the second proposed scheme, multilevel coded modulation is combined with multiple product codes to form two-dimensional product modulation codes. In a product modulation code, the column code is a block modulation code and algebraic codes of various error correcting capabilities are used as the row codes. Methods for constructing good product modulation codes for either the AWGN channel or the Rayleigh fading channel are proposed. A multi-stage decoding algorithm for these codes is devised, which reduces the decoding complexity while achieving good error performance.

Error performance bounds have been derived for both proposed schemes, which along with simulation results show that they achieve good error performance, large coding gains, and high spectral efficiency with reduced decoding complexity. The proposed schemes outperform the ones available in literature both in terms of coding gain and decoding complexity.

As an example, consider a single-level concatenated coded modulation scheme, in which the outer code is the NASA standard (255, 223) RS code over GF(2^8) and the inner code is a 2×2-dimensional trellis 8-PSK modulation code. For inner code construction, we choose the following three binary codes : $A_1 = (2,1,2)$, $A_2 = (2,2,1)$, and $A_3 = (2,2,1)$. These three binary codes are used to form a 2×2-dimensional 8-PSK signal space, denoted $\Lambda_0 = \lambda((2,1,2)*(2,2,1)*(2,2,1))$, which consists of 32 signal points, each signal point consists of two 8-PSK signals. The intra-set distance of Λ_0 is $D[\Lambda_0] = 1.172$. To partition Λ_0, we choose the following binary codes: $B_1 = (2,0,\infty)$, $B_2 = (2,1,2)$, and $B_3 = (2,2,1)$. These binary codes are then used to form a subspace of Λ_0, denoted $\Lambda_1 = \lambda((2,0,\infty)*(2,1,2)*(2,2,1))$, which consists of 8 signal points. The intra-set distance of Λ_1 is 4. The coset code Λ_0/Λ_1 consists of 4 cosets, each coset contains 8 signal points. A rate-1/2 convolutional code of constraint length $\nu = 3$ and minimum free branch distance $d_{\text{B-free}} = 3$ is chosen for the construction of the 2×2-dimensional trellis 8-PSK code. This code is generated by: $G(D) = (1 + D^2, D)$ and has a 4-state trellis diagram. The schematic diagram for constructing the desired 2×2-dimensional trellis 8-PSK code is shown in Figure 1. At each time instant, 4 information bits are encoded into two 8-PSK signals. All the possible signal sequences at the output of the overall encoder form a 2×2-dimensional trellis 8-PSK code. This code has a 4-state trellis diagram in which two **adjacent states** are connected by 8 **parallel branches** and each branch corresponds to a signal point in a coset of Λ_0/Λ_1. The spectral efficiency of the code is $\eta = 2$ bits/signal. The minimum free squared Euclidean distance of the code is 4. Figure 2 shows the bit-error-performance of the overall concatenated coded modulation scheme.

References

[1] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," **IEEE Trans. on Information Theory**, Vol. IT-28, No. 1, pp. 55-67, January 1982.

[2] G.D. Forney, Jr., **Concatenated Codes**, MIT Press, MA, 1966.

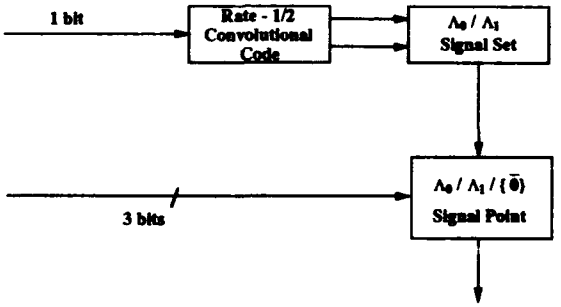


Figure 1 A trellis coded 2 X 2-dimensional 8-PSK encoder

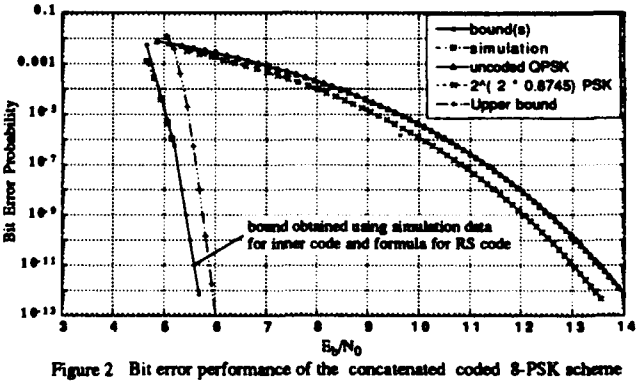


Figure 2 Bit error performance of the concatenated coded 8-PSK scheme

[1]This research was supported by NASA Grant NAG 5-931 and NSF Grant NCR-9115400

MULTILEVEL TERNARY LINE CODES WITH TRELLIS STRUCTURE

Ümit Aygölü and Erdal Panayırcı

Istanbul Technical University, Fac. of Electrical & Electronics Eng., 80626 Maslak, Istanbul, Turkey

Abstract: *In this paper, a combination of the two coding techniques given by Imai-Hirakawa and Pottie-Taylor is applied to the ternary (+,0,-) line code design problem. New ternary coding systems with reduced decoding complexities and improved error performance compared to those obtained by the classical Ungerboeck's trellis coding approach,are obtained. The decoding complexity for high coding rates are reduced by the proper choice of the punctured convolutional component codes for each partitioning level. A spectral null at zero frequency is obtained by the use of a 1/2 rate unit memory convolutional encoder at the last partitioning level which selects line codewords with opposite disparities in an alternated fashion so that the running digital sum values vary in a finite interval.*

Summary

In baseband digital transmission systems, the signal to be transmitted must have zero dc component and as small as possible power spectral components at low frequencies along with having a sufficient timing content. This avoids any dc power feeding over the line, reduces the low-frequency noise and allows the extraction of clock information.If the baseband digital signal is transmitted as a binary unipolar sequence, these requirements are not satisfied.For this purpose, line coding techniques are employed [1]. A line encoder transforms the binary sequences feeded at a rate R bit/sec to its input, into R' symbol/sec rate L-level ($L > 2$) sequences at its output and provides a redundancy of $R'log_2 L - R$ in information rate.

The application of the Ungerboeck's [2] trellis coding technique to the baseband ternary $(+,0,-)$ line code design problem is realized in [3] where the basic requirements for the baseband digital signal transmission and the error performance improvement by trellis coding are considered as an entity during the design phase. For rates $R = n/n+1$ $(n = 1,2,3)$ ternary line encoders are designed based on 2^{n+2}-element alphabet which consists of codewords with 0,+1 and -1 polarities and a proposed codeword assignment model. Coding gains of $3-3.52dB$ are obtained with respect to the classical paired-selected ternary line code.

In this paper, based on the multilevel coding approach, some new ternary line encoders with lower decoding complexity and improved coding gains,compared to those obtained by the Ungerboeck's technique, are proposed. The multilevel coding scheme given first by Imai and Hirakava[4] employes at each signalling interval, one output bit of each of several binary error control encoders to construct the signal to be transmitted. An important advantage of the multilevel coding is the possibility of suboptimum multistage decoding of each code with decoded information transferred from one stage to the next. This allows to reduce the decoding complexity at each stage and therefore for the overall system. Pottie and Taylor[5] have presented a generalized version of the multilevel coding technique where n_i output bits of each $R_i = k_i/n_i$ rate component encoder are used to partition the signal subsets determined at the preceding stage into 2^{n_i} new subsets with fewer signals. In our work, we use a combination of these techniques to obtain increasing minimum subset distances $\Delta_0 \leq \Delta_1 \leq \cdots \Delta_{M-1}$ given by the set partitioning method. Here, M represents the number of coding levels. Our aim is to form the optimal ternary codeword alphabets for several codeword lengths in order to achieve asymptotic coding gains at the lowest possible decoding complexity.

The first step of the multilevel line code design procedure is to construct an alphabet consisting of 2^{n+1} maximally distinct ternary $(+,0,-)$ n-length codewords. For this purpose, we use ternary sequences with only $0,\pm 2$ disparities (word digital sums). Half of the 2^{n+1} codewords are chosen having zero disparity and the rest having equal number of positive $(+2)$ and negative (-2) disparities. the codeword alphabet S_0 is then divided into M-level nonoverlapping subsets to form a partition chain $S_0/S_1/\cdots/S_{M-1}$ with minimum subset distances $\Delta_0 \leq \Delta_1 \leq \cdots \Delta_{M-1}$, respectively. To each partitioning level $S_{i-1}/S_i, i = 1,2,\ldots,M-2$ a binary component code is associated with free Hamming distance d_H, related to the free Euclidean distance d_{fED} of the overall system by

$$d^2_{fED} = \min_{1 \leq i \leq M-2} [\Delta_i^2 d_{Hi}, d^2_{fM-1}] \qquad (1)$$

where d^2_{fM-1} is the free Euclidean distance of the encoder corresponding to the last partitioning level $M-1$. At the partitioning levels $i = 1,2,\ldots,M-2$, we use punctured convolutional codes $C_1, C_2, \ldots, C_{M-2}$ due to their relatively lower decoding complexities. At the last level $M-1$ where the subsets including only one ternary codeword are obtained, we take the basic line requirements into account and use in all cases a $1/2$ rate unit-memory encoder which chooses line codewords with $+2$ and -2 disparities in an alternated fashion. Thus, we restrict the values of the running digital sum(RDS) at each line coding step to $+2,0$ and -2. Note that, at each signalling interval, the two input bits of this encoder are used to determine a ternary line codeword. Therefore, the linear relation between Hamming and Euclidean distances are not valid for this encoding level. The use of this $1/2$ rate encoder results in some slight losses in the number of data bits transmitted per ternary codeword, to obtain systems with good error and complexity properties, compared to the encoders given in [3] using Ungerboeck's approach. Thus, for the sake of comparison, we use the asymptotic coding gain(ACG) defined as,

$$ACG[dB] = 10log\frac{(d^2_{fED}/E_s)_m}{(d^2_{fED}/E_s)_U} - 10log\frac{R_U}{R_m} \qquad (2)$$

where R_U and R_m represent the coding rates(number of data bits per transmitted ternary codeword) of Ungerboeck's type and multilevel ternary line codes, respectively. E_s is the average codeword energy. ACGs up to $2.5dB$ are obtained with significantly reduced decoding complexities.

References

[1] N.Q.Duc, "Line coding techniques for baseband digital transmission", *A.T.R.*, vol.9,1,1975.

[2] G.Ungerboeck,"Channel coding with multilevel/phase signals", *IEEE Trans. on Inf. Theory*,vol.IT-28, January 1982.

[3] Ü.Aygölü and E.Panayırcı," New ternary line codes based on trellis structure", to appear in *IEEE Trans. on Commun.*,1992.

[4] H.Imai and S. Hirakawa,"A new multilevel coding method using error-correcting codes",*IEEE Trans. Information Theory*, vol.IT-23, May 1977.

[5] G.J.Pottie ans D.P Taylor," Multilevel codes based on partitioning", *IEEE Trans. on Inf Theory*,vol.IT-35, January 1989.

On the Construction and Dimensionality of Linear Block Code Trellises

Alan D. Kot and C. Leung

Department of Electrical Engineering
University of British Columbia
Vancouver, B.C. V6T 1Z4 Canada

Abstract

The trellis construction methods of Wolf [1], Massey [2], and Forney [3] for general linear block codes are briefly reviewed. An isomorphism between a trellis constructed using Massey's method and one constructed using Wolf's method is derived. It is confirmed that Wolf's and Massey's trellis constructions also yield minimal trellises. Two simplified methods for minimal trellis construction are presented, along with a method to calculate the trellis dimensions that is an alternative to the methods of [2] and [3]. An improvement is found to a lower bound on the maximum trellis dimension due to Muder [4]. It is shown that when equivalent codes are constructed by permutations of the symbol positions the resulting trellis dimensions are fixed near either end, while in the central portion of the trellis the dimensions vary between an attainable upper bound and a lower bound. From the lower bound on the trellis dimensions in the central portion of the trellis it is seen that only codes (and their duals) that meet a certain condition on their minimum distances can possibly have a trellis with a relatively small number of states.

Summary

Compared to convolutional codes, the trellis representations of linear block codes have been discussed much less frequently, with only a few papers appearing within twenty years of the introduction by Forney of the convolutional code trellis. Of these papers, the introduction of linear block code trellises appears in [1][2][5]. Recently, in [3] and [4], trellis construction and the trellis state-space dimensions were re-examined. Here, we continue this examination of trellis construction and dimensionality for general linear block codes.

Wolf's trellis construction for a general linear (n, k) block code $\mathbf{C}$ begins by generating an *unexpurgated trellis* that represents all uncoded sequences of length n. The trellis states are taken to be a partially formed syndrome vector. The code trellis is then formed by expurgating all paths that do not lead to the zero state at depth n. Massey's trellis construction for a general linear block code assigns a state at depth l in the trellis to be the vector of parity symbols in the codeword *tail*, as determined by the information symbols in the codeword *head*.[1] An isomorphism between a trellis constructed using Massey's method and one constructed using Wolf's method is found by comparing the state assignment equations. To show that these methods produce minimal trellises, we use a condition from [4] that specifies that a state at depth l must be assigned to those heads of the code tree $\hat{T}$ that satisfy the equivalence relation

$$\mathbf{c}^h \sim_t \mathbf{c}'^h$$

where $\sim_t$ indicates that $\mathbf{c}^h$ and $\mathbf{c}'^h$ share the same set of tails.

Based on Wolf's and Massey's trellis state assignment methods, we present two simplified methods for constructing minimal trellises. The simplified trellis construction methods avoid the generation of an unexpurgated trellis followed by expurgation of non-codeword tails, as used in Wolf's method for general linear block codes. Both methods also avoid matrix multiplications at each extension of a head to form the trellis, as used in Massey's method for non-systematic linear block codes or in Forney's method for general linear block codes. The new methods should be useful for complete trellis construction and for reduced-state tree/trellis searches.

A method for calculating the trellis dimensions is presented that is an alternative to those of [2] or [3]. Using this method, an improvement is found to Muder's lower bounds on the maximum trellis dimension (denoted s) for linear block codes. Muder's lower bounds can be summarized into a single expression,

$$s \geq \min\left(d_{min} - 1 - \Delta, d^{\perp}_{min} - 1 - \Delta^{\perp}\right)$$

where d_{min} and $d^{\perp}_{min}$ are the minimum distance of $\mathbf{C}$ and its dual $\mathbf{C}^{\perp}$, respectively; and where $\Delta \triangleq n - k - (d_{min} - 1)$ and $\Delta^{\perp} \triangleq k - \left(d^{\perp}_{min} - 1\right)$. The improved lower bound is

$$s \geq \min\left(d_{min} - 1, d^{\perp}_{min} - 1\right).$$

It is also shown that the trellis dimensions remain fixed near either end of the trellis despite symbol position permutations, and that a lower bound on the *minimum* trellis dimension in the central portion of the trellis (denoted s'), is given by

$$s' \geq \max\left[0, \min\left(d_{min} - 1 - \Delta, d^{\perp}_{min} - 1 - \Delta^{\perp}\right)\right].$$

This establishes that only codes (and their duals) that have a smallest minimum distance $\min\left(d_{min}, d^{\perp}_{min}\right)$ significantly less than the corresponding Singleton bound can possibly have a trellis with few states relative to $\min\left(q^k, q^{n-k}\right)$.

References

[1] J. K. Wolf, "Efficient maximum likelihood decoding of linear block codes using a trellis," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 76–80, Jan. 1978.

[2] J. L. Massey, "Foundations and methods of channel coding," in *Proc. of the Int. Conf. on Info. Theory and Systems*, vol. 65, NTG-Fachberichte, Sept. 1978.

[3] G. D. Forney, "Coset codes — part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1152–1187, Sept. 1988. Appendix A.

[4] D. J. Muder, "Minimal trellises for block codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1049–1053, Sept. 1988.

[5] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, pp. 284–287, Mar. 1974.

[1] The *head* $\mathbf{c}^h$ and the *tail* $\mathbf{c}^t$ of a codeword $\mathbf{c} \in \mathbf{C}$ are the first l symbols, and the last $n - l$ symbols of $\mathbf{c}$, respectively; so that $\mathbf{c} = (\mathbf{c}^h, \mathbf{c}^t)$.

This work was supported in part by a Natural Sciences and Engineering Research Council (NSERC) Scholarship, a B.C. Science Council G.R.E.A.T. Award, a B.C. Advanced Systems Institute Scholarship, and by NSERC Grant OGP0001731.

A new construction method for t-EC/AUED codes based on t-EC codes

Kenji Yoshida† Hajime Jinushi†† Kohichi Sakaniwa††
jinu@ss.titech.ac.jp sakaniwa@ss.titech.ac.jp

† NTT Data Communications Systems Corporation
†† Dept. of Electrical & Electronic Eng., Tokyo Institute of Technology, Tokyo, 152 Japan.

1 Introduction

Recently, many research works have been conducted to look for t (symmetric) error correcting and all unidirectional error detecting (t-EC/AUED) codes based on t-EC codes[1–5]. In this paper, we propose a new construction method for a t-EC/AUED code which is constructed by appending redundant bits, which consist of two parts, to the codewords of a base t-EC code. The first part of redundant bits is constructed systematically and the second by searching algorithms. The mapping from the codewords of a t-EC base code to the two parts of redundant bits is very simple and determined by the Hamming weights of codewords of the base code. It is shown that the proposed t-EC/AUED codes achieve higher information rates and require smaller encoding tables compared to the conventional ones.

We denote the number of $1 \rightarrow 0$ crossovers from $X = (x_0, x_1, \ldots, x_{n-1})$ to $Y = (y_0, y_1, \ldots, y_{n-1})(x_i, y_i \in \{0,1\})$ by $N(X,Y)$, the Hamming weight of X by $W(X)$. For X and $Z = (z_0, z_1, \ldots, z_{m-1})$, we denote the concatenation of X and Z by $XZ = (x_1, x_1, \ldots, x_{n-1}, z_0, z_1, z_{m-1})$. We also denote the cardinality of a set S by $|S|$, the least integer not less than a by $\lceil a \rceil$ and the greatest integer not greater than a by $\lfloor a \rfloor$.

In the following, the t-EC/AUED codes proposed in [1] and [2] are referred to as code C_0 and C_1', respectively.

2 Code Construction

We denote an $(n, k, 2t+1)$ base code by D and define $w_M \triangleq \max\{W(X) \mid X \in D\}$. Let $A(m) = \{A_0^{(m)}, A_1^{(m)}, \ldots, A_{2^m-1}^{(m)}\}$ be a set of binary m-tuples and $B_m(b,t) = \{B_0^{(m)}, B_1^{(m)}, \ldots, B_N^{(m)}\}$ $(N \geq \lfloor w_M/m \rfloor)$ a set of binary b-tuples satisfying

$$N(A_p^{(m)}, A_q^{(m)}) \geq \max\{\lceil (q-p)/2, \rceil, 0\} \quad (1)$$

$$N(B_u^{(m)}, B_v^{(m)}) \geq \min\{t+1, m(v-u)\}, \quad \text{for } u < v. \quad (2)$$

Then, the proposed t-EC/AUED code C_m is defined by

$$C_m = \{X A_p^{(m)} B_u^{(m)} \mid X \in D,\ A_p^{(m)} \in A(m),\ B_u^{(m)} \in B_m(b,t)\}, \quad (3)$$

where

$$W(X) = 2mu + p, \quad 0 \leq p < 2m. \quad (4)$$

Theorem 1 Code C_m is a t-EC/AUED code. ¶

Though a construction method for $A(m)$ is already given in [1], we propose a new systematic construction method for $A(m)$, which introduces a hierarchy into the class of proposed codes $\{C_m\}$. (see Theorem 3 below.)

Lemma 1 Define $A(m) = \{A_0^{(m)}, A_1^{(m)}, \ldots, A_{2^m-1}^{(m)}\}$ by

$$A_0^{(1)} = (1), \quad A_1^{(1)} = (0) \quad (5)$$

$$(A_0^{(m)} \quad \cdots \quad A_{2^m-3}^{(m)} \quad A_{2^m-2}^{(m)} \quad A_{2^m-1}^{(m)})^T$$

$$= \begin{pmatrix} 1 & \cdots & 1 & 0 & 0 \\ A_0^{(m-1)} & \cdots & A_{2^m-3}^{(m-1)} & A_{2^m-4}^{(m-1)} & A_{2^m-3}^{(m-1)} \end{pmatrix}^T, \quad m = 2, 3, \ldots. \quad (6)$$

Then, $A(m)$ satisfies Eq.(1). ¶

It is shown that the proposed code has following properties.

Theorem 2 For given $B_1(b,t)$, if there exists an $(n+1+b, k)$ t-EC/AUED code C_1'[2], we can construct an $(n+1+b, k)$ t-EC/AUED code C_1 by using the same $B_1(b,t)$. The converse is not always the case. ¶

Theorem 3 $C_m (m \geq 1)$ can be regarded as C_1 or C_0. ¶

Theorem 3 states that the information rate of C_0 is not less than that of $C_m (m \geq 1)$ and the information rate of C_1 is not less than that of $C_m (m \geq 2)$. However, it is easy to realize that encoding the proposed code $C_m (m \geq 1)$ requires tables of sizes $\lceil (n+1)/2m \rceil$ and $2m$ for mapping from u to $B_u^{(m)}$ and from p to $A_p^{(m)}$, while C_0 and C_1' require tables of sizes n and $\lfloor (n+1)/2 \rfloor$, respectively. Therefore the size of encoding table for the proposed code $C_m (m \geq 1)$ is about $1/2m$ of that for C_0 and about $1/m$ of that for C_1', if n is large.

3 Construction of $B_m(b,t)$

In order to obtain efficient codes by the proposed method, it is very important to prepare $B_m(b,t)$ which has as many elements as possible for fixed b and t. For lack of space, we only show three algorithms to construct $B_m(b,t)$ among six algorithms we considered. Algorithm 1 is obtained by extending the algorithm for $B_1(b,t)$ given in [3] to $B_m(b,t)$ and Algorithms 2 by modifying Algorithm 1. In Algorithm 1, we denote by $\{X_{w,r} = (x_0^{(r)}, x_1^{(r)}, \ldots, x_{b-1}^{(r)})(r = 1, 2, \ldots, \binom{b}{w})\}$ the set of binary b-tuples of Hamming weight w and numbered with r according to the rule that $r_1 < r_2$ *iff* $\sum_{i=0}^{b-1} x_i^{(r_1)} 2^i < \sum_{i=0}^{b-1} x_i^{(r_2)} 2^i$. Algorithm 1 examine if $X_{w,r}$ can be an element of $B_m(b,t)$ by the function *TEST*. *TEST*$(X_{w,r})$ outputs *true* if $\{B_0^{(m)}, B_1^{(m)}, \ldots, B_u^{(m)}, B_{u+1}^{(m)} = X_{w,r}\}$ satisfies Eq.(2.6) in [1] for $m = 0$ and Eq.(2) for $m \geq 1$, otherwise outputs *false*.

Algorithm 1

```
begin
  B_0^(m) := (11...1); u := 0;
  for w := b - 1 downto 0 do
    for r := 1 to (b choose w) do
      if TEST(X_{w,r}) = true then
        begin
          u := u + 1; B_u^(m) := X_{w,r}
        end;
  B_m(b,t) := {B_0^(m), B_1^(m), ..., B_u^(m)}
end
```

Algorithm 2 Modify Algorithm 1 so as to search $\{X_{w,r}\}$ from $w = 1$ to $w = b$ with the initial condition $B_u^{(m)} = (00\ldots0)$.

Algorithm 3 For a given $B_m(b,t)$, construct $B_0(b+m, t)$ by concatenating $B_m(b,t)$ with $A(m)$ and construct $B_1(b+m-1, t)$ by concatenating $B_m(b,t)$ with the first $m-1$ bits of $A(m)$. ¶

As an example, we show in Table 1 the largest values of $|B_0(b,t)|$ among those obtained by Algorithms 1 through 3 for several values of b and t, in comparison with the best known results so far.

We can see from Table 1 that many better results are obtained by the newly proposed algorithms.

4 Performance of the Proposed Codes

As an example, we show in Table 2 the number of additional bits required to construct SEC/AUED codes from single error correcting base codes by the proposed method in comparison to the best known results so far[1–5](shown in the column *old*). From Table 2, we can see that the proposed codes need less additional bits than the conventional ones.

Table 1: $|B_0(b,t)|$

	$t=1$	$t=2$	$t=3$	$t=4$
1	2	2	2	2
2	4	4	4	4
3	6	6	6	6
4	8	8	8	8
5	[3]14(12)	10	10	10
6	[2]21(16)	[1]13(12)	12	12
7	[3]36(24)	[3]20(16)	14	14
8	[3]58(48)	[1]25(20)	[1]19(16)	16
9	[3]86(72)	[3]38(24)	[3]26(20)	18
10	144	[3]50(32)	[3]30(24)	[1]24(20)
11	248	[3]70(48)	[1]3?(28)	[3]32(24)
12	432	[3]94(72)	[1]<x(32)	[3]36(28)
13	– – –	[3]142(120)	[3]70(40)	[1]43(32)
14	– – –	[3]220(216)	[1]84(48)	[1]54(36)
15	– – –	[3]396(392)	[3]112(72)	[1]65(40)
16	– – –	– – –	[3]154(120)	[1]80(56)
17	– – –	– – –	[3]212(180)	[3]102(72)
18	– – –	– – –	[3]298(264)	[3]124(104)
19	– – –	– – –	[3]496(488)	[3]168(156)
20	– – –	– – –	– – –	216
21	– – –	– – –	– – –	288
22	– – –	– – –	– – –	[3]374(368)

- Values without mark are the results due to [1].
- The new values obtained by the proposed algorithms are marked by a number at the upper left corner which indicates the algorithm used.
- The best known results are also shown in parentheses.

Table 2: SEC/AUED codes

n	$r = m+b$					
	C_0	C_1	C_2	C_3	C_4	old
3	2	2	2	3	4	2
4	3	3	4	3	4	3
6	4	4	4	5	4	4
8	5	5	6	5	6	5
12	5	5	6	7	6	5
14	6	6	6	7	6	6
16	6	6	7	7	8	6
19	6	6	7	7	8	7
20	6	7	7	7	8	7
21	7	7	7	7	8	7
24	7	7	8	8	8	7
32	7	7	8	8	9	7
35	7	7	8	8	9	8
36	8	8	8	8	9	8
48	8	8	9	9	10	8
57	8	8	9	9	10	9
58	9	9	9	9	10	9
72	9	9	10	10	10	9
85	9	9	10	10	10	10
86	10	10	10	10	10	10
96	10	10	10	10	11	10
108	10	10	10	11	11	10
144	11	11	11	11	12	11
216	11	11	11	12	12	11
248	12	12	12	12	12	12
256	12	12	12	12	12	12

References

[1] M. Blaum and H. C. A. van Tilborg, "On t-error correcting/all unidirectional error detecting codes", *IEEE Trans. Comput.*, vol. **38**, no. 11, pp. 1493–1501, November 1989

[2] R. Andrew, "Construction of t-EC/AUED codes", *Electron. Lett.*, vol. **24**, no. 20, pp. 1257–1258, September 1988

[3] Y. Saitoh and H. Imai, "Andrew's t-EC/AUED codes", *Electron. Lett.*, vol. **25**, no. 15, pp. 949–950, July 1989

[4] S. Al-Blassam and B. Bose, "Asymmetric/unidirectional error correcting and detcting codes", in *Proc. AAECC-7*, Toulouse, France, June 1989

[5] F. J. H. Böinck and Henk. C. A. van Tilborg, "Construction and bounds for systematic tEC/AUED codes", *IEEE Trans. Inf. Theory*, vol. **36**, no. 6, pp. 1381–1390, November 1990

Some New Lower and Upper Bounds on Systematic tEC/AUED Codes *

Zhen Zhang, senior member, IEEE, Xiang-Gen Xia and Chungming Tu
Communication Sciences Institute, Dept. of EE-Systems
University of Southern California, Los Angeles, CA 90089-2565

Summary

An $[n,k]$ tEC/AUED code is a tEC/AUED code which is systematic on the first k positions. For an $[n,k]$ tEC/AUED code C, when k is fixed we want to minimize the length n. For fixed k, the lower and upper bounds for the minimized length n have been discussed by many researchers. In this research, we improve numerous existing lower and upper bounds. We first discuss the method to improve the lower bounds.

Let $N(\mathbf{a},\mathbf{b}) = |\{1 \le i \le n : a_i = 1 \wedge b_i = 0\}|$, then it is well-known that a code C is tEC/AUED iff $\forall_{\mathbf{a},\mathbf{b}\in C,\mathbf{a}\neq\mathbf{b}}\ [N(\mathbf{a},\mathbf{b}) \ge t+1]$. This implies that a 0EC/AUED code is an antichain and a tEC/AUED with $t > 0$ is a special antichain. Therefore, the well-known LYM inequality can be applied to these codes. Because the requirement for tEC/AUED codes is much stronger than the one for general antichains, we sharpen the LYM inequality to the so-called weak and strong LYM inequalities as follows.

Define

$$\mathcal{M}_n(m,t,i) \triangleq \sum_{j=0}^{i} \binom{m+t-2i}{t-2i+j}\binom{n-m-t+2i}{j},$$

$$\overline{\mathcal{M}}_n(m,t,0) = \mathcal{M}_n(m,t,0),$$

$$\overline{\mathcal{M}}_n(m,t,i) = max\{\mathcal{M}_n(m,t,i-1),\ \mathcal{M}_n(m,t,i)\},\ i > 0.$$

Let C be a tEC/AUED code, then we have the following results.

Weak LYM inequality for tEC/AUED codes: For each i with $0 \le i \le t$,

$$\sum_{\mathbf{c}\in C} \frac{\mathcal{M}_n(|\mathbf{c}|,t,i)}{\binom{n}{|\mathbf{c}|}} \le 1.$$

Strong LYM inequality for tEC/AUED codes: For each i with $0 \le i \le t$,

$$\sum_{\mathbf{c}\in C} \frac{\overline{\mathcal{M}}_n(|\mathbf{c}|,t,i)}{\binom{n}{|\mathbf{c}|}} \le 1.$$

By appropriately applying the above two inequalities to $[n,k]$ tEC/AUED codes, most existing lower bounds are improved by 1 bit and many of them are improved by 2 bits. For details, one can see [1,2,3].

Next, we have new constructions for $[n,k]$ tEC/AUED codes. Currently the best known constructions for the $[n,k]$ tEC/AUED codes are based on the "*descending tail matrices*" [1]. For each input message sequence, r_1 parity check bits are appended to make it a codeword in a t error correcting linear code, then a descending tail is appended to code the weights of the codewords.

Our idea is to find the inherent relations among the message sequences and design the tail part as a whole. Two methods (the group theoretic method and the linear code syndrome method) are found to efficiently divide the message sequences into "subcodes", and then the tail words are used to code the indecies of the subcodes as well as the weights of the codewords. These new constructions have complexities that are comparable to codes constructed by descending tail method. For more details, one can see [4]. The following table shows a part of the new lower and upper bounds.

t=1		t=2	
k	r	k	r
4	6	4	8
5	6–7	5-6	9–12
6	6–8	7	9–13(15)
7	6–8(9)	8	9–15
8	7–8(9)	9	(9)10–15
9	7–8(10)	10–11	10–16(17)
10–11	7–9(10)	12–13	(10)11–16(17)
12–14	(7)8–10(11)	14	11–16(18)
15	8–10(12)	15	11–19
16–17	8–11(12)	16–18	(11)12–19
18–19	(8)9–11(12)	19-21	12-19
20–25	9–12	22	(12)13–19
26	(9)10–12	23–25	(12)13–20
27–28	(9)10–13	26	(12)13–20(21)
29–30	(9)10–13(14)	27–28	13–20(21)
31	10–13(14)	29–30	(13)14–20(21)
		31	(13)14–20(22)

References

[1]. F.J.H.Böinck and H.C.A.Van Tilborg, "Constructions and bounds for systematic tEC/AUED codes", IEEE Trans. on Information Theory, Vol.36, pp.1381-1390, No.6, Nov. 1990.

[2]. Z. Zhang and X.G. Xia, "LYM-type inequalities for tEC/AUED codes," To appear in IEEE Trans. on Information Theory, Jan. 1993.

[3]. Z. Zhang and X.G. Xia, "LYM-type inequalities for t-antichains," Submitted to Discrete Mathematics.

[4]. Z. Zhang and C.Tu, "On the construction of systematic tEC/AUED codes," Submitted to IEEE Trans. On Information Theory, under revision.

*This research is supported in part by NSF under Grant NCR-8905052.

Coding for Simultaneous Correction and Detection of Skew in Parallel Asynchronous Communications

Mario Blaum and Jehoshua Bruck
IBM Research Division
Almaden Research Center
San Jose, CA 95120

Consider the following communication scheme [1, 2]: a binary vector of length n is transmitted using n parallel wires. Each wire represents a coordinate of the vector. The propagation delay in the wires varies. Arrival of a transition represents a 1 while absence of a transition represents a 0. The problem is to find an efficient communication scheme that will be delay-insensitive.

Let us represent the tracks with the numbers $1, 2, \ldots, n$. After the m-th transition has arrived, the receiver obtains a sequence $\hat{X}_m = x_1, x_2, \ldots, x_m$, where $1 \leq x_i \leq n$, and x_i represents the fact that the i-th transition was received at the x_i-th wire. The set $\{x_1, x_2, \ldots, x_m\}$ is the support (i.e., the set of non-zero coordinates) of a vector and determines uniquely a binary vector.

Verhoeff [2] studied the following problem: assuming that a vector X is transmitted, once reception has been completed, the receiver acknowledges receipt of the message. The next message is sent by the sender only after the receipt of the acknowledgement. The problem is finding a code $\mathcal{C}$ whose elements are messages such that the receiver can identify when transmission has been completed. It is easy to see, as proved in [2], that the codes having the right property are the so called unordered codes, i.e., all its elements are unordered vectors.

Here, we assume that there is no communication between receiver and sender between messages, except, perhaps, when errors are detected. The sender does not wait for acknowledgement before sending the next message. This makes transmission faster, since the waiting period between messages gets shorter. However, if we shorten the waiting period, transitions from Y might start to arrive before reception of X has been completed, a condition called *skew*.

In [1], coding strategies were studied that allow either detection or correction of skew between consecutive messages. Here, our aim is to study codes that can correct a certain amount of skew between messages, and detect an extra amount of skew when the skew correcting capability of the code has been exceeded. We generalize the results in [1].

Consider a transmitted vector X followed by some other vectors, giving a received sequence $\hat{Z}$. There are two parameters that are related to the skew. The first one, denoted $m(X;\hat{Z})$, denotes the index of the last transition in X before the occurrence of skew, i.e., the last transition in X before the arrival of either a transition not in X or a repeated arrival. The second one, denoted $r(X;\hat{Z})$, denotes the index of the last arrival in X. If there is no skew, $m(X;\hat{Z}) = r(X;\hat{Z})$. We are ready now to define the concept of skew of a vector X with respect to a sequence $\hat{Z}$.

Definition 1 Let X be a subset of $\{1, 2, \ldots, n\}$ (equivalently, X is a binary vector of length n). Let $\hat{Z} = x_1, x_2, \ldots, x_j, \ldots$ be a sequence whose elements are in $\{1, 2, \ldots, n\}$, $\hat{Z}_i = x_1, x_2, \ldots, x_i$, and Z_i the set corresponding to $\hat{Z}_i$. Let $m = m(X;\hat{Z})$ and $r = r(X;\hat{Z})$ be as defined above. We say that the skew of X with respect to $\hat{Z}$ is equal to (l_1, l_2) (notation, $\mathcal{S}(X;\hat{Z}) = (l_1, l_2)$), if and only if

$$l_1 = |(Z_r - Z_m) \cap X| \text{ and } l_2 = r - m - l_1.$$

Let $\mathcal{S}(X;\hat{Z}) = (l_1, l_2)$. We say that $\mathcal{S}(X;\hat{Z})$ does not exceed (s_1, s_2), denoted $\mathcal{S}(X;\hat{Z}) \leq (s_1, s_2)$, if $l_1 \leq s_1$ and $l_2 \leq s_2$. Otherwise, we say that $\mathcal{S}(X;\hat{Z})$ exceeds (s_1, s_2) (notation, $\mathcal{S}(X;\hat{Z}) > (s_1, s_2)$).

Definition 2 Let t_1, t_2, s_1, s_2 be 4 non-negative parameters and let $\mathcal{C}$ be a code. We say that $\mathcal{C}$ is (t_1, t_2)-skew-tolerant (ST) $(t_1 + s_1, t_2 + s_2)$-skew-detecting (SD) if, whenever a codeword X in $\mathcal{C}$ is transmitted followed by other codewords giving a received sequence $\hat{Z}$, then, by examining $\hat{Z}$, the code will correctly decode X provided that $(0, 0) \leq \mathcal{S}(X;\hat{Z}) \leq (t_1, t_2)$ and will detect the occurrence of skew when $(t_1, t_2) < \mathcal{S}(X;\hat{Z}) \leq (t_1 + s_1, t_2 + s_2)$.

Notice that a (t_1, t_2)-ST code is a (t_1, t_2)-ST $(t_1 + s_1, t_2 + s_2)$-SD code such that $s_1 = s_2 = 0$, and an (s_1, s_2)-SD code is a (t_1, t_2)-ST $(t_1 + s_1, t_2 + s_2)$-SD code such that $t_1 = t_2 = 0$ (compare with the definition of error correcting/detecting codes that can correct up to t errors and detect up to $t + s$ errors).

Given two binary vectors X and Y of length n, we denote by $N(X, Y)$ the number of coordinates in which X is 1 and Y is 0. The following is our main result:

Theorem 1 Let $\mathcal{C}$ be a code and let $t = \min\{t_1, t_2\}$, $T = \max\{t_1, t_2\}$, $s = \min\{s_1, s_2\}$, $S = \max\{s_1, s_2\}$, $\tau = \min\{t_1 + s_1, t_2 + s_2\}$ and $\rho = \max\{t_1 + s_1, t_2 + s_2\}$. Then, $\mathcal{C}$ is (t_1, t_2)-ST $(t_1 + s_1, t_2 + s_2)$-SD if and only if, for any two distinct codewords X and Y in $\mathcal{C}$, such that $N(X, Y) \leq N(Y, X)$, the following is true:

(a) If $(t_1 - t_2)(s_1 - s_2) \geq 0$, then at least one of the following 3 conditions occurs:

1. $N(X, Y) \geq \tau + 1$.
2. $N(X, Y) \geq T + 1$ and $N(Y, X) \geq \rho + 1$.
3. $N(X, Y) \geq 1$ and $N(Y, X) \geq t_1 + t_2 + S + 1$.

(b) If $(t_1 - t_2)(s_1 - s_2) < 0$, then at least one of the following 4 conditions occurs:

1. $N(X, Y) \geq \tau + 1$.
2. $N(X, Y) \geq T + 1$ and $N(Y, X) \geq \rho + 1$.
3. $N(X, Y) \geq t + 1$ and $N(Y, X) \geq t_1 + t_2 + s + 1$.
4. $N(X, Y) \geq 1$ and $N(Y, X) \geq t_1 + t_2 + S + 1$.

We will prove that the conditions are sufficient by giving a decoding algorithm and we also present codes satisfying the conditions.

References

[1] M. Blaum and J. Bruck, "Coding for Skew-Tolerant Parallel Asynchronous Communications," IBM Research Report, RJ 8268 (75629), July 1991, to appear in IEEE Trans. on Information Theory, March 1993.

[2] T. Verhoeff, "Delay-insensitive codes - an overview," Distributed Computing, 3:1-8, 1988.

Constructions of Skew-Tolerant and Skew-Detecting Codes

Mario Blaum and Jehoshua Bruck
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120
USA

Levon H. Khachatrian *
University of Bielefeld
Department of Mathematics SFB 343
P.O. Box 8640
D-4800 Bielefeld 1
Germany

In [4], a coding solution to the problem of parallel asynchronous communications was presented. After transmission of each codeword, the receiver acknowledges reception of the message through a handshake mechanism. In this way, skew between messages is avoided. From a coding point of view, the problem is identifying the end of a message. As pointed out in [4], the codes that accomplish this task are the so called unordered codes.

A more complicated coding situation occurs when acknowledgement of the message is not allowed. In principle, this is an attractive alternative, since it would allow pipelined utilization of the channel, with increased data throughput. However, the difficulty now is that there might be skew between messages, i.e., signals from a second transmitted vector may arrive before the current vector has been completely received.

Necessary and sufficient conditions for codes that can either detect or correct a certain amount of skew were given in [1]. For further motivation and description of the problem, the reader is referred to [1, 4]. Here, we present constructions of codes that can detect or tolerate skew below a certain threshold.

Given two binary vectors X and Y of length n, we denote by $N(X,Y)$ the number of coordinates in which X is 1 and Y is 0.

In [1] theorems that characterize (t_1,t_2)-skew-detecting and skew-tolerant codes were proven. Here we present the theorems in the form of a definition as follows:

Definition 1 Let t_1 and t_2 be two non-negative integers, and let $t=\min\{t_1,t_2\}$ and $T=\max\{t_1,t_2\}$. We say that a binary code $\mathcal{C}$ of length n is:

1. (t_1,t_2)-skew-detecting (SD) if and only if, for any pair of distinct codewords $X,Y\in\mathcal{C}$, at least one of the following two conditions occurs:

   (a) $\min\{N(X,Y),N(Y,X)\}\geq t+1$

   or

   (b) $\min\{N(X,Y),N(Y,X)\}\geq 1$ and $\max\{N(X,Y),N(Y,X)\}\geq T+1$.

2. (t_1,t_2)-skew-tolerant (ST) if and only if, for any pair of distinct codewords $X,Y\in\mathcal{C}$, at least one of the following two conditions occurs:

   (a) $\min\{N(X,Y),N(Y,X)\}\geq t+1$

   or

   (b) $\min\{N(X,Y),N(Y,X)\}\geq 1$ and $\max\{N(X,Y),N(Y,X)\}\geq t_1+t_2+1$.

We present a general method for constructing (t_1,t_2)-SD and ST codes. The procedure involves adding three tails to the information bits: the first tail encodes the information bits into an $(n',k,2t_1+2)$ error-correcting code; the second tail makes the code satisfy the conditions in Definition 1; the third tail merely unorders the code in a way analogous to the generalization of Berger's construction given in [1]. We also briefly discuss optimality issues of the constructions. More details can be found in [2]

References

[1] M. Blaum and J. Bruck, "Coding for Skew-Tolerant Parallel Asynchronous Communications," IBM Research Report, RJ 8268 (75629), July 1991. to appear in IEEE Trans. on Information Theory, March 1993.

[2] M. Blaum, J. Bruck and L. Khachatrian, "Construction of Skew-Tolerant and Skew-Detecting Codes," RJ 8909 (79996), August 1992, to appear in IEEE Trans. on Information Theory.

[3] L. H. Khachatrian, "Construction of (t_1,t_2)-tolerant Codes," to appear in Proceedings of Dilijan Conference, Sept. 1991.

[4] T. Verhoeff "Delay-insensitive codes - an overview," Distributed Computing, 3:1-8, 1988.

*On leave from the Institute of Problems of Informatics and Automation, Armenian Academy of Sciences.

SUPERIMPOSED CODES IN HAMMING SPACE

Thomas Ericson
Linköping University, Dept. of Electrical Engineering
S-581 83 Linköping, Sweden

Vladimir Levenshtein
Keldysh Institute for Applied Mathematics
Muisskaya Sq.,4, 125047 Moscow, Russia

Binary superimposed codes are considered. The superposition mechanism assumed is addition modulo-2. Various constructions and bounds are derived.

Introduction

The idea of superimposed codes was introduced in 1964 by Kautz-Singleton [1]. The application they had in mind was information retrieval and the superposition mechanism assumed was Boolean sum. Chien-Frazer [2] considered the same problem assuming modulo-2 addition as the superposition mechanism. Later authors have usually emphasized the application of superimposed codes in multiple-access communication [3],[4]. In the present investigation we adhere to that view while adopting the same superposition mechanism assumed by Chien-Frazer: addition modulo-2.

The problem

Let F denote the binary field and let $C \subseteq F^n$ be an n-length block-code over F. For any m; $0 \le m \le T \triangleq |C|$ denote by C^*_m the set of all codewords $x \in F^n$ which can be formed as a sum $x = x_1 + x_2 + \ldots + x_s$ of s distinct codewords x_i from C where $0 \le s \le m$. If they are all distinct – this is the case of interest to us – it is clear that C^*_m is a code of size

$$T^*_m \triangleq |C^*_m| = \sum_{i=0}^{m} \binom{T}{i} .$$

We say that the original code C is a superimposed (n,m,d,T)-code if the induced code C^*_m has minimum distance at least d: $d(C^*_m) \ge d$. The problem is to choose C so as to obtain the best possible trade-off between the parameters (n,m,d,T). A key result is the following.

Theorem 1: If for some k, $0 \le k \le n$, there exists an [n,k,d]-code and a [T,T-k,2m+1]-code, then there exists a superimposed (n,d,m,T)-code.

Proof: Let G be a generator matrix of an [n,k,d]-code and let H be a parity check matrix of an [T,T-k,2m+1]-code. A superimposed (n,d,m,T)-code is given by the rows of the matrix H'G (H' denotes transponse of H). Indeed, any linear combination of rows from the matrix H'G belongs to the [n,k,d]-code generated by G. Moreover, any 2m rows from H'G are linearly independent because the columns of H have exactly this property. Finally, it is obvious that H'G has T rows and n columns.

□

Example: Let G generate the Hamming code [n,k,d] = [15,11,3] and let H generate the Golay code [T,T-k,2m+1] = [23,12,7]. Theorem 1 produces a superimposed code with parameters (n,m,d,T) = (15,3,3,23).

Doubly perfect codes

Let d be odd, d=2t+1. It is clear that C^*_m is subject to any upper bound for binary codes with distance d=2t+1. Therefore the parameters (n,m,d,T) of a superimposed code C are subject to the following obvious bounds,

$$\sum_{i=0}^{m} \binom{T}{i} \le A(n,2t+1) \le \frac{2}{\sum_{i=0}^{t} \binom{n}{i}} .$$

We say that a code C satisfying the first bound is perfect, because for such codes the induced code C^*_m is as large as any binary code with distance d=2t+1. If both inequalities are satisfied with equality we say that the superimposed code C is doubly perfect.

Doubly perfect codes do exist. The above example with (n,m,d,T) = (15,3,3,23) is one example. Further examples are as follows:

n	m	d	T	
11	3	1	23	
15	3	3	23	
23	6	7	13	
$2^{2s+1}-1$	$2^{2s}-s-1$	3	$2^{2s+1}-2s-1$	s=1,2,3,...
2m	m	1	2m+1	m=2,3,...

References

[1] W.H. Kautz and R.C. Singleton, "Nonrandom binary superimposed codes", IEEE Trans. on Inf. Th., IT-10, No.4, pp. 367-377, 1969.

[2] R.T. Chien, W.D. Frazer, "An application of coding theory to document retrieval", IEEE Trans. Inform. Theory, IT-12, No. 2, pp. 92-96, 1966.

[3] A.G. Dyachkov and V.V. Rykov, "Bounds on the length of superimposed codes", Problemy Peredachi Informatsii, vol. 17, No. 2, pp. 26-38, 1981. English translation 1982.

[4] T. Ericson and L. Györfi, "Superimposed codes in R^n", IEEE Trans. Inform. Theory, IT-34, No.4, pp.877-880, 1988.

High-dimensional Symmetric Compacted Code

— Error-correcting of high bit error rate of $10^{-1} \sim 10^{-2}$ —

Masayasu HATA and Ichi TAKUMI

Intelligence and Computer Science, Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, Nagoya, 466 Japan

Concept of new code

In this paper, a long block code is conceived as a code string and it is wound up into a compact sized knot in an n dimensional torus space T^n (Ref.1,2). The code string is winding diagonally to the each dimensional fundamental cycles of the n dimensional torus. Therefore, the digits on the code string are scattered about the fundamental cycles and the time and space distances between digits on the each cycles and among the cycles are greatly expanded and independences between digits and among cycles are assured.

The digits on the each fundamental cycles form a single unit code. These unit codes are mutually independent against the transmission errors through the said independence. Therefore, we can give a modeling for an erroneous route as an error screen with an interval determined by the inverse of the mean bit error rate of the route. If the size of unit code of the fundamental cycle is smaller than the interval, the code can pass through without hit by the error screen.

And so the designing size of the each fundamental cycles of the code should be the same in order to obtain an optimized robustness and the code comes to have geometrical symmetries like a crystalline ball.

Example of proposed code

For this paper, the unit code which is constituting each dimension is a simple short length parity check code of the same length m. The code is constituted by means of product space of cyclic-shifted versions of one-dimensional parity check codes by n-fold orthogonally, which shows the structural symmetries and satisfies the n-dimensional parity check functions. The burst error correcting ability of the code is $(m-1)m^{n-2}$ in length with transmission rate $R = (1 - \frac{1}{m})^n$ of code length of m^n. The random error correcting abilities are also appreciably increased with the dimension n, especially for $n \geq 4$, exceeding the minimum distance limit of the code of $(2^n/2 - 1)$. The proposed code is defined as a quasi-cyclic code. The uncorrectable patterns of error are uniquely showed in the space as a symmetrical solid to the parity check axes, which is the same solid for both burst and random errors.

Results

By computer simulations, we could show that the code may be applicable to a worse quality of bit error rate of order of $10^{-1} \sim 10^{-2}$ and has a better decoding bit error rate for the range of transmission rate of $R = \frac{1}{2} \sim \frac{1}{4}$ than the convolution code.

As an example, the code of $m = 5$, $n = 5$ of length $m^n = 3,125$ and $R = 0.32768 \approx 0.33$ can almost perfectly correct 150 random errors, which correspond to 4.8×10^{-2} of bit error rate, and 200 errors with 99% correction, and can correct a burst error of 500 bits in length.

As a similar code, an n-D cyclic code has been proposed in late 60's and recently again attracted interests from the practical sides of decoding algorithm and the performances, however, the efforts are confined to only two dimensional case and not yet explored for high dimensional code exceeding two. Further, the n-D cyclic code is consisting of different sizes of length of dimension which are primitive each other and has no geometrical symmetries. And therefore, the the proposed code will be superior to the n-D cyclic code's operability for worse BER conditions.

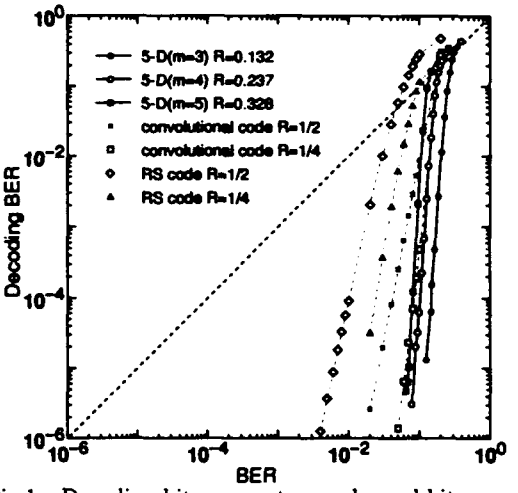


Fig.1 : Decoding bit error rate vs. channel bit error rate

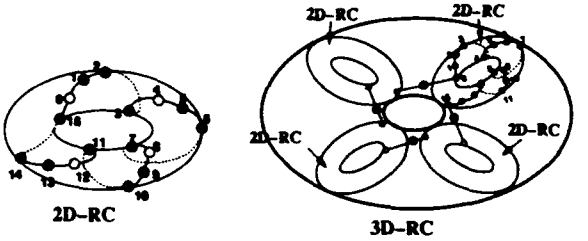


Fig.2 : Torus representation of two and three-dimensional proposed code

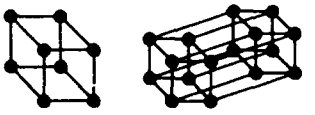


Fig.3 : Undetectable solid

Table 1 : Performances of proposed code

Dim. n	Size m	Length	Trans. rate	Correctable Burst length	Correctable Random Err. 100%	99.9%	99%
3	3	27	0.296	6	3	3	3
	4	64	0.442	12	3	3	3
	5	125	0.512	20	3	3	4
	6	216	0.579	30	3	3	4
4	3	81	0.198	18	7	8	10
	4	256	0.316	48	7	14	18
	5	625	0.410	100	7	22	29
	6	1296	0.482	180	7	29	40
5	3	243	0.132	54	15	35	41
	4	1024	0.237	192	15	95	111
	5	3125	0.328	500	15	200	241
	6	7776	0.402	1080	15	> 380	480

References

[1] Noda T. and Hata M. :"Ring code — code symmetry and uncorrectable errors —," IEICE Trans.,J73-A,2,pp.243-252, Feb. 1990.

[2] Hashimoto K. and Hata M. :"High-dimensional symmetry parity check code capable of correcting $10^{-1} \sim 10^{-2}$ random errors," IEICE Trans., J75-A,8,pp.1257-1266, Aug. 1992.

Some Families of Asymptotically Optimal Optical Orthogonal Codes

O. Moreno, Z. Zhang and P. V. Kumar*

Abstract Three related constructions for families of optical orthogonal codes are presented. All are asymptotically optimum in the sense that in each case, as the length of the sequences within the family approaches infinity, the ratio of family size to the maximum possible under the Johnson bound, approaches unity.

An (n,ω,λ)-optical orthogonal code (OOC) (see [1], [2]) C, $n > 1$, $1 \leq \omega \leq n$, $1 \leq \lambda \leq \omega$, is a family of $\{0,1\}$-sequences of length n and Hamming weight ω satisfying the following auto and cross-correlation conditions:

$$\sum_{k=0}^{n-1} x(k)x(k \oplus_n \tau) \leq \lambda \qquad (1)$$

for all sequences $x(.) \in C$ and all integers $\tau \not\equiv 0 \,(mod\, n)$ and

$$\sum_{k=0}^{n-1} x(k)y(k \oplus_n \tau) \leq \lambda \qquad (2)$$

for all pairs of sequences $x(.), y(.) \in C$ and all integers τ, where $\oplus_n$ denotes addition modulo n.

For a given set of values of n, ω and λ, let $\Phi(n,\omega,\lambda)$, denote the largest possible cardinality of an (n,ω,λ)-optical orthogonal code. Upper bounds for this function and several optimal constructions for $\lambda = 1$ and 2 can be found in [1]-[3]. An easy upper bound derived from the Johnson bound (see [1]) states that

$$\Phi(n,\omega,\lambda) \leq \lfloor \frac{A(n,2\omega-2\lambda,\omega)}{n} \rfloor \leq \frac{(n-1)(n-2)...(n-\lambda)}{\omega(\omega-1)...(\omega-\lambda)}. \qquad (3)$$

In this paper, three constructions ($\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$) for families of OOC's are presented. In every case, the families are asymptotically optimum in the sense that, as the length of the sequence family $\rightarrow \infty$, the ratio of the size of the OOC to that of the maximum permissible as determined by the bound in (3) above, approaches unity.

All three constructions make use of the following two ideas. Let n be an integer that can be expressed as the product $n = n_1 n_2$ of two relatively prime integers n_1 and n_2. Then, from an application of the Chinese remainder theorem, it follows that the construction of sets of $\{0,1\}$ sequences with periodic correlation bounded above by λ is completely equivalent to the task of constructing a collection of arrays whose doubly-periodic correlation is bounded above by λ. Secondly, the sequences in the OOC are required to have constant weight. The sequences in each of the three families $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ when represented in matrix appear as the graph of a function mapping $Z_{n_2} \rightarrow Z_{n_1}$. This guarantees that they all have constant weight (approximately) n_2. The functions in $\mathcal{A}$ and $\mathcal{B}$ are polynomials, whereas, construction $\mathcal{C}$ uses rational functions.

Precise parameters of the three families constructed are tabulated below. Reference [4] appeared after the initial preparation of this paper. The two papers share some material in common such as the idea behind the construction as well as some features of construction $\mathcal{A}$.

REFERENCES

[1] F. R. K. Chung, J. A. Salehi, and V. K. Wei, "Optical Orthogonal Codes: Design, Analysis, and Applications," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 595-604, May 1989.

[2] E. F. Brickell and V. K. Wei, "Optical Orthogonal Codes and Cyclic Block Designs," *Congressus Numerantium*, vol. 58, pp. 175-192, 1987.

[3] H. Chung and P. V. Kumar, "Optical Orthogonal Codes-New Bounds and an Optimal Construction," *IEEE Trans. Inform. Theory*, vol. 36, No. 4, pp. 866-873, July, 1990.

[4] Nguyen Q. A., L. Gyorfi and J. L. Massey "Constructions of Binary Constant-Weight Cyclic Codes and Cyclically Permutable Codes", *IEEE Transactions on Information Theory*, vol. IT-38, No. 3, May 1992.

Fam.	n	ω	λ	Size
$\mathcal{A}$	$p(p-1)$ p prime	$(p-1)$	t $1 \leq t \leq p-2$	$\frac{\sum_{d\|(p-1)} p^{\lfloor t/d \rfloor}\mu(d)}{p-1}$
$\mathcal{B}$	$(q-1)p$ $q = p^a$ p prime $a \geq 1$	$(p-t)$	t $1 \leq t \leq p-1$	$\frac{q}{p}\left(\frac{q^t-1}{q-1}\right)$
$\mathcal{C}$	q^2-1 $q = 2^a$ $a \geq 2$	$q-1$	$2t$ $1 \leq t \leq q-2$	$q \quad t=1$ $\geq q^3+q/2-1 \quad t=2$ $\geq q^5+q^3/2-(5q^2)/6 \quad t=3$ $\frac{q^{2t+1}-q^{2t-1}/2+\mathcal{O}(q^{t+2})}{q^2-1} \quad t \geq 4$

ON BINARY SYNCHRONIZATION ERROR CORRECTING CODES

A.S.J. Helberg, H.C. Ferreira
and W.A. Clarke
Laboratory for Cybernetics
Rand Afrikaans University
P.O. Box 524
AUCKLANDPARK
South Africa
2006

A.J. Vinck
Institut für Experimentelle
Mathematik
Universität GHS Essen
Ellernstrasse 29
4300 Essen 12
Germany

Abstract

We investigate the single bit insertion/deletion correcting codes as proposed by Levenshtein and others. [1 – 4,6]. These synchronization error correcting codes are derived from number theoretic constructions. The weight spectra and Hamming distance properties of these codes are found and a relationship between these properties is established. This relationship is extended to codes that can correct multiple random synchronization errors. From this general relationship, improved bounds on the cardinality of such multiple synchronization error correcting codes are found. From the new relationship between the weight and Hamming distance of synchronization error correcting codes, several new codes are found.

Introduction

In 1965 Levenshtein [1,2] found that a certain code construction technique developed by Varshamov and Tenengol'ts [3] could also yield codes that are capable of correcting single synchronization errors. This work was later extended by several workers [4, 6]. Synchronization errors manifest themselves in the bit stream as the deletion of a valid symbol or the insertion of such a symbol. We first investigated the binary single error correcting codes as developed by Levenshtein to determine some new properties. The weight spectra and the Hamming distance profiles for several short length codes were determined. The dc–free subcodes of the Levenshtein codes were investigated as well as runlength limited concatenatable subsets that have a minimum runlength constraint of 1, i.e. there is at least one zero between ones. The spectra of some of the dc–free codes are also presented.

New bounds on the cardinality

From the investigation of the Levenshtein codes, the relationship between the weight of a codeword and the Hamming distance between other Levenshtein codewords of specific weights was determined. From this relationship follows several propositions that establishe a similar relationship for codes that are capable of correcting two synchronization errors when using a number theoretic construction technique similar to that of Levenshtein. By using the abovementioned relationships we establish the following new upper and lower bounds on the cardinality of double synchronization error correcting codes:

Upper bound:

$$|C| < 2 + \sum_{w=3}^{n-3} (n/w \cdot A(n-1,6,w-1)) \quad n \geq w \geq 1 \qquad (1)$$

where w is the weight (i.e. the number of "ones") of the codeword of length n and $A(n,d,w)$ is the number of codewords of weight w which differ from each other in at least d positions. The bound in (1) is due partly to Johnsson [5] who derived an upper bound for constant weight codes with a certain minimum Hamming distance.

Lower bound:

$$|C| \geq 2^n / \sum_{j=0}^{5} \binom{n-1}{j} \qquad (2)$$

which is a Gilbert – Varshamov type lower bound on the cardinality of linear error correcting codes. These new bounds are compared to the known bounds in Figure 1.

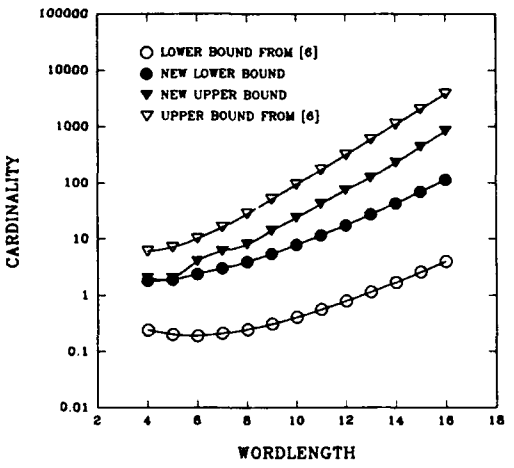


Figure 1 : Bounds on the cardinality of double synchronization error correcting codes.

New codes

We also present general construction techniques for codes that are capable of correcting two adjacent synchronization errors and two random synchronization errors respectively. The cardinality of these codes for short word lengths is also given. By combining certain dc free criteria to the balanced subcodes of the Levenshtein codes, we found a class of "multipurpose" codes which have enhanced dc suppression, are able to correct either one synchronization error or one additive error and require less bandwidth than similar codes.

References

[1] V. I. Levenshtein, "Binary codes capable of correcting deletions. insertions and reversals," (Russian:) Doklady Akademii Nauk SSSR 163(4), pp 845 – 848, 1965: (English:) Soviet Physics–Doklady 10(8), pp 707 –710, 1966.

[2] V. I. Levenshtein, "On perfect codes capable of correcting deletions of a character," Fourth Joint Swedish – Soviet International Workshop on Information Theory, August 27 – September 1, 1989, Gotland, Sweden, pp 199–204.

[3] R. R. Varshamov and G. M. Tenengol'ts, "On asymmetrical error – correcting codes," (In Russian), Avtomatika i Telemehanika, vol 26, N2, pp 288 – 292, 1965

[4] H. D. L. Hollmann, "A relation between Levenshtein–type distances and insertion–and–deletion correcting capabilities of codes," Internal report, Philips Research Laboratories, Eindhoven, The Netherlands, November 9, 1990

[5] S. M. Johnsson, "Improved asymptotic bounds for error–correcting codes," IEEE Transactions on Information Theory, vol IT–17, pp 198 – 205, July 1963.

[6] P. A. Bours, "Bounds for codes that are capable of correcting insertions and deletions," Internal report, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, July 1991.

The Two-way Channel as a Computer Game

ALPHONS H. A. BLOEMEN, HENDRIK B. MEEUWISSEN, AND J. PIETER M. SCHALKWIJK
Group on Information and Communication Theory, Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven E-mail: pnons@ei.ele.tue.nl

Abstract

In his 1961 paper on two-way channels (TWC's) Shannon derived single-letter inner- and outer bounds to the capacity region. The first part of this paper is a survey of earlier results on TWC's in general and the binary multiplying channel (BMC) in particular. The second part is devoted to a new approach to the problem of determining the capacity region of the BMC. Based on Schalkwijk's 1982 idea to represent symmetric, $R_1 = R_2 = R$, coding strategies for deterministic two-way channels as progressive subdivisions of an $M \times M$ square, we developed a computer game $\mathcal{AXE}$ as a development environment for new coding strategies. Playing $\mathcal{AXE}$ is simple and requires no background in information theory.

1 History

In order to approximate the capacity region of a TWC we start off with Shannon's [8] observation (1961) that the capacity region can be found from the per letter rate of increasingly long coding strategies. An initial hurdle was the fact that coding strategies, where the code sequence at each terminal not only depends on the message Θ being transmitted but also on the received sequence Y at that terminal, are hard to visualize.

A breakthrough [5] was made in 1982, when it was discovered that coding strategies for deterministic TWC's could be considered as strategies for subdividing the unit square. For the BMC a subdivision using three types of resolutions (to be referred to as i-, m- and o-resolutions) was found. In the case of equal rates on both directions this constructive coding strategy achieves 0.61914, in excess of Shannon's inner bound of 0.61695. Dueck [2] just previously proved by example that the capacity region of a TWC is in general larger than its inner bound.

A further [6] basic step was taken a year later in 1983. The m-resolution in the strategy [5] mentioned above was not efficient. This m-resolution takes place in an L-shape subregion. By collecting a number of those L-shapes the total resolution information that has to be sent from terminal 1 to terminal 2 and vice versa can be accumulated at each terminal. In the limit, the total resolution information can be transfered at the very rate of the resulting strategy using a technique [6] called *bootstrapping*, thus boosting the 0.61914 rate of the original strategy up to 0.63056. The resulting strategy effectively resembles our original strategy where the m-resolution has been eliminated. As this equivalent strategy is very simple and elegant, 0.63056 was initially thought to be the equal rate capacity of the BMC.

However, repeated trials to find a converse failed and suspicion regarding optimality arose. Accurate calculation of the rate of the bootstrapped strategy yields 0.6305552557. Finally, an improvement to 0.6305552986 was found, by having *two* initial i-resolutions and preserving the efficiency of the postponed o-resolution by a *transparency* condition. Another [7] slight improvement yields the tightest lower bound 0.6305552995 to the equal rate capacity as of to date.

Shannon's [8] upper bound of 0.69121 has been tightened by Zhen Zhang, et al. [11] to 0.64891 for general TWC's. The tightest upper bound as of now for T-channels (TC's), i. e. channels with two inputs and a single common output, found by Hekstra and Willems [3], yields 0.61628 for the BMC. It is our belief that the final result, at least for the BMC, is closer to the best inner bound of 0.6305552995. In order to find better upper bounds it will be necessary to consider the coding strategies in greater detail.

In classical one-way communication we distinguish [1] between the *information theoretic* and the *operational* capacity. Shannon's channel coding theorem shows these two capacities to be equal. The achievable rates of our original strategy [5] and of the bootstrapped strategy [6] are *information theoretic* rates. It was reasoned that these rates were also *operational* as they related to the size of resolution products in the unit square. Rigorous proofs of the *operationality* of the rates in [5] and [6] were given by Tolhuizen [9] and van Overveld [10], respectively.

2 The $\mathcal{AXE}$-program

There are several reasons for trying to find the capacity region of the BMC. First, the BMC was used as the simplest non trivial example of a TWC by Shannon in his original 1961 paper. This [8] paper marks the beginning of *network information theory*, see also Cover [1, chapter 14]. Second, once the BMC is solved, it seems likely that similar methods can be applied to solve all deterministic TWC's. Finally, it does not seem possible to make any progress on general (non) deterministic TWC's as long as the easier deterministic TWC's are not completely understood. As an example, there are 322 *distinct* [4] ternary deterministic TWC's of which 46 are T-channels.

Note that the best [6] lower bound of 0.6305552995 to the equal rate capacity is the result of a continuing line of research building upon our very first [5] strategy. Continuation of this research line does not look promising: it will result in more complicated strategies with more parameters and smaller improvements, and maybe there is an entirely different class of strategies that perform better. Computer search is unfeasable, since the number of possible strategies dwarfs Avogadro's number of $6.02 \cdot 10^{23}$ for relatively small message sets. However, people seem to have a feeling as to the proper shape of resolution products. To help people contructing strategies for binary and ternary TWCs on $M \times M$ squares, F. Hantz and A. Bloemen developed the computer puzzle game $\mathcal{AXE}$. The 'game' is played by editing resolutions already stored in the computer as part of a strategy tree. Any improvements replace the resolutions currently stored. There is no information theoretical background needed to play the game. By putting the game in public domain we hope to get some good coding strategies.

The results obtained so far look very promising. Early results were found by J. van der Leur, who developed a coding strategy resolving all message pairs of a 17×17 square with a rate of 0.61079. A first tuning step is the *save-up* method: square and rectangle-like resolution products are not resolved, but they are put together into a new $M \times M$ square. This technique already provides a coding strategy on the 6×6 square with rate 0.61795, and on the 11×11 square with rate 0.61984. These rates exceed Shannon's [8] inner bound rate of 0.61695 and Schalkwijk's [5] original rate 0.61914, respectively. The second step is *bootstrapping*, similar to [6], to weed out inefficient resolutions. For a 13×13 square, a discrete bootstrap strategy with rate 0.63050 has been constructed. The third step, transforming a discrete bootstrap strategy into a continuous bootstrap strategy will almost surely yield a new lower bound.

[1] Thomas Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[2] Gunter Dueck. The capacity region of the twc can exceed the inner bound. *Information and Control*, 40:258–266, 1979.

[3] Andries P. Hekstra, Frans M.J. Willems. Dependence balance bounds for single-output twcs. *IEEE Trans. on Information Theory*, IT-35(1):44–53, Jan 1989.

[4] Annelies Jacobs. On the capacity regions of ternary deterministic twcs. Master's thesis, Eindhoven University of Technology, Fac. E, Oct 1986.

[5] J. Pieter M. Schalkwijk. The binary multiplying channel - a coding scheme that operates beyond the Shannon inner bound. *IEEE Trans. on Information Theory*, IT-28(1):107–110, Jan 1982.

[6] J. Pieter M. Schalkwijk. On an extension of an achievable rate region for the binary multiplying channel. *IEEE Trans. on Information Theory*, IT-29(3):445–448, May 1983.

[7] J. Pieter M. Schalkwijk. Extending the achievable rate region of the binary multiplying channel. In *Proc. Int. Symp. on Information Theory*, page 302, Budapest, Hungary, Jun 1991.

[8] Claude E. Shannon. Two way communication channels. In *Proc. 4th Berkely Symposium on Mathematics, Statistics and Probability*, volume 1, pages 611–644, 1961. Reprinted in D. Slepian, editor, *Key papers in the Development of Information Theory*, pages 339–372, New York: IEEE Press, 1974.

[9] L. Tolhuizen. Discrete coding for the BMC, bases on Schalkwijk's strategy. In *Proceedings of the BeNeLux Symposium on Information Theory*, volume 6, pages 207–212, Mierlo, The Netherlands, 1985.

[10] Wilhelmina M.C.J. van Overveld. *On the Capacity Region for Deterministic Two-Way Channels and Write-unidirectional Memories*. PhD thesis, Eindhoven University of Technology, Fac. E, Jan 1991.

[11] Zhen Zhang, Toby Berger, and J. Pieter M. Schalkwijk. New outer bounds to capacity regions for twcs. *IEEE Trans. on Information Theory*, IT-32(3):383–386, May 1986.

Ten Good Rate (m-r)/pm Binary Quasi-Cyclic Codes[1]

T. Aaron Gulliver
Dept. of Systems & Computer Engineering, Carleton University
1125 Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6
Vijay K. Bhargava
Dept. of Electrical & Computer Engineering, University of Victoria
P.O. Box 3055, Victoria, B.C., Canada V8W 3P6

Quasi-Cyclic (QC) codes are a generalization of cyclic codes whereby a cyclic shift of a codeword by p positions results in another codeword. The class of QC codes is of interest because it contains many of the best known binary linear codes. The results presented here reinforce the statement that 'Quasi-Cyclic codes are good'.

The blocklength, n, of a QC code is a multiple of p, $n = mp$. Many of the results on QC codes presented in the literature concern those for which a generator matrix can be constructed from $m \times m$ circulant matrices, (with a suitable permutation of coordinates). In this case the generator matrix can be represented as

$$G = (c_0(x), c_1(x), c_2(x), c_3(x), \ldots, c_{p-1}(x)),$$

where the coefficients of the polynomial $c_i(x)$ are defined by the circulant matrix C_i. G is a (pm, m) code, and the dual code H is a $(pm, (p-1)m)$ code. To date most of the results on QC codes are concerned with these rate $1/p$ and $(p-1)/p$ codes. In this paper a generalization of the rate $1/p$ codes to rate $(m-r)/pm$ codes is presented based on the theory of 1-generator QC codes, which are a sub-class of QC codes. The *order* of a 1-generator QC code, V, is defined as

$$h(x) = \frac{x^m - 1}{(x^m - 1, c_0(x), c_1(x), \ldots, c_{p-1}(x))},$$

and k, the dimension of V, is equal to the degree of $h(x)$. If $(x^m - 1, c_i(x)) = 1$, the dimension of V is m, and G is a generator matrix for V. If, $\deg(h(x)) = k < m$, a generator matrix for V can be constructed by deleting $r = m - k$ rows of G. In this paper, codes are constructed with $\deg(h(x)) < k - 1$.

Linear programming is efficient for finding optimal codes if k is small. However, an exhaustive search quickly becomes intractable as k increases. Therefore non-exhaustive techniques must be employed to search for good codes. A greedy exchange algorithm has previously been used with good results because it is computationally simple and therefore able to cover a large number of codes quickly, and so was also used in this case. Although the resulting codes are not guaranteed to be optimal, they can be compared with known bounds to determine if a better code can exist. The results of this search are ten codes which improve the known lower bounds on the minimum distance of binary linear codes as tabulated by Verhoeff.

[1]This research was supported in part by the Natural Science and Engineering Research Council of Canada

ON CERTAIN SUBCODES OF THE BINARY EXTENDED QUADRATIC RESIDUE CODES*

Xuemin Chen, I. S. Reed and T. K. Truong

Communication Sciences Institute
Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089-2565

It is shown in this paper that there exists a class of codes generated by the known self-dual binary extended quadratic residue (EQR) codes. Each code in this new class has an even better error-control rate than its "parent" binary EQR code. Asymptotically, both the information rate and the error control rate of these new codes are shown to be bounded away from zero so that they represent "good" codes.

The Self-dual Subclass of the Binary EQR Codes

Let Q denote the set of quadratic residues modulo a prime integer n, i.e. let $Q = \{i^2 \ (mod\ n) | i \in GF(n),\ i \neq 0\}$. Furthermore, define $q(x) = \prod_{r \in Q}(x + \alpha^r)$ where α is a primitive n-th root of unity in an extension field of $GF(2)$. The cyclic code of length n over $GF(2)$ with the generator polynomial $q(x)$ is called a binary QR code and denoted by $\mathcal{Q}$.

Let C denote the extended code of the binary QR code $\mathcal{Q}$. Since $\mathcal{Q}$ has an odd minimum distance d, C has the minimum distance $d+1$. Evidently, such an extended code C is of the form, $(n+1, \frac{n+1}{2}, d+1)$. Thus, the information rate of C is $\frac{1}{2}$.

Next, the self-dual class of the extended binary QR codes is obtained.

Definition 1 : Let $C^{\perp}$ denote the dual code of C. If $C^{\perp} = C$, the code C is called a self-dual code. Furthermore if all weights are divisible by 4, the code C is called a doubly even self-dual code.

Lemma 1 : Let C be a binary EQR code $(n+1, \frac{n+1}{2}, d+1)$. Then if $n = 8m - 1$, C is a doubly even self-dual code.

A New Class of Binary Linear Codes

A construction theorem for the new codes is given in this section.

Theorem 1: For a given binary self-dual EQR code $(n + 1, \frac{n+1}{2}, d+1)$, there exists a binary linear code (N, K, D) such that $N = n - d$, $K = \frac{n+1}{2} - d$, and $D \geq d + 1$.

An infinite class of binary linear codes is constructed by means of Theorem 1. The construction technique of these codes consists primarily of two steps. First certain $d+1$ linearly dependent columns are found in the generator matrix of a self-dual binary EQR code $(n + 1, \frac{n+1}{2}, d + 1)$. Then the columns and rows, associated with such the corresponding d marked columns of the generator matrix, are punctured and deleted. Finally, one more column is removed to leave a matrix which is the generator matrix of the new code. The parameters of several new codes which are generated by Theorem 1 are listed in Table 1. In such a table, the given distance D is found by a computer search. The result shows that the error control rate of each new code is better than that of its corresponding self-dual extended binary QR code. The information rate of each new code satisfies the inequalities $\frac{1}{4} \leq \frac{K}{N} < \frac{1}{2}$.

Asymptotic Bounds

The asymptotic bounds on the information rate and the error-control rate of the new codes are found in this section. To describe these bounds compactly, the following notation is used. First, let

(N, K, D)	Rate $\frac{K}{N}$	Rate $\frac{D}{N}$
$(4, 1, 4)$	0.2500	1.0000
$(16, 5, 8)$	0.3125	0.5000
$(24, 9, 8)$	0.3750	0.3333
$(36, 13, 12)$	0.3611	0.3333
$(60, 25, \geq 12)$	0.4167	≥ 0.2000
$(64, 25, \geq 12)$	0.3906	≥ 0.2500
$(84, 33, \geq 20)$	0.3928	≥ 0.2381
$(108, 45, \geq 20)$	0.4167	≥ 0.1852
$(132, 57, \geq 20)$	0.4318	≥ 0.1515

Table 1: A list of codes of this new class

$R(N)$ and $E(N)$ denote the information rate and the error-control rate of the new code, respectively, i.e. let $R(N) = \frac{K}{N}$ and $E(N) = \frac{D}{N}$. Next, define the limit superior and limit inferior of $E(N)$ to be, respectively,

$$\delta_u = \limsup_{M \to \infty, N > M} E(N), \quad \delta_l = \liminf_{M \to \infty, N > M} E(N). \tag{1}$$

In a similar manner for $R(N)$, define

$$R_u = \limsup_{M \to \infty, N > M} R(N), \quad R_l = \liminf_{M \to \infty, N > M} R(N). \tag{2}$$

Finally, for each real number $\delta_l \leq \delta \leq \delta_u$ let

$$\mathcal{R}(\delta) = \sup\{\liminf_{M \to \infty, N > M} \mathcal{R}(N, d_N)\} \tag{3}$$

denote the outer supremum taken over all sequences $\{d_N\}$ for which $d_N/N \to \delta$ and $\mathcal{R}(N, d_N) = N^{-1} log_2 M(N, d_N)$ where $M(N, d_N)$ is the largest possible number of codewords in a code of length N with a minimum distance of at least d_N.

The main theorem on the asymptotic bounds for the new punctured subcode is presented as follows.

Theorem 2 : Let $(N, K(N), D(N))$ denote the new punctured code developed in Theorem 1. Then

1)
$$\liminf_{M \to \infty, N > M} E(N) \geq 0.1236; \tag{4}$$

2)
$$R_u \leq 0.4382, \quad R_l \geq 0.4 \tag{5}$$

where R_u and R_l are defined in (2); and

3)
$$R(N) > \frac{1}{2} - \frac{1}{2}E(N). \tag{6}$$

Corollary 3 (Asymptotic bounds). For the new punctured code $(N, K(N), D(N))$ of Theorem 1 one has the following inequalities :

$$0.4 \leq R(N) \leq 0.4382, \tag{7}$$

$$0.1236 \leq E(N) \tag{8}$$

and

$$\frac{1}{2} - \frac{1}{2}E(N) \leq R(N) \leq \mathcal{R}(\delta) \tag{9}$$

for N sufficiently large where $\mathcal{R}(\delta)$ is defined in (3).

*This work is supported by the NSF under Grant NCR-9016340.

ON COVERING POLYNOMIALS FOR BINARY CYCLIC CODES[1]

Wonjin Sung and John T. Coffey
Electrical Engineering and Computer Science Department,
University of Michigan, Ann Arbor, MI 48109

The covering polynomial method of decoding cyclic codes is described in [1] and is essentially a modification of error trapping. While this method is a simple and effective way to decode many cyclic codes, determining the optimum set of covering polynomials for general code length n, code dimension k, and error weight τ is not an easy problem (it is Research Problem 16.12 in MacWilliams and Sloane [2]).

When the rate of a code satisfies $R < 2/\tau$, all error patterns can be trapped by monomials, and Wei [4] presented a smallest covering set for this class of codes when $\tau = 2$ or 3. We extend Wei's result to higher τ, and propose an algorithm to find optimum covering sets.

Proposition 1 : For (n, k, τ) binary cyclic codes with rate $R < 2/\tau$, the number of optimal covering monomials for *even* τ is given by

$$c = \left\lceil \frac{k - \lceil \frac{n}{\tau} \rceil + 1}{p} \right\rceil + 1, \text{ where } p = \left\lceil \frac{2n}{\tau} \right\rceil - k$$

Further, $\{ 0, x^{n-k-1+\lceil \frac{n}{\tau} \rceil}, x^{n-k-1+\lceil \frac{n}{\tau} \rceil + p}, x^{n-k-1+\lceil \frac{n}{\tau} \rceil + 2p}, \cdots, x^{n-k-1+\lceil \frac{n}{\tau} \rceil + (c-1)p} \}$ is a covering set.

Note that our choice of optimal monomials above are all located in one half of the information positions. The number of covering polynomials is very much dependent on the *interval patterns* which represent how error bits are spaced in an error vector. Two error patterns are said to have the same interval pattern if they are cyclic shifts of each other. Let $v = (v_1, v_2, \cdots, v_\tau)$ denote the interval pattern, and $v_m = \max_i v_i$. Then $\lceil \frac{n}{\tau} \rceil \le v_m \le n - (\tau - 1)$.

Proposition 2 : For (n, k, τ) binary cyclic codes with rate $R < 2/\tau$, *the following procedure gives covering monomials that are* sufficient to trap all error patterns of weight τ, and we conjecture that the set is minimal.

1. Let $z_1 = \lceil \frac{n}{\tau} \rceil, M_1(x) = x^{n-k-1+z_1}, j = 2$.
2. Assume $z_{j-1} \le v_m \le z_j - 1$, where $z_j (> z_{j-1})$ is unknown, and covering monomials $M_1(x), \cdots, M_j(x)$ are used. $M_j(x) = x^{n-k-1+z_j}$ when j is odd, $M_j(x) = x^{n-z_j}$ when j is even. Beginning from the condition $v_m \le z_j - 1$, we can get sets of upper bounds on lengths of $v_{m+1}, \cdots, v_\tau, v_1, \cdots, v_{m-1}$ that make any consecutive interval fractions (v_i, v_{i+1}) are *not* covered by any monomials used. Call the sums of upper bounds $\Sigma_1, \cdots, \Sigma_l$, where l is the number of possible combinations of interval fractions. Find the largest z_j that satisfies $\Sigma_i < n$, all i.
3. If $z_j > k$, then $\{0, M_1(x), \cdots, M_{j-1}(x)\}$ is the covering set, *size of the set* $c = j$, *stop*.
   Otherwise, let $j = j + 1$, go to step 2.

[1]Supported in part by NSF Grant NCR-9115969

For cyclic codes with rate $R < 3/\tau$, all error patterns can be trapped by binomials or monomials.

Proposition 3 : The number of binomials c required to trap every error pattern of weight $\tau = 3$ for an (n, k) code with rate $R < 1$ is given by

$$\left\lceil \frac{I(n,k,3)}{n-k} \right\rceil \le c \le \sum_{i=1}^{\min(\lfloor \frac{n}{3} \rfloor, k)} \left\lceil \frac{\min(n-3i, 2k-n+1+i)}{n-k} \right\rceil$$

where $I(n, k, \tau)$ is the number of interval patterns of length n and error weight τ, whose largest fraction is not greater than k.

The above equation gives reasonably tight bounds on the number of binomials. For example, $n = 21, k = 15$ gives $9 \le c \le 10$, $n = 21, k = 18$ gives $c = 21$, and $n = 27, k = 24$ gives $c = 36$.

Among the attractive features of the algorithm is that it can be used to decode past the guaranteed error correcting power of the code, up to complete hard-decision decoding. The method can also be extended easily to soft-decision decoding, where we find the closest soft decision codeword, subject to a maximum number of hard errors. In [3] we consider the decoding of binary cyclic codes of length 31 or less, including the decoding of error patterns of weight $t+1$ or higher. We classify the error patterns to be trapped as (1) all error patterns of weight τ (important in soft decision decoding), (2) all coset leaders of weight τ, and (3) all unique coset leaders of weight τ. Using a combination of analysis, exhaustive search, computational shortcuts, and a greedy algorithm, we determine the number of covering polynomials needed for codes of our consideration and summarize these results in tables which we do not include in this summary. Simulation of the performance of soft-decision decoding using covering polynomials was also performed to show that approximately $1.5 \sim 2.0$ dB gain is achieved at a bit error rate of 10^{-4} for soft-decision using covering polynomial that trap error pattern of weight $t+1$ or less.

References

[1] T. Kasami, "A decoding procedure for multiple-error-correcting cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 134–138, 1964.

[2] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, New York, 1977.

[3] W. Sung and J. T. Coffey, "Maximum likelihood error-trapping decoding of binary cyclic codes," *presented at the Thirtieth Annual Allerton Conference on Communication, control, and Computing, Monticello, Illinois, Sep. 30 - Oct. 2, 1992.*

[4] V. K. Wei, "An error-trapping decoder for nonbinary cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 538–541, 1984.

A NOVEL APPROACH FOR CONSTRUCTION OF ALGEBRAIC GEOMETRIC CODES FROM AFFINE PLANE CURVES

G. L. Feng and T. R. N. Rao

The Center for Advanced Computer Studies
University of Southwestern Louisiana, Lafayette, LA 70504

Abstract

The current algebraic geometric (AG) codes are based on the theory of algebraic geometric curves. In this paper, we present a novel approach for construction of AG codes without any background in algebraic geometry. Given an affine plane irreducible curve and its all rational points, based on the equation of this curve, we can find a sequence of monomial polynomials x^iy^j. Using the first r polynomials as a basis of dual code of a linear code called AG code, the designed minimum distance d of this AG code can be easily determined. For these codes a fast decoding procedure with complexity $O(n^{7/3})$, which can correct errors up to $\lfloor (d-1)/2 \rfloor$, is also shown. By this approach it is neither necessary to know the genus of curve nor find a basis of differential form. This approach can be easily understood by most engineers. Some examples are also shown, which indicate that the codes constructed by this approach are better than the current AG codes from same curves.

Summary

First we introduce a new method to determine the minimum distance bound for linear codes. Let $H \triangleq \{ \mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_r, \ldots \}$ be a sequence in F_q^n, where $\mathbf{h}_r = (h_{r1}, h_{r2}, \ldots, h_{rn})$ and let S(r) be the linear space over F_q spanned by the first r vectors of H. Let $\hat{H} \triangleq \{ \hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_u \}$ and $S(r,u)$ be the linear space over F_q spanned by the first r vectors of H and all vectors of $\hat{H}$. When $u = 0$, that means $\hat{H} = \varnothing$ and $S(r,u) = S(r)$. In this paper, we are only interested in such H and $\hat{H}$:

$$\text{for } i<j,\ i+j \le r,\ \mathbf{h}_j \notin S(i),\ \mathbf{h}_{i,j} \notin S(r-1,u),\ \text{and } \mathbf{h}_{i,j} \in S(r,u), \quad (1)$$

where $\mathbf{h}_{i,j} \triangleq (h_{i1} \cdot h_{j1}, h_{i2} \cdot h_{j2}, \ldots, h_{in} \cdot h_{jn})$. Let $\mathbf{H}_r = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_r]^T$ and $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \cdots, \hat{\mathbf{h}}_u]^T$. Then $\mathbf{H}^*_r \triangleq \begin{bmatrix} \hat{\mathbf{H}} \\ \mathbf{H}_r \end{bmatrix}$ can be a parity check matrix of a linear code C_r over F_q. When $u = 0$, $\mathbf{H}^*_r$ is reduced to $\mathbf{H}_r$.

In order to construct AG codes, we prefer to construct directly some simple $\{ H, \hat{H} \}$ for a given curve. For convenience, we restrict H and $\hat{H}$ to some special vectors, that is, $\mathbf{h}_r = (p_r(x_1,y_1), p_r(x_2,y_2), \ldots, p_r(x_n,y_n))$, where (x_i,y_i) are rational points and $p_r(x,y)$ is a monomial polynomial $x^{a_r}y^{b_r}$. For simplicity, denote $H = \{ x^{a_1}y^{b_1}, x^{a_2}y^{b_2}, \ldots, x^{a_r}y^{b_r}, \ldots \}$. In the same way, $\hat{H} = \{ x^{c_1}y^{d_1}, x^{c_2}y^{d_2}, \ldots, x^{c_u}y^{d_u} \}$. In order to construct H and $\hat{H}$ such that (1) is satisfied let us define *order* of polynomial $f(x,y)$. For each polynomial $f(x,y)$, it is associated with an integer o_f, which satisfies the following conditions:

$$o_{f+g} = o_f, \quad \text{if } o_f > o_g, \quad \text{and} \quad o_{f \cdot g} = o_f + o_g. \qquad (2)$$

For convenience, let $\underline{f}$ be o_f. We have $\underline{x^ay^b} \triangleq a \cdot \underline{x} + b \cdot \underline{y}$.

Let I be the set of the orders of polynomials in H, that is, $I \triangleq \{ \underline{x^{a_i}y^{b_i}} \mid i = 1, 2, \ldots \}$. If an integer $p \notin I$ and $0 \le p \le$ the order of last polynomial in H, p is called a *gap* of I. Let the number of all gaps of I be g^*, g^* is called the *genus* of H (or I). Let $g' = g^* + u$. Later we will see that the action of g' is as the same as g in current AG codes.

If $\{ H, \hat{H} \}$ satisfies (1), then we have:

Theorem 1: If $r \ge g^*$, the minimum distance of code C_r defined by $\begin{bmatrix} \hat{\mathbf{H}} \\ \mathbf{H}_r \end{bmatrix}$, is at least $r-g^*+1$. The value of $r-g^*+1$ is called *designed minimum distance*.

Thus, construction of good AG codes is now reduced to finding H and $\hat{H}$ for a given affine plane curve, such that **(1*) (1) is satisfied and (2*) $g^* + u$ is as small as possible**. In the following, for two classes of affine plane curves and for those curves which can be transformed into these classes of curves, we give solutions of H and $\hat{H}$, which satisfy (1*) and (2*).

Type I of Affine Plane Curves: $f(x,y) = x^a + y^b + g(x,y) = 0$, where $\gcd(a,b) = 1$ and $a > b >$ deg $g(x,y)$.

Type II of Affine Plane Curves: $f(x,y) = x^ay^c + y^{b+c} + g(x,y) = 0$, where $\gcd(a,b) = 1$ and $a+c,\ b+c >$ deg $g(x,y)$.

Example : Let $f(x,y) = x^5y^2 + y^9 + x^2 = 0$ over GF(2^5). We have $\underline{x} = 7$, $\underline{y} = 5$. By this new approach, we obtain:

$H = \{ 1, y, x, y^2, xy, x^2, y^3, xy^2, x^2y, y^4, x^3, xy^3, x^2y^2, y^5, x^3y, xy^4, x^4, x^3y^2, y^6, x^3y^2, xy^5, x^4y, x^2y^4, y^7, \ldots \}$,

$I = \{ 0, 5, 7, 10, 12, 14, 15, 17, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, \ldots \}$, and $g^* = (a-1)(b-1)/2 = 12$.

$\hat{H} = \{ x^5, x^6, x^7, x^8, x^5y, x^6y, x^7y, x^8y \}$, and $u = 8$.

Thus, we have $g' = 12 + 8 = 20$. From this example, when $r > 12$, i.e. $r + u > g'$, $d \ge r - 12 + 1 = r + u - g' + 1$, where $r + u$ is the number of check bits in C_r. But from [1, Example 6], the genus is 26.

Remark : There are many affine plane curves, which do not belong to these two types, can however be transformed into any of these types. For example, an affine Hermitian curve is $w^{r+1} + v^{r+1} + 1 = 0$ over GF(r^2). It can be transformed into $x^{r+1} - y^r - y = 0$ by $x = w \cdot \frac{s}{v-s}$ and $y = \frac{s}{v-s} - t$, where $t^r + t = s^{r+1} = -1$.

A fast decoding procedure for these AG codes can be easily realized by the fast decoding in [2-4]. The decoding procedure can correct any $\lfloor (d-1)/2 \rfloor$ or fewer errors with complexity $O(n^{7/3})$, where d is designed minimum distance determined by the above theorems.

References

[1] J. Justesen, K. J. Larsen, A. Havemose, H. E. Jensen, and T. Hóholdt, "Construction and decoding of a class of algebraic geometric codes." *IEEE Trans. on Information Theory* 35 (1989) 811-821.

[2] G. L. Feng and T. R. Rao, "Decoding algebraic-geometric codes up to the designed minimum distance." to appear in *IEEE Trans. on Information Theory* Jan. 1993.

[3] G. L. Feng, V. K. Wei, T. R. Rao, and K. K. Tzeng, "True Designed-Distance Decoding of a Class of Algebraic-Geometric Codes, Part I: A New Theory without Riemann-Roch Theorem," to appear in *IEEE Trans. on Information Theory*.

[4] G. L. Feng, V. K. Wei, T. R. Rao, and K. K. Tzeng, "True Designed-Distance Decoding of a Class of Algebraic-Geometric Codes, Part II: Fast Algorithms and Block Hankel Matrices," to appear in *IEEE Trans. on Information Theory*.

The Automorphism Groups of the Delsarte-Goethals Codes

Claude Carlet

INRIA, France

The Delsarte-Goethals codes $\mathcal{DG}(m,d)$ ($m=2t+2 \geq 4$, $2 \leq d \leq t+1$), introduced by P. Delsarte and J.-M. Goethals in [3], are generalizations of the Kerdock codes. Nonlinear and distance invariant, they are the best codes known for their parameters, and possess formal duals (cf [4], [5]).

For $d = t+1$, $\mathcal{DG}(m,d)$ is the Kerdock code $\mathcal{K}(m)$. The automorphism groups of the Kerdock codes are known (cf [1], [2], [6]). We study the automorphism groups of those Delsarte-Goethals codes which are not Kerdock codes : the $\mathcal{DG}(m,d)$ codes, with $m=2t+2 \geq 6$, $2 \leq d \leq t$.

We first recount some definitions and properties.

m' is the integer m - 1; G, G' and F denote the Galois fields of orders 2^m, $2^{m'}$ and 2 (respectively), and tr the trace function from G' to F. G'* is the set $G' \setminus \{0\}$.

The Reed-Muller code of order 1, R(1, m), is the set of all the affine forms on the F-space G. The Reed-Muller code of order 2, R(2, m), is the set of all the boolean functions f on G (ie the functions from G to F) such that the function φ_f defined by :

$$\forall (x, y) \in G^2,\ \varphi_f(x, y) = f(0) + f(x) + f(y) + f(x + y)$$

(where + denotes the addition in F) is bilinear. φ_f is called the symplectic form associated with f.

φ_f is the zero symplectic form if and only if f belongs to R(1, m). A coset of R(1, m) in R(2, m) is the set of all the boolean functions which admit the same associated symplectic form.

The weight w(f) of a boolean function on G is the size of its support: $w(f) = \left| \{x \in G / f(x) = 1\} \right|$.

A function on G is called balanced if its weight is equal to 2^{m-1}. If f belongs to R(2, m), then it is balanced if and only if its restriction to the kernel $\{x \in G / \forall y \in G,\ \varphi_f(x, y) = 0\}$ of its associated symplectic form is not constant (cf [7]).

A function in R(2, m) is bent if and only if it satisfies one of the following equivalent properties :

- its weight is $2^{m-1} \pm 2^{m/2-1}$
- its associated symplectic form φ is non-degenerate
- the sum $\sum_{x,y \in G} (-1)^{\varphi(x,y)}$ is equal to 2^m.

If f is bent anf g belongs to R(1, m), then f+g is bent.

G is identified with G' x F (as a linear space). Let $\theta(x, \varepsilon)$ be the boolean function on G defined by : $\theta(x, \varepsilon) = tr(\sum_{i=1}^{t} x^{2^i+1}) + \varepsilon\ tr(x)$

and, for any element $v = (v_1, \ldots, v_{t-d+1})$ of G'^{t-d+1} where $2 \leq d \leq t$, let $\rho_v(x, \varepsilon)$ be the function defined by : $\rho_v(x, \varepsilon) = tr(\sum_{i=1}^{t-d+1} v_i\, x^{2^i+1})$.

The function θ is bent (cf [7] p.460, th 18). Its associated symplectic form is the following function on G^2 :

$$((x, \varepsilon),(y, \eta)) \rightarrow tr(x)\, tr(y) + tr(xy) + \varepsilon\, tr(y) + \eta\, tr(x).$$

If u and u' are two distinct elements of G', then the function $\theta(ux, \varepsilon) + \theta(u'x, \varepsilon)$ is bent.

ρ_v is not bent since $\rho_v(x, \varepsilon)$ is independant from ε.

The Delsarte-Goethals code $\mathcal{DG}(m,d)$ is the non-linear subcode of R(2, m) whose elements are all the functions :

$$(x, \varepsilon) \rightarrow \theta(ux, \varepsilon) + \rho_v(x, \varepsilon) + l(x, \varepsilon)$$

where u ranges over G', v over G'^{t-d+1}, and l over R(1, m).

We denote by $\mathcal{C}(m', d)$ the linear code of length $2^{m'}$ whose elements are the functions on G' : $x \rightarrow tr(\sum_{i=1}^{t-d+1} v_i\, x^{2^i+1}) + l(x)$ where $v = (v_1, \ldots, v_{t-d+1})$ ranges over G'^{t-d+1} and l over R(1,m').

If C is a set of boolean functions on a set G, an automorphism of C is any permutation ϕ on G such that, for any f in C, $f \circ \phi$ is an element of C. More generally, we will call homomorphism of C any mapping from G to G satisfying the same condition.

We prove that the automorphism group of $\mathcal{DG}(m,d)$ is the set of all the permutations on G of the type :

$(x, \varepsilon) \rightarrow (a\, x^{2^k} + b,\ \varepsilon + \delta)$, $(a, b \in G', a \neq 0, \delta \in F, k = 0, \ldots, m-2)$.

So, it is the same as that of the Kerdock code of same length.

Previously, we need to characterize the linear homomorphisms of the code $\mathcal{C}(m', d)$ (ie those linear mappings from G' to G' which are homomorphisms of the code $\mathcal{C}(m', d)$).

THEOREM 1

The linear homomorphisms of $\mathcal{C}(m', d)$, $m' = 2t+1 \geq 5$, $2 \leq d \leq t$, are :

- *the permutations on G' : $x \rightarrow a x^{2^k}$ where a ranges over G'*, and k = 0, ..., m' - 1*
- *the functions $x \rightarrow b\, tr(ax)$ where a and b range over G'.*

THEOREM 2

The automorphisms of $\mathcal{DG}(m,d)$ ($m=2t+2 \geq 6$, $2 \leq d \leq t$) are all the permutations on G :

$$(x, \varepsilon) \rightarrow (a x^{2^k} + b,\ \varepsilon + \delta)$$
$$G' \times F \rightarrow G' \times F$$

where a ranges over G', b over G', δ over F, and k = 0, ..., m - 2.*

To achieve the proof of this theorem, we prove two lemmas :

LEMMA 1

For any element (u, u', u") of G'^3 and any element v of G'^{t-d+1} ($2 \leq d \leq t$), the functions $(x, \varepsilon) \rightarrow \theta(u x, \varepsilon) + \theta(u' x, \varepsilon) + \theta(u'' x, \varepsilon)$ and $(x, \varepsilon) \rightarrow \rho_v(x, \varepsilon)$ belong to the same coset of R(1, m) if and only if :
1) v = 0
2) u + u' + u" = 0
3) one of the elements u, u', u" is equal to 0.

LEMMA 2

Let v be any element of G'^{t-d+1} ($2 \leq d \leq t$) such that, for any non-zero element w of G', the function : $(x, \varepsilon) \rightarrow \theta(wx, \varepsilon) + \rho_v(x, \varepsilon)$ is bent. Then v is equal to 0.

COROLLARY

The automorphisms of the shortened Delsarte-Goethals code $\mathcal{DG}(m,d)$ of length $(2^m - 1)$ are all the permutations on $G' \times F \setminus \{(0, 0)\}$: $(x, \varepsilon) \rightarrow (a x^{2^k}, \varepsilon)$, where a ranges over G'* and k = 0, ..., m' - 1.*

REFERENCES

(1) E. R. Berlekamp, "Coding theory and the Mathieu groups" Inform. contr. Vol. 18, pp 40-64, 1971.
(2) C. Carlet, "The automorphism groups of the Kerdock codes", Journal of Information and Optimization Sciences, vol 12 (1991) n°3
(3) P. Delsarte, J.-M. Goethals, "Alternating bilinear forms over GF(q)", J. Combin.Theory, 19 A (1975) 26-50 [15, 21, A]
(4) J.M. Goethals, "Nonlinear codes defined by quadratic forms over GF(2)", Information and control, 31 (1976) 43-74.
(5) F. B. Hergert, "On the Delsarte-Goethals codes and their formal duals" Discrete Mathematics 83 (1990) p. 249-263, North Holland
(6) W. M. Kantor, "Spreads,translation planes, and Kerdock sets I, II", SIAM J. Alg. Disc. Math. 3 (1982), 151-165 and 308-318
(7) F. J. Mac Williams and N. J. Sloane, "The theory of error-correcting codes", Amsterdam, North Holland (1977).

ALGEBRAIC DECODING OF ZETTERBERG AND DUMER-ZINOVIEV CODES

S.M Dodunekov
Institute of Mathematics
Bulgarian Academy of Sciences
1113 Sofia, Bulgaria

J.E.M. Nilsson
Institute for Communications Technology
German Aerospace Research Establishment (DLR)
D-8031 Oberpfaffenhofen, Germany

Abstract

We consider two families of exceptionally good double-error correcting codes: the Zetterberg binary codes and the Dumer-Zinoviev quaternary codes. The Zetterberg codes are the best known family of double-error correcting binary linear codes. They are longer than the Bose-Chaudhuri-Hocquenghem double-error correcting codes of the same redundancy. The quaternary Dumer-Zinoviev codes are the only known q-ary double-error correcting codes which asymptotically meet the Hamming bound for $q > 3$.

We derive simple criteria to decide whether 1, 2 or 3 errors have occurred when one of these codes is used for data transmission. Based on these criteria new decoding algorithms are proposed, which are faster and simpler to implement than the known ones. The main improvements compared with the known algorithms are two. First, a quadratic equation only has to be solved when two errors have occurred. Secondly, some calculations, especially the inversion, can be carried out in a field considerably smaller than the ground field.

Summary

In this paper we present algebraic decoding algorithms for two classes of double-error correcting codes: the Zetterberg binary codes [1] and the Dumer-Zinoviev quaternary codes [5].

Let $n = 2^{2s} + 1, s > 1$ and let α be a primitive n-th root of unity in the finite field $GF(2^{4s})$. The Zetterberg code C_s is a binary cyclic code of length n generated by the minimal polynomial $g_s(x)$ of α over $GF(2)$. The code C_s has dimension $k = n-4s$, minimum Hamming distance 5 and covering radius 3. The Zetterberg codes are the best known family of double-error correcting binary linear codes. They are longer than the BCH double-error correcting codes of the same redundancy.

The known decoding algorithm [2] requires to solve a quadratic equation in order to decide whether 2 or 3 errors have occurred. We derive a simple criterion which makes it possible to determine in advance the number of errors and suggest a new algorithm with considerably lower time and space complexity.

Let e(x) be an error vector and denote by $S_i = e(\alpha^i)$ the syndromes. Let $Tr(\epsilon) = \epsilon + \epsilon^2 + \epsilon^{2^2} + ... + \epsilon^{2^{2s-1}}$ be the trace function from $GF(2^{2s})$ to $GF(2)$. Set $\gamma = S_1 S_{-1}$.

Lemma 1 *$\gamma = 1$ iff one error has occurred.*

Lemma 2 *$Tr(\gamma^{-1}) = 1$ iff two errors have occurred.*

Based on the above lemmas the following algorithm has been established [3][4].

Step 1. Calculate $S_1 = r(\alpha)$ and go to step 2.

Step 2. If $S_1 = 0$ then no error has occurred. Otherwise go to step 3.

Step 3. Calculate $\gamma = S_1^n$; if $\gamma = 1$ there is a single error with locator S_1. Otherwise go to step 4.

Step 4. Calculate γ^{-1} and $Tr(\gamma^{-1})$; if $Tr(\gamma^{-1}) = 1$ go to step 5. Otherwise three errors have occurred.

Step 5. Two errors. Solve the equation $\delta^2 + \delta + \gamma^{-2} = 0$ and correct two errors on positions $\alpha^j = S_1\delta^{1/2}, \alpha^i = \alpha^j + S_1$.

Notice that since $\gamma \in GF(2^{2s})$ the computation of γ^{-1} can be done in $GF(2^{2s})$. The results about decoding complexity for the codes C_2 and C_3 show that the new algorithm has considerably lower time and space complexity compared to the known one whenever 2 or 3 errors have occurred.

Similarly, we establish a new decoding algorithm for the irreducible Dumer-Zinoviev codes [6]. Notice that these are the only known q-ary double-error correcting codes which asymptotically meet the Hamming bound for $q > 3$.

References

[1] L.H. Zetterberg. *Cyclic Codes from Irreducible Polynomials for Correction of Multiple Errors.* IRE Trans.Inf.Theory, vol.8, pp.13-20, 1962.

[2] P. Källquist. *Decoding the Zetterberg Codes.* In Proceedings of: Fourth Joint Swedish-Soviet workshop on Information Theory, August 27-Sept.1 1989, Gotland, Sweden, pp.305-309.

[3] S.M. Dodunekov, J. Nilsson. *Algebraic Decoding of the Zetterberg Codes.* IEEE Trans. Inform. Theory, vol. IT-38, pp. 1570-1573, 1992.

[4] S.M. Dodunekov, J. Nilsson. *Algebraic Decoding of the Zetterberg Codes.* Internal report, Linköping University, Sweden, LiTH-ISY-I-1255.

[5] I.I. Dumer, V.A Zinoviev. *Some new maximal codes over GF(4).* Problems of Information Transmission. v.14, 1978, pp.174-181. (Translated from Problemy Peredachi Informatsii).

[6] S.M. Dodunekov, J. Nilsson, V.A. Zinoviev. *Algebraic Decoding of the Quaternary Irreducible Double-Error Correcting Dumer-Zinoviev Codes.* Internal report, Linköping University, Sweden, LiTH-ISY-I-1284.

On a Fast Decoding Algorithm for Goppa Codes Defined on Certain Algebraic Curves With at Most One Higher Order Cusp

Norifumi KAMIYA and Shinji MIURA

C&C Information Technology Research Laboratories
NEC Corporation, 4-1-1 Miyazaki
Miyamae-ku, Kawasaki-shi 216, Japan
tel 044-856-2141
fax 044-856-2235
e-mail kamiya@ibl.cl.nec.co.jp

Abstract

We propose a fast decoding algorithm for a class of geometric Goppa codes defined on certain algebraic plane curves, associated with Artin-Schreier extensions of $F_q(x)$, introduced by Stichtenoth [3]. Although we do not attempt here to treat all the class of curves introduced by Stichtenoth, we do include certain elliptic, hyperelliptic and Hermitian curves. These curves are defined by the homogeneous equation $Y^a Z^{b-a} + Y Z^{b-1} = X^b$ over an arbitrary finite field F_q of characteristic p, where a and b are relatively prime integers such that $a = p^\nu (\nu \in N^*)$, $a < b$ and the zeros of $y^a + y$ form an additive subgroup of F_q of order p^ν. The main step of the proposed algorithm is to solve a key equation studied by Porter, Shen and Pellikaan [1]. For this purpose, we derive explicit formulas for certain differential forms, which are used to construct the syndrome of the codes defined on the above-mentioned curves, and propose a modified version of Sakata's algorithm [4]. Further, we prove, in work inspired by Shen's study [2], that the Porter-Shen-Pellikaan key equation for codes defined on the curves treated here can be solved by using our modified Sakata algorithm with complexity $O(d_{des}^2 a + g^2 a)$, where d_{des} is the designed minimum distance and g is the genus of the curve. The proposed decoding algorithm may be regarded as an extension of Shen's algorithm [2] for Hermitian codes to a wider class of codes. For certain hyperelliptic codes, this algorithm can decode up to $\lfloor (d_{des} - 1)/2 \rfloor$ errors with complexity $O(n^2)$, where n is the word length of the code.

References

[1] Porter S. C., Shen B. -Z. and Pellikaan R. : "Decoding geometric Goppa codes using an extra place," to appear in IEEE Trans. Inform. Theory.

[2] Shen B. -Z. : "Codes from Hermitian curves and an iterative decoding algorithm," preprint. September, 1991.

[3] Stichtenoth H. : "Self-dual Goppa codes," Journal of Pure and Applied Algebra 55, pp.199–211, 1988

[4] Sakata S. : "Extention of the Berlekamp-Massey algorithm to N dimensions," Inform. and Comp., vol.84, pp.207–239, 1990

THE EXPANSION FACTOR OF ERROR-CONTROL CODES

Ali S. Khayrallah
Electrical Engineering Department
University Of Delaware
Newark, DE 19716

Abstract: We investigate the expansion factor of a linear block code. The expansion factor $\Delta(C)$ of an $(n,k;d)$ code with codebook C is the maximum over all generator matrices $\mathbf{G}$ whose row space is C, of the minimum of $w(\mathbf{xG}) - w(\mathbf{x})$ over all non-zero inputs $\mathbf{x}$. One can view $\Delta(C)$ as a measure of the "continuity" of the code. It indicates how well the code preserves and expands the distance relations of the input. We show that the expansion factor is bounded as $d - k \le \Delta(C) \le d - 1$. We also relate it to the weight distribution of the code, and the output length n. Finally we find the expansion factor for a number of codes, including Hamming, Equidistant, Golay, and BCH codes.

Introduction

Consider the problem of error-control coding over the q-ary symmetric channel with crossover probability $\varepsilon/(q-1)$ using linear block codes. Given an $(n,k;d)$ code with alphabet A and codebook C, let $S(C)$ be the set of all generator matrices whose row space is C. The expansion factor $\delta(\mathbf{G})$ of $\mathbf{G} \in S(C)$ is defined as

$$\delta(\mathbf{G}) \triangleq \min_{\substack{\mathbf{x} \in A^k \\ \mathbf{x} \neq \mathbf{0}}} (w(\mathbf{xG}) - w(\mathbf{x}))$$

The expansion factor $\Delta(C)$ of the code is given by

$$\Delta(C) \triangleq \max_{\mathbf{G} \in S(C)} \delta(\mathbf{G})$$

We also pick $\mathbf{G}^*$ in $S(C)$ with $\delta(\mathbf{G}^*) = \Delta(C)$ and call it the expansion matrix of C. The expansion factor indicates how well the code preserves and expands the distance relations of the input. This is important when the code is a stage in a cascade of codes, and generally whenever the input distances are meaningful. It can be helpful to think of $\Delta(C)$ as a measure of "continuity", and $\mathbf{G}^*$ as the most "continuous" generator matrix in $S(C)$. Thus given C and C' with the same parameters except for $\Delta(C) > \Delta(C')$, one should choose C and use its $\mathbf{G}^*$.

To further clarify the notion of expansion factor, consider an encoder described by a generator matrix $\mathbf{G}$ with expansion factor $\delta(\mathbf{G})$, and a pure error detection decoder. An information word $\mathbf{x}$ is mapped into codeword $\mathbf{y}$, which is sent over a noisy channel. If the channel output is a codeword $\mathbf{y}'$ with corresponding information word $\mathbf{x}'$, then

$$d(\mathbf{x},\mathbf{x}') \le d(\mathbf{y},\mathbf{y}') - \delta(\mathbf{G})$$

This suggests that for a fixed codebook C, given a choice of generator matrices, one should pick $\mathbf{G}^*$ to minimize the number of input errors caused by channel errors. An upper bound to the bit error probability as a function of $\delta(\mathbf{G})$ follows from the above inequality. The bound is similar to the one in [2].

Results

We list our theoretical results. The proofs are omitted in this summary.

Proposition 1: The expansion factor $\Delta(C)$ of an $(n,k;d)$ code C is bounded by

$$d - k \le \Delta(C) \le d - 1$$

Let the weight distribution b_i be the number of codewords of weight no greater than i, $0 \le i \le n$.

Proposition 2: The weight distribution b_i of an $(n,k;d)$ code C with expansion factor $\Delta(C) \ge \delta_2$ satisfies

$$b_i \le b_i^{\delta_2} \triangleq \begin{cases} 1 & 0 \le i \le d-1 \\ \sum_{j=0}^{i-\delta_2} \binom{k}{j} (q-1)^j & d \le i \le k+\delta_2 \\ q^k & k+\delta_2+1 \le i \le n \end{cases}$$

The next result combines Proposition 2 and the fact that the total weight of a linear code is given by $n(q-1)q^{k-1}$ (the basis of the Plotkin bound [5].)

Proposition 3: The length n of an $(n,k;d)$ code C with expansion factor $\Delta(C) \ge \delta_3$ satisfies

$$n \ge N(k;d;\delta_3) \triangleq \frac{W^{\delta_3}}{(q-1)\,q^{k-1}}$$

where

$$W^{\delta_3} \triangleq \sum_{i=d}^{k+\delta_3} (b_i^{\delta_3} - b_{i-1}^{\delta_3})\, i.$$

Proposition 2 yields an upper bound to $\Delta(C)$ in terms of the weight distribution. Proposition 3 yields a looser upper bound to $\Delta(C)$ in terms of n.

Examples

We find the expansion factor and the expansion matrix for several Hamming, Golay, and BCH codes ([4], [5].) We also discuss examples of expansion codes [1], for which the upper bound of Proposition 1 is achieved ($\Delta(C) = d - 1$), and equidistant codes [5], for which the lower bound of Proposition 1 is achieved ($\Delta(C) = k - d$.) We compare the values of $\Delta(C)$ with the estimates found from Propositions 2 and 3. For instance, the familiar (7,4;3) Hamming code has $\Delta(C) = 0$, and Proposition 2 and 3 yield the upper bounds $\delta_2 = \delta_3 = 1$. And for the (15,5;7) BCH code, $\Delta(C) = 4$, $\delta_2 = 4$, and $\delta_3 = 5$.

Related Work

The notion of expansion factor can be modified by taking into account the fact that, on normal channels, error patterns of large weight occur with very small probability. Thus one can define a bounded expansion factor, where only low weight codewords are included in the computation of $\delta(\mathbf{G})$ and $\Delta(C)$. This idea is related to the work in [3].

Bibliography

[1] A. S. Khayrallah, "Expansion channel codes: Performance bounds and examples," in Proceedings of the Conference on Information Sciences and Systems, 1992.

[2] A. S. Khayrallah, "Per-letter error probability of expansion channel codes," in Proceedings of the Allerton Conference on Communication, Control, and Computing, 1992.

[3] A. S. Khayrallah, "Bounded expansion codes for error control," in Proceedings of the joint DIMACS/IEEE workshop on coding and quantization, 1992.

[4] S. Lin and D. J. Costello, Error control coding: Fundamentals and applications, Prentice-Hall, 1983.

[5] W. W. Peterson and E. R. Weldon, Error-correcting codes, second edition, MIT Press, 1972.

A METHOD FOR COMPUTING SHOT-NOISE CUMULATIVE DISTRIBUTIONS AND DENSITIES

John A. Gubner
Department of Electrical and Computer Engineering
University of Wisconsin–Madison
Madison, WI 53706–1691
e-mail: gubner@engr.wisc.edu

Shot-noise processes, also known as filtered point processes, constitute an important class of mathematical models used to understand physical phenomena ranging from the measurement of nerve impulses in the brain, to the formation of images on film exposed under low-level illumination, to the electric current generated by photodiodes used in optical communication systems. Hence, it is unfortunate that in most cases, shot-noise densities must be obtained by using special techniques such as contour integration to numerically invert their characteristic functions. The purpose of this presentation is to suggest a new method for computing both shot-noise cumulative distributions and densities. In fact, as shown in [1], the method is quite general and can be used to recover any continuous cumulative distribution from it characteristic function without numerical integration.

Consider the real-valued process $\{Z_t\}$ given by

$$Z_t \;=\; \sum_\nu A_\nu h(t - T_\nu),$$

where the $\{T_\nu\}$ are points of a Poisson process with nonnegative intensity $\lambda(\cdot)$ and $\{A_\nu\}$ is an independent, identically distributed, nonnegative "gain" sequence. We assume that the sequences $\{A_\nu\}$ and $\{T_\nu\}$ are independent of each other. The deterministic function h is the system impulse response or point spread function, depending on the application.

Let t be fixed and set $g(\tau) \triangleq h(t-\tau)$. We now focus on the random variable

$$Y \;\triangleq\; Z_t \;=\; \sum_\nu A_\nu g(T_\nu).$$

Let $F(y)$ denote the cumulative probability distribution of Y. We assume that F is continuous everywhere except the origin, where it has a jump discontinuity of size e^{-B}. Let Γ_0 denote the measure defined by

$$\Gamma_0(C) \;=\; \int_{\{\tau: g(\tau)\in C,\, g(\tau)\neq 0\}} \lambda(\tau)\, d\tau.$$

If Γ_0 is a finite measure with density γ_0, then (i) $B = \mathsf{P}(A_\nu > 0)\cdot\Gamma_0(\mathbb{R})$, (ii) the function γ, defined by

$$\gamma(\theta) \;\triangleq\; \mathsf{E}\Big[\frac{\gamma_0(\theta/A_\nu)}{A_\nu} I_{\{A_\nu>0\}}\Big],$$

is integrable, and (iii) its Fourier transform,

$$\tilde\psi(\omega) \;\triangleq\; \int_{-\infty}^{\infty} e^{j\omega\theta}\gamma(\theta)\, d\theta,$$

is well defined. It is shown in [1] that

$$G^{\mathrm{c}}(y) \;\triangleq\; \lim_{L\to\infty} \sum_{n=-\infty}^{\infty} b_n[e^{\tilde\psi(n\pi/L)} - \tilde\psi(n\pi/L) - 1]e^{-jn\pi y/L}, \quad (1)$$

where $b_n = -j/n\pi$ for n odd, $b_0 = 1/2$, and $b_n = 0$ otherwise, is well defined and that

$$\mathsf{P}(Y > y) \;=\; \begin{cases} e^{-B}[G^{\mathrm{c}}(y) + \eta(y)], & y \ge 0, \\ e^{-B}[G^{\mathrm{c}}(y) + \eta(y) + 1], & y < 0, \end{cases}$$

where

$$\eta(y) \;\triangleq\; \int_y^\infty \gamma(\theta)\, d\theta.$$

Furthermore, if G^{c} is absolutely continuous, then the cumulative distribution $F(y) = 1 - \mathsf{P}(Y > y)$ has density

$$f(y) \;=\; e^{-B}\left[\delta(y) + \gamma(y) - \frac{d}{dy}G^{\mathrm{c}}(y)\right].$$

In many instances [1], the functions η and γ can be computed in closed form, and $\tilde\psi$ can be expressed in terms of special functions. In these cases, we only need to worry about G^{c}. We approximate (1) by taking L finite and replacing the infinite sum with a finite sum. In examples we have considered, $\tilde\psi(\omega)$ decays no slower than $1/\sqrt{\omega}$; since b_n decays like $1/n$, the terms of the series decay like $1/n^2$. The series therefore converges to a continuous function. This implies that the Gibbs phenomenon will not be present. Samples from the Fourier series are computed with a fast Fourier transform, and a cubic spline is then fit to the samples. To approximate $\frac{d}{dy}G^{\mathrm{c}}(y)$, we simply differentiate the cubic spline between its knots. The result of applying this technique to [1, Example 2] is shown below in Fig. 1.

References

[1] J. A. Gubner, "On the computation of shot-noise probability distributions," *IEEE Trans. Inform. Theory*, submitted.

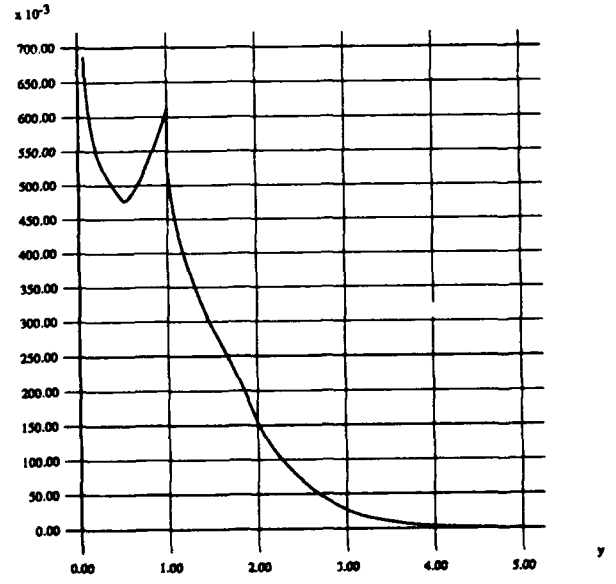


Fig. 1. Approximation of $f(y)$. Impulse at origin not shown.

Distributions and Expectations of Random Variables of Interference Type

L.L.Campbell[1], P.H.Wittke[2], A.L.McKellips[1]
[1]Department of Mathematics and Statistics
and
[2]Department of Electrical Engineering
Queen's University
Kingston, Ontario, Canada K7L 3N6

Summary

Cochannel and intersymbol interference can often be modeled as the sum of an infinite series of random variables with weights which decay rapidly. One important example is provided by interference which results from passing data through a causal filter consisting of lumped elements. For many practical applications, this model yields a random variable with a distribution which is singular (concentrated on a set of measure zero) but diffuse (non-atomic), such as the Cantor distribution. Because this model is intermediate between the case of random variables with a density function and the case of discrete random variables, there are no well-developed mathematical tools for calculating expectations. A trapezoidal rule for evaluating expectations is developed in this paper. Upper and lower bounds on expected values are also given. Finally, with a view to future applications, some of the mathematical properties of the associated distribution function are examined.

We consider here modeling interference as a random variable

$$Z = \sum_{k=1}^{\infty} \beta_k X_k$$

where $\{X_1, X_2, \ldots\}$ is a sequence of independent identically distributed random variables, each of which can take on one of M possible values, and where $\{\beta_1, \beta_2, \ldots\}$ is a known sequence, possibly vector-valued. In the applications envisaged here, X_k represents the k-th interfering information digit, while $\beta_k = h(kT)$, where $h(t)$ is the channel impulse response function and T is the sampling interval. Thus the terms in the sequence $\{\beta_k\}$ typically decay fairly rapidly, and the sum is not one to which the central limit theorem applies. Indeed, for fairly typical channel parameters, the distribution can be of Cantor type, concentrated on a nondenumerable set of Lebesgue measure zero [1]. Such distributions have neither density functions nor discrete probabilities.

For error probability calculations on communication channels, one typically wishes to calculate $E[g(Z)]$ for some smooth function g. Because of the special nature of the random variable, we cannot employ either of the two standard tools, involving an integral of a density function, or an infinite series with weights equal to the probabilities, to compute $E[g(Z)]$. It is possible to express Z as a discontinuous function defined on the unit interval $[0,1)$ and to write $E[g(Z)]$ as an integral on this interval. This integral can be approximated arbitrarily well by a trapezoidal integration rule The only condition which must be assumed on the coefficients is that the series $\Sigma\beta_k$ converge absolutely. If they satisfy the stronger condition $\beta_k = O(r^k)$ for $r<M^{-1}$ then the distribution is singular. Fairly simple but tight upper and lower bounds for $E[g(Z)]$ can also be obtained with the aid of Jensen's inequality, under mild restrictions on the coefficients and on the convexity of the function g.

Some graphs of error probability as a function of signal to noise ratio and channel bandwidth are given to illustrate the possibilities. Typically, for small bandwidth, the interference is the dominant effect, while for large bandwidth the noise dominates. For some receiver structures, there is an intermediate bandwidth at which the error probability is smallest. These results extend and improve the results of Wittke, Smith, and Campbell [1].

Because the distribution of Z is rather pathological, it would probably be useful to understand its properties better. As was mentioned above, the set of possible values of Z is frequently *nondenumerable*, but of Lebesgue measure zero. In these circumstances, the Hausdorff (fractional) dimension of this set provides a finer measure of its size. A calculation of this dimension is difficult in general, but for the case $\beta_k = O(r^k)$, for $r<M^{-1}$, the dimension is bounded above by

$$-(\log M)/(\log r).$$

This bound approaches one as $r \to 1/M$ and it approaches zero as $r \to 0$. The calculation of this dimension provides some additional insight into the significance of a result of Garsia [2] about entropy and singularity of infinite convolutions. Also, when the distribution function is singular, but continuous, its derivative can be evaluated as a generalized function (Schwartz distribution) which is neither an integrable function, nor a series of impulse functions. The properties of this derivative can also be related to the Hausdorff dimension mentioned above.

Acknowledgement

This research was supported by the Natural Sciences and Engineering Research Council of Canada through Grants A2151 and A3391, and through a scholarship.

References

[1] P.H.Wittke, W.S.Smith, and L.L.Campbell, "Infinite series of interference variables with Cantor-type distributions," IEEE Trans. Inform. Theory, vol.34, pp. 1428-1436, Nov.1988

[2] A.M.Garsia, "Entropy and singularity of infinite convolutions," Pacific. J. Math. vol.13, pp. 1159-1169, 1963.

The Asymptotic Equivalence of Investing with and without Replacement

Thomas M. Cover

Abstract and Summary

Consider the following scenarios for a sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ of price relatives corresponding to the history of a finite collection of stocks over a period of n investment periods: 1) Nothing is known about the sequence of vectors; 2) The vectors in the sequence are known, although the order is not known and the vectors are drawn independently with replacement from this set; 3) The collection of vectors is known, although the order is unknown, but the vectors are drawn without replacement from this set.

Clearly the amount of wealth that can be generated in these scenarios increases as the amount of information increases. For example, in scenario 2 one knows the empirical distribution of the market, whereas in 1, one does not. In 3, end-play can be used.

We shall argue, for bounded vector sequences, that the universal portfolio algorithm [1]

$$\hat{\mathbf{b}}_{k+1} = \frac{\int \mathbf{b} \prod_{i=1}^{k} \mathbf{b}^t \mathbf{x}_i d\mathbf{b}}{\int \prod_{i=1}^{k} \mathbf{b}^t \mathbf{x}_i d\mathbf{b}}$$

for scenario 1 will perform as well to first order in the exponent as the best algorithms in scenarios 2 and 3. Thus even end-play on a known collection of vectors of price relatives cannot outperform this universal portfolio based on no knowledge whatsoever, at least to first order in the exponent.

The growth rate of wealth in all three scenarios is given, to first order in the exponent, by the doubling rate W^* (a generalization of entropy rate), which is given by

$$W^* = \max_{\mathbf{b}} \frac{1}{n} \sum_{i=1}^{n} \log \mathbf{b}^t \mathbf{x}_i,$$

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is the sequence of vectors of price relatives for the n trading days, and the maximization is over all portfolios

$$\mathbf{b} = (b_1, b_2, \ldots, b_m), b_i \geq 0, \sum_{i=1}^{m} b_i = 1.$$

Thus the wealth

$$S_n = \prod_{i=1}^{n} \mathbf{b}_i^t \mathbf{x}_i$$

for the best portfolio algorithm $\mathbf{b}_i$ in each scenario is given by

$$S_n = 2^{nW^* + o(n)}.$$

References

[1] T. Cover. Universal Portfolios. *Mathematical Finance*, 1(1): 1-29, January 1991.

T. Cover, Stanford University, Stanford, Calif., 94305. email: cover@isl.stanford.edu. This material is based upon work supported by the National Science Foundation under Grant No. NCR-8914538-02.

STATE PRICES AND GIBBS STATES

Michael J. Stutzer
Dept. of Finance
Carlson School of Management
University of Minnesota

ABSTRACT

The foundation for the theory of asset prices in the absence of riskless arbitrage opportunities is the existence and use of normalized Arrow-Debreu state prices, also called a *risk neutral probability distribution*. Under this distribution, an asset's price is predicted to be the risklessly discounted, present value of its future payoff. The Bayesian, information theoretic view of inference directs us to use a generalized exponential distribution solving a constrained entropy problem (i.e. a Gibbs state), as an estimator of these risk-neutral probabilities. Use of the Gibbs state provides simple derivation of powerful asset pricing predictions, and also uncovers an isomorphism between statistical mechanics and asset pricing, paving the way for future development of new asset pricing predictions which exploit the isomorphism.

The paper uses only simple mathematics to make these points, and is self-contained, in the hope that it will stimulate additional interdisciplinary interest in financial economics.

SUMMARY

It is well-known that prices of contingent claims in complete and arbitrage-free securities markets can be computed using normalized Arrow-Debreu state prices, also called *risk neutral probability measures*. This paper uses simple mathematics to explore the value of a Bayesian approach, called the maximum entropy formalism (MEF), in selecting a risk neutral probability measure in situations of incomplete financial markets. The investigation is conducted within what is perhaps the simplest possible multiperiod setting, i.e. a discrete time approximation to a correlated exponential Wiener vector process. The resulting risk neutral probability measure is from an exponential family called *canonical distributions* or *Gibbs states*.

Gibbs states have a form which facilitates passing to the continuous time limit. Doing so shows that the limiting Gibbs state is parametrized by a vector of parameters which, when normalized, are the portfolio weights in the familiar mean-variance efficient *tangency portfolio* of the observed risky assets. This limiting Gibbs state is used to produce a multi-beta, approximate arbitrage pricing theory, which linearly restricts asset excess returns and covariances with N observable traded factors. The coefficient vector in the linear relation is the vector of the factors'weights in the *canonical mean-variance efficient portfolio* of the N traded factors and the riskless asset. Large deviations theory is then used to provide a frequentist rationale for the canonical distribution, which in turn is used to describe the isomorphism between the statistical mechanics of large physical systems and the arbitrage-free pricing of contingent claims in continuous time.

To illustrate the potential for exploiting this isomorphism to generate testable predictions in complex circumstances, we used it to predict the change in a country's riskless interest rate following integration of its bond market into that of another "country" (e.g. the rest of the world). The prediction is that the post-integration interest rate will be a weighted average of the countries' pre-integration interest rates, with the weighting dependent on the respective countries' tangency portfolios.

On the Optimality and Stability of Exponential Twisting in Monte Carlo Estimation [1]

John S. Sadowsky[2], Purdue University
School of Electrical Engineering, West Lafayette, IN 47907-1285

Let $P(\cdot)$ be a probability distribution on $\mathbf{R}$, and let $\mathcal{P}_P(\cdot)$ denote the i.i.d. distribution for the sequence $\{X_k\}$ with marginal $P(\cdot)$. Define $S_n = \sum_{k=1}^n X_k$ and consider the probability $p_n = \mathcal{P}_P(S_n \geq \gamma n)$ where $\gamma > \mathrm{E}_P[X_k]$. By Cramér's theorem we have $p_n \sim_{LD} e^{-\mathrm{I}(\gamma)n}$ where $\mathrm{I}(\gamma)$ is the convex conjugate of $\Lambda(\alpha) = \log(\mathrm{E}_P[e^{\alpha X}])$. The notation $a_n \sim_{LD} e^{\beta n}$ means $\lim_{n\to\infty} \log(a_n)/n = \beta$. This should not be confused with $a_n \sim b_n$, which means $\lim_{n\to\infty} a_n/b_n = 1$.

This paper considers the problem of estimating p_n via the Monte Carlo technique of *importance sampling*. Let $\mathcal{M}_P$ denote the family of all distributions $Q(\cdot)$ such that $P(\cdot) \ll Q(\cdot)$. Independent i.i.d. n-tuples $\left(X_1^{(\ell)}, .., X_n^{(\ell)}\right)$, $\ell = 1, .., L_n$, are sampled from the i.i.d. sequence distribution $\mathcal{P}_Q(\cdot)$ and applied to the *sample mean estimator*

$$\hat{p}_n \quad = \quad \frac{1}{L_n} \sum_{\ell=1}^{L_n} 1_{E_n}\left(X_1^{(\ell)}, .., X_n^{(\ell)}\right) \prod_{k=1}^{n} \frac{dP}{dQ}\left(X_k^{(\ell)}\right)$$

where $E_n = \{(x_1, .., x_n)\colon \sum_{k=1}^n x_k \geq n\gamma\}$ and $1_{E_n}(\cdot)$ is the indicator function. Write $Z_n = 1_{E_n}(X_1, .., X_n) \prod_{k=1}^n \frac{dP}{dQ}(X_k)$. Then, provided $Q(\cdot) \in \mathcal{M}_P$, we have $\mathrm{E}_Q[\hat{p}_n] = \mathrm{E}_Q[Z_n] = p_n$, which is to say that $\hat{p}_n$ is an *unbiased* estimator for p_n.

$P^{(\alpha)}(dx) = \exp(\alpha x - \Lambda(\alpha))\, P(dx)$, whenever $\Lambda(\alpha) < \infty$, is the *exponentially twisted* distribution for twisting parameter α. We show that $P^{(\theta)}(\cdot)$, where θ solves $\Lambda'(\theta) = \gamma$, has very strong *nonparametric* asymptotic optimality properties as a sampling distribution.

For a fixed integer $\nu \geq 2$ suppose that we set L_n to *stabilize the νth error moment*; that is, set L_n so that $|\mathrm{E}_Q[(\hat{p}_n - p_n)^\nu]| \sim c\, p_n^\nu$ with $0 < c < \infty$. For example, for $\nu = 2$ we set $L_n \sim v_n(Q)/\epsilon^2 p_n^2$, where $v_n(Q) = \mathrm{var}_Q[Z_n]$, in order to achieve $\mathrm{var}_Q[\hat{p}_n] \sim \epsilon^2 p_n^2$. In general, we show that stabilization of the νth error moment requires sampling cost of exponential order. Specifically,

$$L_n \quad \sim_{LD} \quad \exp(\, a_\nu(\gamma; Q)\, n\,)$$

where $a_\nu(\gamma; Q) \geq 0$ for all $Q(\cdot) \in \mathcal{M}_P$. Moreover, for all integers $\nu \geq 2$

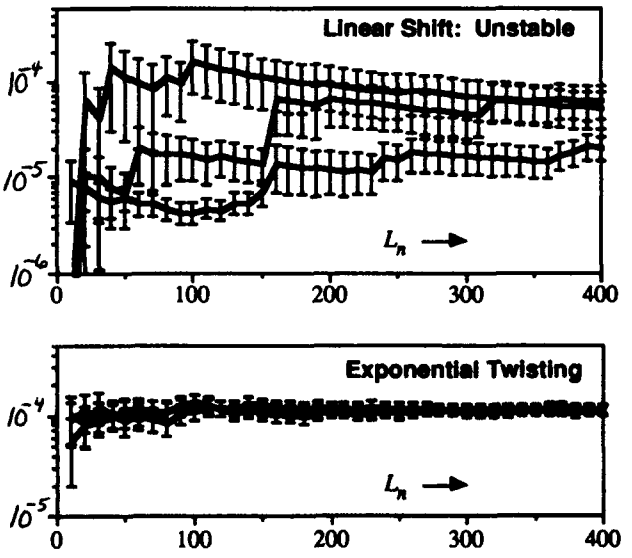
$$a_\nu(\gamma; Q) \;=\; 0 \quad \textit{if and only if} \quad Q(\cdot) \;=\; P^{(\theta)}(\cdot).$$

This extends the original work by Bucklew, Ney and Sadowsky (*J. Appl. Prob.*, March 1990) that originally obtained the result for the case $\nu = 2$.

Moreover,we show here that $P^{(\theta)}(\cdot)$ also asymptotically minimizes the error moments *sample variance estimator* in the same sense as above. This result impacts directly on the practical issue of setting $L_n \sim v_n(Q)/\epsilon^2 p_n^2$. We generally do not know either p_n or $v_n(Q)$ a priori. Both must be estimated. In practice, L_n is increased, either continuously or in batches, until $L_n \geq \hat{v}_n/(\epsilon^2 \hat{p}_n^2)$ where $\hat{v}_n$ is the sample variance. Our result that $P^{(\theta)}(\cdot)$ asymptotically minimizes the error moments of both $\hat{p}_n$ and $\hat{v}_n$ lends much credibility to this practical stopping rule.

The stability of suboptimal estimators is also addressed. Suppose that one must accept the exponential sampling cost $L_n \sim_{LD} \exp(\, a_2(\gamma; Q)\, n\,)$ of a suboptimal $Q(\cdot)$. If $a_\nu(\gamma; Q) > a_2(\gamma; Q)$ for some $\nu > 2$, then the νth error moment of sample mean and the $(\nu/2)$th error moment of sample variance will be unstable. In particular, $a_4(\gamma; Q) > a_2(\gamma; Q)$ implies that the variance of the sample variance will be unstable, and this in turn implies instability and poor convergence of the practical stopping rule. We say $Q(\cdot)$ is *completely asymptotically stable* if $a_\nu(\gamma; Q) = a_2(\gamma; Q)$ for all integers $\nu \geq 2$. A new result proved here is that the entire parametric family $\{P^{(\alpha)}(\cdot)\colon\ \alpha \geq 0\} \subset \mathcal{M}_P$ is completely asymptotically stable. This result has important practical significance because in some multidimensional applications it can be difficult to precisely determine the optimal twisting parameter vector.

A simple example illustrates the importance of the stability issue. Take $P(\cdot)$ to be the Laplacian distribution with p.d.f. $p(x) = e^{-|x+1|}/2$, fix $n = 30$ and $\gamma = 0$. Then $p_n = \mathcal{P}_P(S_{30} \geq 0) \approx 10^{-4}$. Consider two sampling distributions: the *linear shift* with p.d.f. $q(x) = e^{-|x|}/2$ and $P^{(\theta)}(\cdot)$. The figures below present numerical results of three independent runs of each estimator, plotted with empirical standard deviation error bars ($\pm\sqrt{\hat{v}_n/L_n}$). The linear shift results clearly exhibit unstable behavior. Error bars often do not overlap, and there is a tendency to underestimate p_n, in some cases, by a full order of magnitude. In contrast, observe that a single horizontal line at 10^{-4} would pierce all of the error bars of the $P^{(\theta)}(\cdot)$ estimates.



Finally, one might ask what cost $L_n \sim_{LD} \exp(\, a_\infty(\gamma; Q)\, n\,)$ is required to simultaneously stabilize all error moments (i.e., all $\nu < \infty$)? For the Laplacian example we compute $a_\infty(0; Q) = 1.226$. However, for the ordinary Monte Carlo estimator $P(\cdot) = P^{(0)}(\cdot)$, which is completely asymptotically stable, we compute $a_\infty(0; P) = a_2(0; P) = 0.226$. Thus, surprisingly, the sampling cost required to completely stabilize the linear shift estimator is substantially *bigger* than that of ordinary Monte Carlo!

[1] To appear in *IEEE Trans. on Information Theory*, Jan. 1993.
[2] This work was supported by the National Science Foundation, grant No. 9003007-NCR.

Convergence of Probability Measures for Continuous Sample Paths of Multidimensional Random Field Simulations Using Trigonometric Series

Robert Patton Leland
Department of Electrical Engineering
University of Alabama
Tuscaloosa, AL 35487-0286

We wish to simulate continuous sample paths of Gaussian random fields that are either homogeneous or have homogeneous increments, and show weak convergence of the sample path probability measures on $C[0,1]^d$.

A zero mean Gaussian random field is homogeneous if its covariance function R is shift invariant, $R(\vec{\rho}_1,\vec{\rho}_2) = E[X(\vec{\rho}_1)X(\vec{\rho}_2)] = R(\vec{\rho}_1+\vec{\rho},\vec{\rho}_2+\vec{\rho})$. In this case R has a spectral representation $R(\vec{\rho}) = \int_{R^d} \exp(i\vec{\lambda}\cdot\vec{\rho})\, d\Phi(\vec{\lambda})$. In order to ignore slowly varying effects, we can also describe the random field by its structure function $D(\vec{\rho}_1,\vec{\rho}_2) = E[(X(\vec{\rho}_1) - X(\vec{\rho}_2))^2]$. A zero mean Gaussian random field has homogeneous increments if $D(\vec{\rho}_1,\vec{\rho}_2)$ is shift invariant. Homogeneous increment random fields have the spectral representation [6]

$$D(\vec{\rho}) = 2\int_{R^{d-}} (1-\cos(\vec{\lambda}\cdot\vec{\rho}))d\Phi(\vec{\lambda}), \qquad \int_{R^{d-}} \frac{|\vec{\lambda}|^2}{1+|\vec{\lambda}|^2} d\Phi(\vec{\lambda}) < \infty$$

We simulatea zero mean homogeneous Gaussian random field $X(\vec{\rho})$, $\vec{\rho}\in \mathbf{R}^d$, by a random trigonometric series

$$X_n(\vec{\rho}) = \sum_{k=1}^{n} a_k^n \cos(\vec{\lambda}_k^n\cdot\vec{\rho}+\theta_k) \tag{1}$$

where the weights a_k and frequencies $\vec{\lambda}_k^n$ are known and $\{\theta_k\}$ is an iid sequence of random phases uniformly distributed on $[0,2\pi]$. This yields an approximate covariance function

$$R_n(\vec{\rho}) = \sum_{k=1}^{n} (a_k^n)^2 \cos(\vec{\lambda}_k^n\cdot\vec{\rho}) \tag{2}$$

For proper choices of a_k^n and $\vec{\lambda}_k^n$, R_n will approximate R, as the finite sum in (2) approximates the integral in the spectral representation of R. Such approximations were considered for $d=1$ in [3], and for random fields in [4].

We simulate a homogeneous increment random field, $X(\vec{\rho})$, as

$$X_n(\vec{\rho}) = \sum_{k=1}^{n} a_k^n\{\cos(\vec{\lambda}_k^n\cdot\vec{\rho}+\theta_k) - \cos(\theta_k)\} \tag{3}$$

The weights a_k^n and frequencies $\vec{\lambda}_k^n$ are chosen to approximate the structure function $D(\vec{\rho})$ by $D_n(\vec{\rho})$.

$$D_n(\vec{\rho}) = E[(X_n(\vec{\rho}') - X_n(\vec{\rho}'+\vec{\rho}))^2] = \sum_{k=1}^{n} (a_k^n)^2(1-\cos(\vec{\lambda}_k^n\cdot\vec{\rho}))$$

For the random fields $X_n(\vec{\rho})$ in (1) and (3), the even moments of the increments obey

$$E[(X_n(\vec{\rho}') - X_n(\vec{\rho}'+\vec{\rho}))^{2m}] \le \frac{2m!}{m!2^m} D_n(\vec{\rho})^m$$

If $R_n \to R$ (or $D_n \to D$) pointwisely, and a condition on $D(\vec{\rho})$ is satisfied, then the probability measures for X_n on the space of continuous functions $C[0,1]^d$ converge weakly to that for X. Convergence of the finite dimensional distributions does not imply the sample path measures converge weakly. Billingsley [1] gives a necessary and sufficient condition.

Theorem 1 *If the finite dimensional distributions for the random fields X_n, $P(X_n(\vec{\rho}_1)\cdots X_n(\vec{\rho}_m))$ converge to the finite dimensional distributions for a random field X and for every $\eta > 0$ there is an a and N such that $P_n(|X(0)| > a) < \eta, \forall n > N$ and for every η, $\epsilon > 0$ there is a δ and N such that $P_n(w_x(\delta) > \epsilon) < \eta$, $\forall n > N$ where $w_x(\delta) = \max_{|\vec{\rho}-\vec{\rho}'|<\delta} |X(\vec{\rho}) - X(\vec{\rho}')|$ then the corresponding probability measures P_n on $C[0,1]^d$ converge weakly to the measure P for X.*

The next theorem is our main result on weak convergence.

Theorem 2 *Let X be a homogeneous increment (possibly homogeneous) separable Gaussian random field on $[0,1]^d$ with structure function $D(\vec{\rho})$. Let X_n be the simulation for X in (1)-or (3). If $D(\vec{\rho}) \le M|\vec{\rho}|^\beta$ for some M, $\beta > 0$, and the weights and frequencies are chosen so $D_n(\vec{\rho}) \to D(\vec{\rho})$ (and $R_n(\vec{\rho}) \to R(\vec{\rho})$ for X homogeneous) pointwisely, $D_n(\vec{\rho}) \le M|\vec{\rho}|^\beta$ for all $n > N$, and $\lim_{n\to\infty} \max_k |a_k^n| = 0$, then the probability measures for X_n on $C[0,1]^d$ converge weakly to the probability measure for X, as $n \to \infty$.*

Sketch of Proof (see [2] for details)

The finite dimensional distributions for X_n converge to those of X since R_n (or D_n) converges, and the limit is jointly Gaussian by the Lindeberg-Levy version of the Central Limit Theorem. The conditions in Theorem 1 are satisfied for $D(\vec{\rho}) \le M|\vec{\rho}|^\beta$. Using a multidimensional version of the Kolmogorov Lemma, Yadrenko [5] (Theorem 2, p. 108) the condition on $w_x(\delta)$ can be demonstrated. □

For the structure function to obey $D(\vec{\rho}) \le M|\vec{\rho}|^\beta$ it is sufficient that for some $0 < \beta \le 2$, $M > 0$

$$\int_{R^{d-}} |\vec{\lambda}|^\beta \frac{|\vec{\lambda}|^2}{1+|\vec{\lambda}|^2} d\Phi(\vec{\lambda}) < \infty \tag{4}$$

To construct X_n with structure function $D_n(\vec{\rho})$ so that $D(\vec{\rho})$ and $D_n(\vec{\rho})$ are simultaneously bounded by $M|\vec{\rho}|^\beta$, we partition a sufficiently large region of a half space of R^d into small cubes, and let

$$a_k^n = (4\int_{cube_k} d\Phi(\vec{\lambda}))^{\frac{1}{2}}$$

when the cube does not touch on or contain any non-integrable singularities. The frequencies $\vec{\lambda}_k^n$ can be taken as the center points of the cubes. If there is a non integrable singularity at the origin, for cubes touching the origin, we take

$$a_k^n = (4\int_{I_k^n} \frac{|\vec{\lambda}|^2}{|\vec{\lambda}_k^n|^2} d\Phi(\vec{\lambda}))^{\frac{1}{2}}$$

For a large enough number of sufficiently small cubes, the structure function is approximated as required.

The FFT algorithm quickly generates random fields using trigonometric series if the desired sample points all lie on a rectangular grid. If this is not the case, Gaussian-Legendre quadrature yields good approximations of the integrals for $R(\vec{\rho})$ and $D(\vec{\rho})$, and can be modified to handle singularities. Gaussian-Legendre quadrature was used to simulate random fields with the Von Karman spectrum, and with $D(\vec{\rho}) = C_n^2|\vec{\rho}|^{5/3}$, $\vec{\rho}\in R^2$.

[1] Patrick Billingsley, *Weak Convergence of Measures: Applications in Probability*, Society for Industrial and Applied Mathematics, Philadelphia, 1971.

[2] R.P. Leland, 'Simulation of Continuous Sample Paths of Random Fields Using Trigonometric Series', Multidimensional Systems and Signal Processing, Vol. 2, pp. 23-43, 1991.

[3] S. O. Rice, "Mathematical Analysis of Random Noise", Bell System Technical Journal, Vol. 24, p. 46-156, January 1945.

[4] M. Shinozuka, C.-M. Jan, "Digital Simulation of Random Processes and Its Applications", Journal of Sound and Vibration, 25, pp. 111-128, 1972.

[5] M. I. Yadrenko, *Spectral Theory of Random Fields*, Optimization Software, New York, 1983.

[6] A.M. Yaglom,"Some Classes of Random Fields in N-Dimensional Space, Related to Stationary Random Processes", Theory of Probability and its Applications, Vol. II, No. 3, pp. 273-319, 1957.

WAVELET APPROXIMATION OF DETERMINISTIC AND RANDOM SIGNALS: CONVERGENCE PROPERTIES AND RATES

Stamatis Cambanis
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

Elias Masry
Department of Electrical and
Computer Engineering
University of California at San Diego
La Jolla, CA 92093-0407

Summary

Multiresolution signal decomposition and wavelet orthonormal bases of $L_2(-\infty,\infty)$ have received increasing attention in recent years in the mathematical and in the signal and image processing literatures, see [1]. A multiresolution decomposition of $L_2(-\infty,\infty)$ is an increasing sequence $\{V_f\}_{-\infty}^{\infty}$ of closed subspaces of $L_2(-\infty,\infty)$ with dense union, empty intersection, and certain translation and scaling properties [1].

The approximation of a function $f\in L_2(-\infty,\infty)$ at resolution 2^{-f} is the orthogonal projection $\hat{f}_f$ of f on V_f which is computed by using a wavelet orthonormal basis for V_f, $\{\phi_{f,k}(t)=2^{f/2}\phi(2^f t-k)\}_{k=-\infty}^{\infty}$, generated by a scale function $\phi\in L_2(-\infty,\infty)$ by means of dilations and translations. The simplest example is the Haar basis where $\phi(t)=1_{[0,1]}(t)$ has compact support and is discontinuous. There are scale functions which are k-times continuously differentiable with compact support [1].

The approximation of any $f\in L_2(-\infty,\infty)$ at resolution 2^{-f} thus has the orthonormal series representation

$$\hat{f}_f(t)=\sum_{k=-\infty}^{\infty} a_{f,k}\,\phi_{f,k}(t),$$

which converges in $L_2(-\infty,\infty)$ norm, and whose coefficients are $a_{f,k}=\int_{-\infty}^{\infty} f(t)\phi_{f,k}(t)\,dt$. The L_2 approximation error at resolution 2^{-f} is

$$e_f^2=\|f-\hat{f}_f\|_2^2=\int_{-\infty}^{\infty} f^2(t)\,dt-\sum_{k=-\infty}^{\infty} a_{f,k}^2$$

and $e_f^2\to 0$ as $f\to\infty$.

An n^{th} order asymptotic expansion for the approximation error e_f^2 as $f\to\infty$ is established in [2] for functions $f\in L_2(-\infty,\infty)$ with n derivatives. Under certain additional conditions we have

$$e_f^2=\frac{C_2}{2^{2f}}+\cdots+\frac{C_{2\lfloor n/2\rfloor}}{2^{2\lfloor n/2\rfloor f}}+o\left(\frac{1}{2^{nf}}\right)$$

where for $k\ge 1$,

$$C_{2k}=\frac{(-1)^{k+1}}{(2k)!}\int_{-\infty}^{\infty}[f^{(k)}(t)]^2\,dt\int_{-\infty}^{\infty}(u-v)^{2k}\,\phi(u)\phi(v)\,du\,dv.$$

The quality of the approximation can be improved by using a scale function with vanishing moments: if $\int_{-\infty}^{\infty}(t-\mu)^j\phi(t)\,dt=0$ for some μ and $j=1,\cdots,2(p-1)$ where $1\le p\le n/2$ then the dominant term in e_f^2 has order 2^{-2pf}.

We also consider in [2] the wavelet approximation at resolution 2^{-f} of stationary and nonstationary second-order random processes. All stationary and most nonstationary second-order processes do not have sample paths in $L_2(-\infty,\infty)$ and thus they do not fit the standard framework of $L_2(-\infty,\infty)$ wavelet representation. However, with probability one, the sample functions of mean-square continuous stationary and nonstationary random processes are square integrable over every finite interval. We therefore consider the wavelet approximation of such processes, at resolution 2^{-f}, over a finite interval, say $[0,T]$, and we use a scale function ϕ with compact support, say $[0,N]$. A substantial simplification occurs if we use a slightly larger data interval $[-N2^{-f},T+N2^{-f}]$ to compute the coefficients $\{a_{f,k}\}$. Then the approximation is

$$\hat{X}_f(t,\omega)=\sum_{k=-\infty}^{\infty} a_{f,k}(\omega)\phi_{f,k}(t),\quad 0\le t\le T,$$

where the sum is actually finite involving, for each $t\in[0,T]$, the terms with $2^f t-N\le k\le 2^f t$, and

$$a_{f,k}(\omega)=\int_{-\infty}^{\infty} X(t,\omega)\phi(t)\,dt=\frac{1}{2^{f/2}}\int_0^N X\left(\frac{s+k}{2^f},\omega\right)\phi(s)\,ds.$$

We provide an n^{th} order asymptotic expansion for the integrated mean-square approximation error,

$$e_{f,T}^2=E\int_0^T[X(t,\omega)-\hat{X}_f(t,\omega)]^2\,dt.$$

For stationary processes whose covariance function $R(\tau)=R(\tau,0)$ has n one-sided derivatives at 0, but need not be differentiable at 0, we obtain the n^{th} order asymptotic expansion, as $f\to\infty$,

$$\frac{1}{T}e_{f,T}^2=\sum_{j=1}^{n}\frac{C_j}{2^{jf}}+o\left(\frac{1}{2^{nf}}\right)$$

where

$$C_j=\frac{1}{j!}\{-R^{(j)}(0+)\}\int_0^N\int_0^N|u-v|^j\,\phi(u)\phi(v)\,du\,dv,$$

and the term $o(2^{-nf})$ does not depend on T. Thus generally the dominant term is of order 2^{-f}. When the stationary process has p quadratic-mean derivatives and the moments of ϕ of order $1,\cdots,2(p-1)$ vanish $(1\le p\le n/2)$, then the dominant term is of order 2^{-2pf}. For nonstationary processes, a similar asymptotic expansion is provided, and the dominant term is generally of order 2^{-f} (it is not clear in this case whether a scale function can be matched to a q.m. differentiable process in order to speed up the rate of convergence). This is also the case for deterministic functions which do not belong to $L_2(-\infty,\infty)$ but are square integrable over finite intervals.

For nonstationary processes with finite mean energy over the entire real line: $E\int_{-\infty}^{\infty}X^2(t,\omega)\,dt=\int_{-\infty}^{\infty}R(t,t)\,dt<\infty$ we obtain an n^{th} order asymptotic expansion for the integrated mean-square error

$$e_f^2=E\int_{-\infty}^{\infty}[X(t,\omega)-\hat{X}_f(t,\omega)]^2\,dt,$$

with properties similar to those for deterministic functions in $L_2(-\infty,\infty)$.

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[2] S. Cambanis and E. Masry, "Wavelet approximation of deterministic and random signals: convergence properties and rates," *Univ. of North Carolina, Center for Stochastic Processes Tech. Rep.* No. 352, Nov. 1991.

ON THE MINIMUM EXPECTED DURATION OF A COIN TOSSING GAME

Inchi Hu* Santosh S. Venkatesh†

ABSTRACT

The following coin tossing game is analysed: A store of N fair coins is given and it is desired to achieve M heads in a round of tosses of the coins. To allow for unfavourable sequences of tails, restarts are permitted at any epoch in the game where, in any restart, all coins are returned to store and tosses are begun anew from tabula rasa. A restart strategy is a prescription which specifies when a restart should be made. It is desired to estimate the minimum expected duration of the game over all restart strategies, and to find an optimal strategy which minimises the expected duration of the game. This simple coin tossing game, proposed by R. L. Rivest, has cryptographic roots and is linked to issues in the factoring of integers.

It is shown that there exists an optimal deterministic strategy which minimises the expected duration of the game, and a backward induction algorithm is derived which efficiently yields the optimal strategy. The properties of the optimal strategy are characterised, and some sub-optimal strategies analysed. In particular, it is shown that if the desired number of heads M is less than or equal to one-half the number of coins N in the store, then the minimum expected duration of the game grows linearly in N; if, on the other hand, M exceeds one-half N, then the minimum expected duration of the game grows exponentially fast in N.

*Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104

†Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104; *electronic mail*: venkateshee.upenn.edu

A Simulation Study of Forward Error Correction for Lost Packet Recovery in B-ISDN/ATM

Nihat Cem Oğuz
Electrical and Electronics Engineering Department
Bilkent University
Ankara, 06533, Turkey

Ender Ayanoğlu
AT&T Bell Laboratories
101 Crawfords Corner Road 4F-507
Holmdel, NJ 07733-3030, USA

ABSTRACT

We present the results of a simulation study for a virtual circuit connection over an ATM network where forward error correction is performed at both the ATM cell level and the packet data unit (PDU) level. A main conclusion of this study is that at low loads ATM cells from the same source dominate in the switch buffers, while at high loads there is a mixing of ATM cells from different sources. For the latter case, ATM cell level coding performs better, while for the former, PDU level coding performs better. The combination of the two techniques has the best overall performance.

1 Introduction

In broadband integrated services digital network using the asynchronous transfer mode (B-ISDN/ATM), the end-to-end propagation delays will typically be much larger than the duration of a packet. Consequently, retransmissions associated with the conventional error detection and *Automatic Repeat reQuest* (ARQ) mechanisms will increase the delay of a packet intolerably, especially for loss and delay sensitive high-speed applications. Therefore, one may use *Forward Error Correction* (FEC) to improve reliability without increasing end-to-end delay. In FEC, redundant information is sent along with the original data so that the receiver can avoid retransmissions by recovering lost information using this redundancy. However, there is a trade-off in using FEC: adding redundancy increases the load in the network, and in turn, the loss rate. FEC can be useful only when the former effect prevails.

In this work, we simulated a long-distance *virtual circuit* (VC) connection over an ATM network, and quantified the improvement in delay-throughput performance achieved by using FEC. In ATM, the basic unit of transport, switching, and queueing is a 53-byte *cell*. ATM cells are grouped into variable size *Packet Data Units* (PDUs) at the adaptation layer. Some of the PDUs arrive at the receiver with missing cells due to buffer overflows at congested nodes. By adding parity cells to each PDU, some of the lost cells can be recovered. A PDU is considered *lost* and retransmitted if its missing cells cannot be recovered by the FEC mechanism. Our principal motivation is the fact that the nature of the cell loss process strongly affects the performance of FEC. Adding parity cells to each PDU is effective when cells are lost "randomly." It is not effective when the cells are lost in bursts. In such cases, interleaving and buffer management techniques can be used. To combat burst losses, we employ FEC over PDUs, in addition to FEC over consecutive cells, and our results indicate this is effective. Unlike employing buffer management techniques in intermediate nodes, this method is simpler to implement as it does not require processing at the network nodes and can be employed selectively, e.g., only for delay and loss sensitive applications.

2 Forward Error Correction in ATM Networks

In the method of FEC over consecutive ATM cells, the encoder appends M_A independent parity cells to each group of N_A information-bearing cells. We consider this block of $N_A + M_A$ cells as a PDU. Since the receiver determines the positions of lost cells by means of sequence numbers, it is possible to design an *erasure channel* code so that up to M_A erased cells per PDU can be recovered. A lost PDU with more than M_A erased cells is retransmitted upon a time-out or a retransmission request from the receiver.

In ATM networks, once a node is congested, it remains in this state for some time resulting in consecutive cell losses. Increasing M_A is not a good solution to this problem since it leads to higher cell loss rates and limits the throughput. A better solution is to use a code over PDUs in addition to FEC over consecutive cells; each block of N_P PDUs is followed by M_P independent parity PDUs. We call this block of $N_P + M_P$ PDUs a *coding block*. This can be viewed as a two-dimensional code if consecutive cells are arranged in the form of a matrix each row of which contains cells of one PDU. The issues related to the construction of parity cells for both row and column coding are the same, and therefore, up to M_P lost PDUs per coding block can be recovered.

3 Simulation Results

In the simulations, we consider a five-hop VC connection over a wide geographical area. The end-to-end propagation delay is taken to be 10,240 *slots*, where a slot is defined to be the unit time needed to serve a cell at 155 Mb/sec, approximately equal to the delay of a US-wide or a Europe-wide network at this rate. In each one of the four intermediate nodes, there is a non-blocking 8 x 8 ATM switch with output buffers of capacity $B = 256$ cells. To measure the coding gain, we perform FEC on the forward traffic belonging to a *tagged* source-destination pair. While passing through the network nodes, the tagged traffic interferes with the *untagged* cells belonging to other source-destination pairs. We assume that the tagged PDUs of N_A cells arrive at the source according to a Poisson process with rate p/N_A, where p is the *network load*. $N_A + M_A$ cells of a tagged PDU are transmitted in successive slots under the control of a window flow control mechanism with PDU *permits*. This mechanism prevents the network from getting into a state of ever increasing congestion since it limits the number of PDUs on the VC. The transmitter stores each tagged PDU until an acknowledgement message (ACK) is received from the destination. In the case of receiving a negative acknowledgement message (NACK), the NACK'ed PDU is retransmitted. In real life, ACKs and NACKs flow back through a similar, possibly the same, path as the forward traffic follows, and hence, are subject to loss as well as random delay. We assume for the sake of programming simplicity that they flow back through dedicated lossless, constant-delay channels. We still maintain a timeout mechanism. Similarly, we assume independent Poisson arrivals of N_A-cell untagged PDUs with rate p/N_A at the 7 untagged input ports of each node. Each untagged PDU chooses the tagged output port independently with probability *1/8* and its cells depart from the VC at the downstream nodes independently with probability *7/8*. Also, the tagged and the untagged cells are served at the same priority level. We also assume that the receiver has a cell level memory: the successful cells of lost PDUs are stored. This feature has a strong impact on the overall network performance since it decreases the amount of work from one retransmission cycle to the next.

In the simulations, we fixed the parameters N_A and N_P as 16 and 256, respectively, and measured *the average PDU delay*, which was defined as the average time that a tagged PDU spent in the network, as a function of p. The results for the uncoded ($M_A = M_P = 0$), only cell coded ($M_A = 4$, $M_P = 0$), only PDU coded ($M_A = 0$, $M_P = 4$), and both cell and PDU coded ($M_A = M_P = 4$) cases are compared. The parameters $M_A = 4$ and $M_P = 4$ were chosen according to the results of two optimizations in which we tried $M_A \in \{0, 1, 2, 3, 4, 6, 8, 12\}$, and $M_P \in \{0, 2, 4, 6, 8, 10, 12, 14, 16, 24, 32\}$, respectively. In the PDU coded cases, the averages were computed over information-bearing PDUs so as to make a meaningful comparison with the uncoded case.

The results show the trade-off between using cell coding and PDU coding. For low p, where cells of a single connection dominate in the output buffers, losses occur in rare bursts for buffer capacities as large as 256, and hence, PDU coding outperforms cell coding. For high p, the frequency of burst losses increases, and many cells from distinct connections interfere at the switch outputs resulting in random losses. Therefore, cell coding starts to perform better as p increases. The joint code outperforms only cell coding or only PDU coding for almost all p, except for a small degradation around $p = 0.45$, which is due to the individual performance degradation in cell coding. The results of the optimizations over M_A and M_P, and the details of transmitter and receiver implementations are available from the authors.

Finally, note that if the successful cells of the lost PDUs were not stored at the destination, the average PDU delays would be much higher for high network loads. With a cell level memory at the destination, the lost PDUs have to arrive at the receiver with fewer and fewer number of cells in successive retransmission cycles. However, when there is no such memory, the PDUs are subject to the same probability of loss in successive transmission cycles.

4 Summary and Conclusions

We have presented the results of a simulation study, showing that the use of forward error correction improves the performance of broadband networks. We have concentrated on the performance over a virtual circuit connection over an ATM network. The FEC technique is based on transmitting parity packets, which are constructed by using an erasure channel code, along with information-bearing packets. Although this may increase the network load leading to higher packet loss rates and limit the network throughput, retransmissions are avoided provided that sufficiently many packets reach the destination. In particular, we have considered two types of coding: coding over consecutive ATM cells and coding over consecutive fixed-length PDUs. The simulation results obtained have confirmed our a priori expectation that coding over PDUs would be effective for burst cell losses, and indicated that, by using FEC with correct parameters, it is possible to reduce the average PDU delays approximately to the extent of a half.

A Robust Error Control System For Broadcast Channels

S. Ram Chandran, TeleSciences, Fremont, CA. Shu Lin, Univ. of Hawaii.

Abstract

We have proposed a robust error-control system for a broadcast environment. The system is a combination of FEC and ARQ schemes. We use a cascaded coding scheme with a binary inner code and an interleaved non-binary outer code. The decoding policy is chosen so that we make optimum use of outer code's error correcting capability. The system also has a parity retransmission feature. We use a dynamic programming optimization method to optimize the throughput. We show that this system achieves very high reliability and high channel utilization even at extremely high bit-error-rates ($\leq 10^{-2}$). The scheme is suitable for high speed data transfer even when the channels are noisy.

1 Summary

In this paper we present a robust error-control system for broadcast channels. The communication environment for the proposed scheme consists of a single transmitter broadcasting messages to R receivers. As a special case, this system can be used for point-to-point communications also. The goal of this system is to facilitate high speed data transfer even when the channels are very noisy. The coding scheme used here is similar to the work of Kasami et. al. [1]. We have modified their cascaded coding scheme and also added a parity retransmission feature to it. The coding scheme is obtained by cascading two error-correcting codes: the inner code and the outer code. The inner code is a binary code designed for simultaneous error correction and detection. The outer code is obtained by interleaving a non-binary code with symbols from Galois Field $GF(2^l)$. This code is designed for correcting symbol errors and erasures. The errors handled by this code are either caused by the channel or the inner code decoder, whereas the erasures are introduced by the inner code decoder only. The interleaving facilitates burst-error correction.

In addition, we also have a parity retransmission feature for the recovery of incorrectly received messages. The retransmission scheme is a combination of *type-1* and *type-2* hybrid ARQ schemes [3]. We use a selective-repeat mode of retransmission and the protocol is similar to the one proposed by Chandran and Lin [2]. In this retransmission protocol, the number of copies of a message at any given stage of retransmission is chosen to optimize the throughput, using dynamic programming optimization. This optimization scheme takes into account the number of previous transmissions of the message and also the number of receivers that are yet to acknowledge the message, thereby allowing us to achieve the maximum possible throughput.

Parity blocks for retransmissions are formed based on the original data block and a half rate *invertible* code. This code is obtained by shortening the inner code. Therefore, we can use the inner-code encoder and decoder for encoding and decoding of this code. Moreover, this code has the same error-correcting capability as the inner code but is used on a much smaller portion of the data block therefore it is very powerful. The data and parity transmissions both use the same format hence their decoding procedures are very similar. The original message can be recovered from the parity block also, by inversion, if the decoding is successful. Otherwise, the parity and data blocks are combined for further error correction using the half-rate code. If the parity block fails to recover the original message by this combined decoding, then another retransmission is requested. The next retransmission is a data block. Thus the retransmissions are alternate repetitions of data and parity blocks.

The key idea of this coding scheme is that the decoding information is passed on from inner-code decoder to the outer-code decoder [1]. Data and parity blocks consist of codewords from inner and outer codes arranged in an array. Decoding process of these blocks consist of inner decoding followed by outer decoding on each of these codewords. The decoding policy is chosen in such a way that the error correcting capability of the outer code is utilized optimally.

We provide a complete analysis of the performance of this coding scheme. We show that even with very simple inner and outer codes, we can achieve a probability of block decoding error of 10^{-10} at extremely high bit-error-rates ($\leq 10^{-2}$). At lower bit-error-rates, the achievable probabilities of block decoding error are negligibly small. Moreover, such high reliability is achieved at very acceptable levels of channel utilization. The corresponding probabilities of decoding failure (retransmission) is small. Using the proposed scheme, we can achieve up to 20% channel utilization with a hundred receivers at a Bit-Error-Rate= 10^{-2}.

References

[1] T. Kasami, T. Fujiwara, T. Takata and S. Lin, "A Cascaded Coding Scheme for Error Control and It's Performance Analysis," IEEE Trans. Information Theory, vol. 34, No. 3, pp. 448-462, May 1988.

[2] S. Ram Chandran and S. Lin, "Selective Repeat ARQ Schemes for Broadcast Links," IEEE Trans. Comm., vol. 40, No. 1, January 1992, pp. 12-19.

[3] S. Lin and D. J. Costello, Jr., "Error Control Coding: Fundamentals and Applications," Prentice-Hall, 1983.

THE CAPACITY PER CHANNEL OF A BROAD CLASS OF NOISEFREE CDMA IS FOR VERY DIVERSE TASKS CLOSE TO $1/e$ UNDER VERY LOOSE CONSTRAINTS

Sándor Csibi

Dept. of Telecom., Tech. Univ. of Budapest, Stoczek u 2, Budapest, H-1111; e-mail: h179csi@ella.hu

The capacity (i.e., the max. total throughput) per channel of code division multiple access with superimposed codes as hop sequence sets is considered. The access to a channel is controlled slot-by-slot by a binary sequence for time hopping. One out of Q hop frequencies is selected slot-by-slot by the symbols of a Q-ary sequence in the considered (simplest) version of frequency hopping. The number of (single server) channels $K = 1$ for time, and $K = Q > 1$ for frequency hopping. The length of the hop sequence (the frame length) in bits is denoted by N in the binary as well as in the Q-ary case. m bits are conveyed during an active slot. Random delay, erasures, and erasure correction are assumed, but no independent additive noise.

Task **A**: z bits per frame from at most M window active sources have to be served without error out of T potential sources. Task **B**: a Poissonian source population has to be served, with a demand rate λ, each source having at most once a packet of y bits to send next to a demand ($z = y/\nu$ bits are conveyed per frame). For simplicity for Task **B** the same single hop sequence is assumed at each potential source, and the ϵ-capacity per channel $C(\epsilon)$ (with a probability $1 - \epsilon$, and error free transmission up to M window active sources) is considered.

It was shown by Massey [1982], Bassalygo and Pinsker [1983], Tsybakov and Likhanov [1983], Massey and Mathys [1985], for the basic Task **A**, that $C \to 1/e$ and by Csibi [1991], for task **B** that $C(\epsilon) \to 1/e$, and $\epsilon \to 0$, as $z \to \infty$. under distinct constraints, all for time hopping.

It can be shown that for investigating the joint possibilities of separation and errorfree decoding the disjunction of all possible cyclic shifts of the binary (resp. that of the binary representation of the Q-ary) hop sequences, modulo N, is of our interest. Accordingly, the capacity study may be formulated, in the noisleless case, as an extremal additive set problem, with disjunction as addition. (For the basic formulation and a basic bibliography of such problems see, e.g., a survey by Sós [1989].)

It can be expected already from the aforementioned studies that while the additive set models of our interest for Tasks **A** and **B** (time as well as frequency hopping, slot synchronous as well as slot asynchronous arrivals) are under general circumstances essentially distinct, one may get close to the very same capacity limit $1/e$ under loose (however more or less individually tailored) conditions. One of our purposes, in this paper, is to prove that this is really the case. Another aim is to show what a peculiar interrelation holds among the additive families underlying the various distinct versions of the multiple access problems considered; and try to point out, also qualitatively, under what kind of constraints (on the multi-user objectives) can the capacity be kept close to $1/e$.

Transparent upper and lower bounds are given on C (resp., $C(\epsilon)$), for the aforementioned eight problems, under distinct constraints on M, m and ν (and also on T, for Task **A**), w.r.t. z. Both quantities are close to $1/e$, for an approriately large value of z (e.g., $z \geq 100$ bits), and approach to $1/e$ (with $\epsilon \to 0$ for $C(\epsilon)$) as $z \to \infty$.

As a matter of fact the capacity (resp., ϵ-capacity) per channel may be kept close to the limit $1/e$ for very distinct classes of the considered joint separation and transmission problems if the objective is an essential point-to-point message transmission together with joint frame separation (and far not just separation). This feature is consistent with the fact that the families of the extremal sets corresponding to the aforementioned eight distinct extremal set problems (with points defined by a tradeoff between the resources for joint frame separation and point-to-point-to-point transmission, and with a shortest hop sequence of length N') have got a peculiar joint structure, resembling to the corolla of a petaled flower. Viz., the capacity (resp. ϵ-capacity) sets for these eight distinct problems have got a common limit point $1/e$: $C \to 1/e$ (resp. $C(\epsilon) \to 1/e$ with $\epsilon \to 0$) as $z \to \infty$.

The constraints for Tasks **A** as well as **B** are met, e.g., by the RS hop sequence constructions due to A, Györfi, and Massey [1992] of actual interest. Thus one may get by zMe/N a good idea about the absolute efficiency $\eta = N'/N$ of such (and also of other) well implementable constructions.

ON THE DELAY IN A MULTIPLE ACCESS SYSTEM WITH LARGE PROPAGATION DELAY

BRUCE HAJEK

Coordinated Science Laboratory and the
Department of Electrical and Computer Engineering
University of Illinois, Urbana, Illinois 61801, USA

N. B. LIKHANOV and B. S. TSYBAKOV

Institute for Problems in Inforation Transmission
19 Emolovoi Street, GSP-4
Moscow, Russia 101447

SUMMARY

Recent advances in optical technology have made it possible to transmit at very high data rates. Consequently, the propagation delay for a packet of information is long compared to the length of a packet. For example, consider a wide area network with a single star topology, such that the stations are located 50 kilometers from the hub, packets are 1000 bits long and the transmission rate is a gigabit per second. The propagation delay from one station to another is dictated by the speed of light in glass. The delay is about 500 microseconds, roughly 500 times as long as the transmission time of one packet. In contrast, classic protocols such as the ALOHA protocol were investigated with a propagation delay roughly 12 times the packet length in mind.

The particular model discussed in this paper is now described. Newly generated packets arrive according to a Poisson process with rate λ. Time is divided into slots of unit length, where time is normalized so that one packet can be transmitted in one slot. We denote by slot i the time interval $[i, i+1)$. Those packets with generation times in the set B_i are transmitted during slot i. We require that $B_i \subset [0, i)$, for a packet can't be transmitted until the first full slot after it arrives. The outcome of slot i, denoted by $\theta_i = \theta(B_i)$, satisfies $\theta_i \in \{0, 1, 2\}$. If no packets are transmitted in slot i, then $\theta_i = 0$. If one packet is transmitted in slot i, then the packet transmission will be successful and $\theta_i = 1$. If two or more packets are transmitted in slot i, then the packets will collide and the transmission will not be successful.

There are two, often the same, propagation delays associated with the model—the propagation delay of feedback and the propagation delay in the forward channel. The propagation delay of feedback is denoted by the positive integer N. The outcome θ_i is assumed to be announced to all stations by time $i+N$. Thus, we require B_{i+N} to be a function of $(\theta_0, \theta_1, \cdots, \theta_i)$. The usual model, in which the outcome of slot i is known by the beginning of slot $i+1$, corresponds to $N = 1$. We define the transmission delay of a packet to be the number of whole slots that elapse between the time the packet is generated until the beginning of the slot in which the packet is first successfully transmitted. With this definition, the delay is a nonnegative integer value, and it does not include the forward propagation delay.

By finding a lower bound on the probability that a typical packet will be successfully transmitted within the first $N/2$ slots, we also find a lower bound on the mean delay suffered by a typical packet.

Proposition 1 *Under any random access algorithm,*

$$P(\frac{N}{2}) \geq (2^{-\ln 2})^{\frac{1}{2}} > (.618)^{\frac{1}{2}}, \tag{1}$$

where $P(\frac{N}{2})$ is probability that a typical packet suffers delay $\frac{N}{2}$ or more. In particular, the mean delay suffered by a typical packet is at least $0.5N(.618)^{\frac{1}{2}}$.

An upper bound on the achievable mean delay of a typical packet is obtained by considering specific random access algorithms. Given $k \geq 1$ let $\lambda_{\max}(k) = \max\{G[1-(1-\exp(-kG))^k] : G \geq 0\}$. Given λ with $0 < \lambda < \lambda_{\max}(k)$, let G_o be the minimum positive solution to the equation

$$\lambda = G_o[1-(1-\exp(-kG_o))^k]. \tag{2}$$

Finally, let $\gamma_o = (1-\exp(-kG_o))^k$ and $d_o(k, \lambda) = \gamma_o/(1-\gamma_o)$.

Proposition 2 *There exists a family of random access algorithms parameterized by k, λ, N so that if $D(k, \lambda, N)$ is the average delay (exclusive of the forward propagation delay) of a typical packet, then $\lim_{N\to\infty} D(k, \lambda, N)/N = d_o(k, \lambda)$.*

Table 1: Comparison of lower and upper bound on the coeficient of N in the mean delay, for some values of λ and optimal values of k.

λ	lower bound $0.5(2^{-\ln 2})^{\frac{1}{2}}$	k^*	upper bound $d_o(k^*, \lambda)$
0.05	0.0000336	16	0.0000716
0.10	0.00410	7	0.00861
0.15	0.0203	5	0.0506
0.20	0.0452	3	0.138
0.25	0.0731	2	0.293
0.30	0.101	1	0.631
0.35	0.127	1	1.05

ACKNOWLEDGEMENT

The effort of B. Hajek was supported by the National Science Foundation under National Science Foundation Contract NCR 90-04355.

CONSTRUCTIONS OF PROTOCOL SEQUENCES FOR MULTIPLE ACCESS COLLISION CHANNEL

László Györfi and István Vajda*

A, Györfi and Massey [1] have given a general way to construct constant-weight cyclically permutable codes. A cyclically permutable code CPC(N, T, d_c) is a binary block code with block length N, size T and positive cyclic minimum distance d_c. The cyclic minimum distance d_c of a code is defined as the minimum Hamming distance from a codeword to its own cyclic shifts or to some cyclic shift of another codeword.

Let α be a primitive element of $GF(p^r)$, where p is a prime number and $r \geq 1$. A primitive BCH code V of length $n = p^r - 1$ is then defined by the parity-check polynomial $h(x) = \text{l.c.m.}\{M_0(x), M_1(x), \ldots, M_{k-1}(x)\}$, where $M_j(x)$ is the minimal polynomial of α^j over $GF(p)$, and $3 \leq k < p-1$, $j = 0, 1, \ldots, k-1$. V is given by the direct sum

$$V = V_0 + V_1 + V_2 + \ldots + V_{k-1},$$

where V_j is the code over $GF(p)$ of length n with parity check polynomial $M_j(x)$, $j = 0, 1, \ldots, k-1$. Because $M_1(x)$ is primitive polynomial, V_1 contains an m-sequence $\mathbf{c}^*$.

Consider the following subcode of V:

$$\hat{V} = \{\mathbf{c}^*\} + V_2 + \ldots + V_{k-1}.$$

If the pulse-position-modulation (PPM) code consists of all weight-one sequences of length p, then let B^* be the cyclic concatenation of $\hat{V}$ and the PPM code, defined in [1]. It is shown that B^* is a binary constant-weight cyclically-permutable code with length $p(p^r - 1)$, size $p^{(k-2)r}$, cyclic minimum distance $d_c \geq 2(p^r - 1 - (k-1)p^{r-1})$.

The set $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_T\}$ of binary sequences is said to be a (T, M, N, σ) protocol sequence set if these sequences all have length N and, when used as protocol sequences for multiple access collision channel without feedback, have the property that each active user can be identified by the receiver, the receiver can synchronize and each active user achieves at least σ successful packet transmissions during the protocol sequence length, provided that at most M out of the T users are active. For any integer σ with $1 \leq \sigma \leq w$, a binary constant-weight$-w$ cyclically-permutable code CPC(N, T, d_c) is a (T, M, N, σ) protocol-sequence set for

$$M = \min\left\{T, \left\lfloor \frac{w-1}{w - d_c/2} \right\rfloor, \left\lfloor \frac{w-d}{w - d_c/2} \right\rfloor + 1\right\}$$

where $\lfloor . \rfloor$ denotes rounding down to the nearest integer ([1]).

If the total information transmission rate R_{sum} is defined by

$$R_{sum} = \frac{M\sigma}{N} \, (packets/slots).$$

and the code B^* is used as a (T, M, N, σ) protocol sequence set then the parameters are as follows:

$$T = p^{(k-2)r}, \; N = p(p^r - 1),$$

$$M \geq \frac{w - \sigma}{(k-1)p^{r-1}}, \; R_{sum} \geq \frac{\sigma(w-\sigma)}{N(k-1)p^{r-1}},$$

the maximum of which is obtained for $\sigma = w/2$ under the condition $w/2 - 1 \geq (k-1)p^{r-1}$. Choosing $\sigma = w/2$, we get

$$R_{sum} \approx \frac{1}{4(k-1)}$$

for large p. The ratio of the total population T to the block length N is

$$\frac{p^{(k-2)r}}{p(p^r - 1)}.$$

For $k = 3$, this ratio is $\approx p^{-1}$. For fixed $k > 3$, this ratio is a monotone increasing function of r and is $\approx p^{(k-3)r-1}$.

[1] N. Q. A, L. Györfi, J. L. Massey "Constructions of binary constant-weight cyclic codes and cyclically permutable codes" *IEEE Trans. on Information Theory*, vol. 38, pp. 490-499, May 1992

*Technical University of Budapest, Stoczek u. 2, H-1521 Budapest, Hungary

CERTAIN GENERALIZATIONS ON THE COLLISION CHANNEL WITHOUT FEEDBACK

Thomas J. Ketseoglou

SIEMENS AG

SYSTEMS ENGINEERING

P.O. BOX 700073

D-8000 MUNICH 70

GERMANY

ABSTRACT

In this paper, certain generalizations on the collision channel without feedback are presented, based on the original work of Massey and Mathys. We are concerned with situations in which, given a collision in a slot, the channel capacity is a non-zero, decreasing function of the number of users involved. This corresponds, for example, to spread-spectrum type of signaling. Due to this model, concatenated coding schemes are employed to efficiently exploit the time-varying nature of the channel, upon using reliability information from the inner code decoding process, at the outer code decoding process. Results concerning capacity and coding/decoding tradeoffs will be presented.

SUMMARY

The collision channel without feedback, was thoroughly investigated by Massey and Mathys in [1]. Massey and Mathys showed that, the asymptotic throughput of this channel, as the number of users tends to infinity, approaches e ^ -1, independently of the kind of network operation (synchronous or asynchronous). In their constructive proof, [1], they presented specific protocol sequences, and maximum-erasure-burst-correcting (MEBC) codes to achieve this maximum throughput.

In this paper, we attempt certain generalizations on [1]. Our main scope is to consider systems in which, collisions do not, in general, lead to a totally useless channel of zero capacity. In this case, the channel capacity available during a collision, is a function of the number of interfering users during the collision. This model applies, for example, in Code Division Multiple Access (CDMA) systems, and in capture systems, in which, there are power variations in the received signal powers of different users.

We show that, the protocol sequence construction presented in [1], can be adopted effectively by our model, to provide efficient channel accessing, independently of time shifts between users signals. Due to the time-varying, non-zero channel capacity during a collision, more complex coding schemes, like for example, concatenated schemes with generalized minimum distance decoding [4] will be considered. This is dictated by the fact that packets are not fully destroyed in a collision. Thus, an inner code will output reliability information for decoding the (outer code) superpackets. In fact, due to the "softer" collisions in our case, the overall coding/decoding problem becomes more complex.

We will study three scenarios suitable for application of the above mentioned model: Frequency-Hopping Spread-Spectrum (FH/SS), Direct-Sequence Spread-Spectrum (DS/SS), and unspread signaling in which, power variations in the arriving signals arise due to path loss and fading.

Finally, because the code alphabet size required for achieving capacity in this generalized version of collision channel is very high, we examine a potential application of binary codes together with interleaving [5], as a means of achieving "realistic" but non-optimum throughput in the channel.

REFERENCES

[1] J. L. Massey and P. Mathys, "The Collision Channel Without Feedback", *IEEE Transactions on Information Theory*, Vol. IT-31, No. 2, pp. 192-204, March 1985.

[2] M. B. Pursley, " Frequency-Hop Transmission for Satellite Packet Switching and Terrestrial Packet Radio Networks", *IEEE Transactions on Information Theory*, Vol. IT-32, pp. 652-667, September 1986.

[3] J. N. Hui, "Throughput Analysis for Code Division Multiple Accessing of the Spread Spectrum Channel", *IEEE Journal on Selected Areas in Communications*, Vol. SAC-2, No. 4, pp. 482-486, July 1984.

[4] G. D. Forney, "Concatenated Codes", Cambridge, MA: MIT research monograph: No. 37, MIT PRESS 1966.

[5] S. Laufer, and J. Snyders, "Feedforward Multiple Access Satellite Communications", *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 6, pp. 1003-1011, August 1992.

CAPACITY AND CODING FOR T ACTIVE USERS OUT OF M ON THE COLLISION CHANNEL

Brian Hughes
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, Maryland 21218

Abstract

The problem of designing codes for M users that permit any $T < M$ users to transmit at the same time is investigated for the collision channel. Twelve communication problems are considered that vary according to the degree of synchronisation among users, the receiver's knowledge of the active users, and the desired reliability of the code. For each problem, the T-of-M user capacity region is determined and constructive coding schemes that approach any rate in this region are presented. Applications to random access communications are discussed.

Summary

The *collision channel* without feedback models the communication of many transmitters with a common receiver through a shared packet broadcasting channel. Massey and Mathys [1] determined the T-user capacity regions of this channel for asynchronous and slot-synchronous users, and also gave constructive codes that approach all rates in these regions.

This paper considers the design of codes for M users that permit any subcollection of up to T of the M to transmit at the same time. Twelve communication problems are posed that vary according to whether the users are *synchronous, slot-synchronous,* or *asynchronous*; the active users are *known* or *unknown* in advance to the receiver, and the error is desired to be *zero-error* or *arbitrarily small error.*

Theorem 1: For the T-of-M user collision channel, all slot-synchronous and asynchronous capacity regions coincide. This common capacity region consists of all $\mathrm{R} = (R_1, \ldots, R_M)$ such that

$$R_i \le \min_{j_1,\ldots,j_{T-1} \neq i} q_i \, (1 - q_{j_1}) \cdots (1 - q_{j_{T-1}}) \, ,$$

for some $\mathrm{q} = (q_1, \ldots, q_M)$ satisfying $0 \le q_i \le 1$, $q_{[1]} + \cdots + q_{[T]} = 1$, $q_{[k]} = q_{[T]}$ for $T \le k \le M$, and where $q_{[i]}$ is q_i arranged in decreasing order. ◇

Remark: The slot-synchronous, known user, capacity region was obtained earlier in [2].

For the T-of-M capacity region $\Re$, the *symmetric capacity* is $C_{sym} = \sup\{ \, r \, : \, (r/T, \ldots, r/T) \in \Re \, \}$.

Supported in part by ARO Grant DAAL03-89-K-0130.

Corollary: The T-of-M user symmetric capacity is

$$C_{sym} = (\, 1 - 1/T \,)^{T-1} \quad \text{packets/slot,}$$

regardless of whether the users are slot-synchronous or asynchronous and whether or not the active users are known in advance to the receiver, and whether for arbitrarily small error or zero-error. ◇

Since C_{sym} does not depend on M, adding users does not reduce capacity.

Theorem 2: For the T-of-M user collision channel, all synchronous capacity regions coincide. This common capacity region is the set of all $\mathrm{R} = (R_1, \ldots, R_M)$ that satisfy

$$R_i \le \min_{j_1,\ldots,j_{T-1} \neq i} \mathrm{E}\{ \, Z_i (1 - Z_{j_1}) \cdots (1 - Z_{j_{T-1}}) \, \}$$

for some binary random variables $Z_1, \ldots, Z_M$. ◇

Theorem 3: In the synchronous case, the T-of-M symmetric capacity is

$$C_{sym} = T \binom{M-T}{K-1} \binom{M}{K}^{-1} \quad \text{packets/slot,}$$

where $K = \lceil (M+1)/T \rceil - 1$, regardless of whether or not the receiver knows in advance the set of active users, and regardless of whether small error or zero error is desired. ◇

Remark: Symmetric, uncoded TDMA achieves a symmetric rate of T/M. This is optimal if and only if $T \le M < 2T$.

For each capacity region, constructive codes that approach all rates in these regions are given. Applications to random-access communications will be discussed.

References

[1] J. L. Massey and P. Mathys, "The collision channel without feedback," *IEEE Transactions on Information Theory*, IT-31 (2), pp. 192-204, March 1985.

[2] B. S. Tsybakov and N. B. Likhanov, "Packet switching in a channel without feedback," *Problemy Peredachi Informatsii*, vol. 19 (2), pp. 69-84, April-June 1983.

A Model for the Approximation of Interacting Queues that Arise in Multiple Access Schemes

Eytan Modiano and Anthony Ephremides
Electrical Engineering Dept.
University of Maryland
College Park, MD

In this paper we present a new approximate model for the analysis of systems of interacting queues which often arise in multiple access network protocols. This new model is a refinement of an existing model developed in [1] for the ALOHA multiple access protocol. We begin by applying this model to the analysis of a multiple-node broadcast algorithm for a mesh network, which was presented in [2]. We then show how our model can be used to study the performance of the ALOHA multiple access protocol.

A multiple-node broadcast is a common task in the execution of parallel algorithms in a network of processors, where every processor may have a message to be broadcast to all other processors. In [2] an algorithm was developed which performs periodic, synchronized, broadcast cycles, where during each cycle only a small number of nodes are allowed to broadcast their message. Consider an N by N mesh, where each node has exogenous packets arriving (to be broadcast) independently according to a Poisson random process and placed in infinite-capacity queues. Our broadcast algorithm works as follows: We partition the mesh into N vertical rings, such that each node belongs to exactly one ring. At the beginning of every broadcast cycle each ring selects, at random, up to d packets to be broadcast throughout the mesh. The broadcast of the d packets from each ring is performed and has a fixed duration of $(d+1)(N-1)$ time slots. Clearly, the queues at the N nodes on each ring are highly dependent on each other. In fact, the queue sizes of the N nodes on each ring form an N-dimensional infinite Markov chain. Obtaining analytic expressions for the steady-state behavior of such a system is very difficult. Even a numerical evaluation of such systems can be computationally prohibitive. A similar difficulty arises in the analysis of the Aloha multiple access protocol and no exact analysis for packet delay is known, for that case either. Several approximate models have been proposed for the analysis of ALOHA which may be useful in analyzing our system.

In [1], Ephremides and Saadawi developed an approximate model for a system of interacting queues for analyzing the ALOHA protocol. In their model they approximate a system of N infinite queues as a single dimensional infinite Markov chain representing the state of one user together with an N-dimensional finite Markov chain representing the state of the rest of the system. They use parameters from the solution of one chain in analyzing the other and solve the two chains together using an iterative algorithm. This two-chain approach tracks the interaction between the different users in a system model that can be analyzed. We develop a similar approximate model for the system of interacting queues in the mesh broadcast case.

One Markov chain in our model, termed the user chain, represents the queue size for a single user. It is, therefore, an infinite chain. Packets arrive according to a Poisson random process and depart only when this node is chosen for service. We denote the probability that this node is chosen for service by P_s and show that the delay, D, can by expressed as

$$D = \frac{S}{2} + \frac{\lambda S(2-\lambda S)}{2(P_s - \lambda S)}$$

where S is the cycle duration which is equal to (d+1)(N-1). The missing ingredient in this expression, P_s, is the one term that can be obtained from the other chain in our model, termed the system chain.

The system chain represents the number of non-empty nodes on one ring (the ring containing our node of interest). Clearly, this chain consists of N+1 states. The transition probabilities between these states can be expressed in terms of parameters from the user's chain. If S_i denotes the i^{th} state of the system with i non-empty and (N-i) empty nodes and if P_i denotes the steady-state probability of S_i, then P_s can be expressed as

$$P_s = \frac{\sum_{i=1}^{i=d} P_i + d\sum_{i=d+1}^{i=N} \frac{1}{i} P_i}{1 - P_0}$$

Since the system chain equations depend on parameters from the user's chain and visa versa, the two chains are solved together using an iterative algorithm. The results from our approximate model compare very well with simulation, particularly when arrival rates are low.

In order to improve the accuracy of the model, we expanded the system chain to include the identity of the individual queues and their status (empty or non-empty). This adjustment to the system model proved to dramatically improve the performance of our approximation. However, with this change the system chain consists of 2^N states and is difficult to solve for all but very small values of N. To overcome this shortcoming of the expanded model, we limited the system chain so that it merely represents the dentity and state of one user (our user of interest) along with the number of non-empty nodes on the ring. This modification permits a more accurate derivation of the probability of success for the user chain. This is because the probability of success is defined to be the probability that the user is chosen to be served given that it is non-empty. Therefore, when the system chain contains the state of our user, we can compute the probability of success by conditioning on the user state being non-empty. It turns out that this new model is just as accurate as the previous model (containing the identities of all of the users) but since this new chain has only 2(N+1) states it is much easier to analyze.

Since the improved model offers such an improvement to the original model with a minimal additional complexity, we were motivated to develop a similar modification for the ALOHA multiple access protocol. In the ALOHA case we consider a finite number of users, each accepting packets that arrive independently according to a Bernoulli random process, competing for the use of a single channel. If a terminal is empty (has no packets), a newly arrived packet is transmitted immediately. The transmission is successful if and only if no other user attempts transmission during the same slot, otherwise a collision occurs and the terminal enters the blocked state. When in the blocked state, the terminal attempts re-transmission with probability p. In case of success the terminal becomes unblocked. An unblocked terminal can be in one of two states; idle (when its queue is empty), or active (when its queue is not empty). An active terminal transmits a packet with probability one.

The state of any single user can be specified by its queue size and by the indication of whether it is in the blocked or active states. A complete description of a N-terminal system requires the analysis of a 2N-dimensional infinite Markov chain. Again, such chains are known to be very difficult to analyze. We therefore resort to an approximation.

As was stated earlier, in [1] an approximation was developed which modeled an N-dimensional infinite Markov chain as a one-dimensional infinite chain representing the state of a single user together with a N-dimensional finite chain representing the number of blocked and active users in the entire system. In [3] an improvement to the above model was proposed which expanded the system chain to include the identity of all N users. That expanded model was shown to perform far better than the model in [1]; however, the expanded system chain contained 3^N states and was very difficult to analyze for all but very small values of N. We therefore develop a new system chain, similar to the one developed for the multiple-node broadcast algorithm, which includes the state of only one user together with the number of active and blocked users in the entire system.

Our analysis shows that this refined model performs very well at low arrival rates and offers an improvement over the original model in which the system chain contained no information about the individual terminals; however, it does not perform as well as the improved model which contained the identity of all N users. The differences are most noticeable when the arrival rates are high (close to saturation).

References

[1] T. N. Saadawi and A. Ephremides, "Analysis, Stability, and Optimization of Slotted ALOHA with a Finite Number of Buffered Users," *IEEE Transactions on Automatic Control*, June, 1981.

[2] E. Modiano and A. Ephremides, "Efficient Routing Schemes for Multiple Broadcasts in a Mesh" *Twenty-Sixth annual Conference on Information Sciences and Systems*, Princeton, NJ, March 1992.

[3] A. Ephremides and R. Z. Zhu, "Delay Analysis of Interacting Queues with an Approximate Model" *IEEE Transactions on Communications*, Feb., 1987.

ANALYSIS OF THE EXHAUSTIVE CYCLE-GATED SERVICE SCHEME

Irfan Ali *Kenneth S. Vastola*
Department of Electrical Computer and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180.
ali@networks.ecse.rpi.edu vastola@ecse.rpi.edu

The Exhaustive Cycle-Gated (ECG) Service Scheme is used in a single-server multiqueue system. It works as follows: at the beginning of each cycle, when the server reaches queue Q_1, a "poller" is dispatched from Q_1. This poller visits queues in a sequential order $Q_1,Q_2,Q_3,...,Q_N$. When the poller reaches a queue, it marks all the customers present in the queue for service in that cycle. The poller incurs zero delay at each queue; however, there is fixed switch-over time from one queue to the next. The server serves the queue in the same order. At each queue it serves all the *marked* customers. Meanwhile, the arriving customers to all queues have to wait for the next cycle to receive service. When the server returns to Q_1, a new cycle begins. The queues are of infinite length and customer arrival processes to the queues are nonsymmetric independent Poisson processes. The service time of customers has a general distribution which is the same for all the queues (though this can be generalized to different service distributions at stations).

The ECG scheme is related to an existing service scheme, the Gated Sequential Service (GSS) scheme [1], which is used in the Fasnet protocol for high-speed fiber optic bus Local Area Networks. However, in the GSS scheme only the head-of-line customer present at each queue at the beginning of a cycle is served during that cycle. Hence, in our terminology, the service scheme is a 1-limited cycle-gated scheme. The ECG scheme is also related to a recently proposed and analyzed scheme, the Globally Gated (GG) service scheme [2]. In the GG scheme there is a global clock which at the beginning of each cycle, gates the customers in all the queues. These are the customers which receive service when the server arrives to the queue. The GG scheme is impractical for high-speed networks as it is very difficult to maintain a global clock.

We analyze the ECG scheme and derive closed-form expressions for the moment generating function, mean and variance of the waiting time and the number of customers served at each queue at steady state. For the analysis, we employ a space-time diagram to model the system as it elegantly captures the 2-dimensional nature of the problem. From the space-time diagram, the waiting time of each customer is shown to be composed of three components. Equations for these are derived, which in turn gives the Laplace transform of the waiting time of customers at individual stations. We then highlight some properties of the Exhaustive Cycle-Gated (ECG) service scheme. We show that the ECG scheme is similar to window-gated access schemes, in that only those customers are served in a cycle whose arrival time falls within a time window of limited duration which is the same for all stations. The ECG scheme also leads to a natural prioritization of the queues, $EW_1 < EW_2 < ... < EW_N$, where EW_i is the mean waiting time of a customer in Q_i. Moreover, under general nonsymmetric load distribution, the ratio of the mean waiting time at Q_N to that at Q_1 is less than $(1+3\rho)$, i.e $EW_N/EW_1<(1+3\rho)$, where ρ is the normalized load on the system ($\rho<1$). Thus the maximum unfairness is bounded. We also show that the average waiting time of customers arriving to the system (averaged over all the queues) is independent of the spatial distribution of the load in the system. The number of customers served at a station during each cycle is proportional to the arrival rate at that station.

Comparing the ECG scheme to the exhaustive polling disciple, we find that the average waiting time is higher for the ECG scheme. However, it has been widely accepted in polling literature that the exhaustive service discipline is unfair to lightly loaded queues in nonsymmetric traffic arrival scenarios. By considering several cases for low switch-over time between queues, we show that the ECG scheme is more fair to the lightly loaded stations in that they have lower mean waiting times than in the exhaustive service discipline. Extensive numerical results for the ECG scheme as incorporated into Fasnet are given in [3]. Simulation results included therein validate our analysis.

An extension of our work is to consider a variation of the ECG service scheme in which the polling and service order reverses from one cycle to the next. For this service scheme—the Exhaustive Reversing Cycle Gated (ERCG) scheme—we use similar space-time modelling techniques for the analysis of the system. Details of the analysis can be found in [4].

The analysis can also be applied to models for internal mail delivery systems [5]. In these models a clerk picks up, sorts and delivers mail to a closed loop of offices. The mail picked up in a round is sorted in the mailroom and delivered in the next round. The marking of customers (mail) for service is the same as in the ECG scheme; however, the server collects the customers from all the queues and serves them in the mailroom rather than serving them at the individual queues as in the ECG scheme.

References

[1] F. Tobagi and M. Fine, "Performance of unidirectional broadcast local area networks: Expressnet and Fasnet," IEEE Jour. Select. Areas Commun., Vol. SAC-1, No. 5, pp. 913-926, Nov. 1983.

[2] O. J. Boxma, H. Levy and U. Yechiali,"Cyclic reservation scheme for efficient operation of multiple-queue single-server systems,"Annals of Oper. Research, Vol. 35, pp. 187-208,1992.

[3] Irfan Ali and K. S. Vastola, "Performance of exhaustive cycle-gated access in high speed bus networks," Proc. Globecom '92, Orlando, Florida, Dec. 6-9, 1992.

[4] Irfan Ali, Ph.D Thesis, Electrical Computer and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy NY, in preparation.

[5] Irfan Ali and K. S. Vastola, "Analysis of models for internal mail delivery systems," Proc. 5th Advanced Tech. Conf., U.S. Postal Service, Washington D.C., Nov. 30-Dec. 2, 1992.

BLIND WIENER FILTERING: ESTIMATION OF A RANDOM SIGNAL IN NOISE USING LITTLE PRIOR KNOWLEDGE

ABHIJIT A. SHAH and DONALD W. TUFTS *
Dept. of Electrical Engineering, The University of Rhode Island, Kingston, RI 02881, USA

1 Introduction

In this paper we extend an existing signal estimation method [1] so as to estimate the samples of a segment of a stationary random signal embedded in noise where the correlation structure of the process is unknown. The method assumes little prior information and can be applied as a pre-processing step of "cleaning up" the data.

The extension of the method is based on the idea of reducing the rank of the signal model in order to lower the mean-squared estimation error. Rank reduction is a general method for reducing the complexity in linear statistical models in order to lower the mean-squared error in the estimate and to decrease the sensitivity to measurement errors. By using reduced-rank models we lower the variance of the estimate at the expense of introducing bias in the estimate and, by having the right tradeoff between bias and variance, the overall mean-squared estimation error is reduced. In the past, this idea of using rank-reduced models to obtain biased estimates with lower mean-squared errors has been explored for modeling stationary signals with known correlation structure [2], and for solving linear least squares problems [3]. In this paper we analyze the use of a reduced-rank signal model in the context of bias/variance tradeoff for signal vector estimation when the correlation structure of the signal is unknown.

2 Signal Estimation Method

Consider an observed data vector $\tilde{\boldsymbol{y}}_{(L\times 1)}$ containing a signal component $\boldsymbol{y}$ and a noise component $\boldsymbol{w}$. We are interested in estimating the signal component $\boldsymbol{y}$ from the observed data vector $\tilde{\boldsymbol{y}}$. The correlation structure of the signal $\boldsymbol{y}$ and the noise $\boldsymbol{w}$ is unknown.

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{w} \tag{1}$$

The first step of the method is to form a Toeplitz data matrix $\tilde{\mathbf{Y}}_{m\times n}$ from the observed data vector $\tilde{\boldsymbol{y}}$. The formed data matrix $\tilde{\mathbf{Y}}$ can also be represented as a sum of $\mathbf{Y}$, the signal-only Toeplitz matrix and $\mathbf{W}$ the noise Toeplitz matrix.

$$\tilde{\mathbf{Y}} = \mathbf{Y} + \mathbf{W} \tag{2}$$

In the second step of the algorithm the Singular Value Decomposition (SVD) of the Toeplitz data matrix $\tilde{\mathbf{Y}}$ is formed as follows

$$\begin{aligned} \tilde{\mathbf{Y}} &= \tilde{\mathbf{U}}_s\tilde{\mathbf{\Sigma}}_s\tilde{\mathbf{V}}_s^H + \tilde{\mathbf{U}}_o\tilde{\mathbf{\Sigma}}_o\tilde{\mathbf{V}}_o^H & (3)\\ &= \tilde{\mathbf{Y}}_r + \tilde{\mathbf{Y}}_o, & (4)\end{aligned}$$

From the rank r approximation $\tilde{\mathbf{Y}}_r$, the signal component is recovered. The residual matrix $\tilde{\mathbf{Y}}_o$ contains vestiges of the signal component which are traded off in order to reduce the overall mean squared estimation error.

In the third and the final step of the algorithm each element of the estimated signal vector $\hat{\boldsymbol{y}}$ is obtained by a linear combination of the elements of the matrix $\tilde{\mathbf{Y}}_r$. This step of signal recovery can be represented as a matrix filtering operation

$$\hat{\boldsymbol{y}} = \mathbf{A}\tilde{\boldsymbol{y}}_r', \tag{5}$$

where $\tilde{\boldsymbol{y}}_r'$ is a column vector with mn elements obtained by concatenation of the columns of $\tilde{\mathbf{Y}}_r$. The matrix $\mathbf{A}$ consists of filter weights. See [4, 5] for details.

In this paper we assume that an appropriate effective rank for the signal model has been determined from the data or from prior knowledge. An example of determining the effective rank directly from the data is presented in [6].

3 Applications

The generalization of the method can be readily used in the applications of the original method, e.g. (1) Adaptive detection [7, 6] where one can temporarily treat strong interference as a signal to be estimated and then subtracted from the data and (2) Data-adaptive improvement of SNR as a pre-processing step for estimating the values of signal parameters. The performance of approximate maximum likelihood estimation of signal parameters can be improved by first estimating the waveform of the signal component [8].

References

[1] D. Tufts, R. Kumaresan, and I. Kirsteins, "Data adaptive signal estimation by singular value decomposition of a data matrix," *Proc. IEEE*, vol. 7, pp. 684–685, 1982.

[2] L. L. Scharf and D. W. Tufts, "Rank reduction for modelling stationary signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 350–355, Mar. 1987.

[3] A. Thorpe and L. L. Scharf, "Reduced rank methods for solving least squares problems," in *Twenty-third Annual Asilomar Conf. on Signals, Sy stems and Computers*, Nov. 1989.

[4] A. A. Shah and D. W. Tufts, "Estimation of the signal component of a data vector," in *Proc. ICASSP 92*, (San Francisco, CA), pp. 393–396, IEEE, March 1992.

[5] D. W. Tufts and A. A. Shah, "Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix," *IEEE Trans. Acoust., Speech, Signal Processing*. To appear in April 1993.

[6] D. W. Tufts, D. H. Kil, and R. R. Slater, "Reverberation suppression and modeling," in *Proc. Ocean Reverberation Symposium*, (LaSpezia, Italy), May 1992.

[7] I. P. Kirsteins, *Analysis of Reduced Rank Interference Cancellation*. PhD thesis, University of Rhode Island, Kingston, RI, 1990.

[8] R. J. Vaccaro, D. W. Tufts, and Y. Ding, "Improved subspace-based parameter estimation," in *Proc. 1992 Conf. Information Sciences and Systems*, (Princeton, NJ), March 1992.

*This work was supported in part by the Office of Naval Research under grant N00014-90-J1283 and by the SDIO/IST, managed by the Army Research Office under contract DAAL03-86-K-0108, Donald W. Tufts, Principal Investigator.

A METHOD OF MULTI-DIMENSIONAL BLIND EQUALIZATION

Hiroyoshi Oda and Yoichi Sato
Department of Information Science, Faculty of Science, Toho University
2-2-1 Miyama, Funabashi 274, Japan

Abstract: The blind equalization problem for the multi-channel data transmission is investigated. The algorithm is based on the principle of distribution matching, i.e., the total system must be transparent when both the transmitted signal and the equalizer output depend on the same distribution, where the transmitted signal is assumed to be IID sequence. The difficulty in its extention toward multi-dimensional cases is to reduce simultaneously cross-interference between channels and inter-symbol interference in each channel. The cost function, which measures the distance between the joint-distribution of equalizer output vector(z_k) and that of transmitted vector (a_k), is able to solve the difficulty. The proposed algorithm is closely related to the minimum entropy deconvolution (MED), whose cost function measures the distance from the Gaussian to the distribution of equalizer output. By extending kurtosis used in MED theory to multi-dimensional cases, we derive another cost function which appears to be equivalent to the first proposed cost function excepting power normalization.

In multi-dimensional cases, the transmitted signal a_k, the received signal y_k and the equalizer output z_k are series of vector, and the channel response H_k and the equalizer W_k are series of matrix. For example, the equalization in multi-carrier data transmission is such a typical model. In these cases, a_k should be assumed to be multi-dimensional IID(independent identically distributed), i.e., independent both in the time series of each element of the vector and among elements. Therefore, the equalizer must eliminate not only intersymbol interference in each channel but also cross-interference between channels. Letting T_k be the matrix series of total response, our destination is written as $T_0 = I$ (unit matrix), $T_k = 0$ (zero matrix) ($k \neq 0$). Two types of algorithms for blind equalization have been reported. One is based on the strategy to force the probability distribution of equalizer output to that of the transmitted signal [1],[2]. The other aims to eliminate the time-dependency in equalizer output [3]. In this paper, we extend the first algorithm toward multi-dimensional cases. In one-dimensional cases, we have a cost function as $E[(z_k - \gamma \mathrm{sign}(z_k))^2]$, where sign$(\cdot)$ is signum function and γ is constant value given by $\gamma = E[a_k^2]/E[|a_k|]$. It is shown that this cost function measures the distance between the distribution of a_k and z_k, and to minimize it makes the total system transparent. An extension to multi-dimensional cases is easily given by rewriting z_k in vector form z_k. However, this can not remove the inter-channel interferences. For instance, such a significant problem occurs, as plural channels degenerate into single channel and some of channels are dropped-out. To solve this, we propose a new cost function as

$$J_1 = AE[(\|z_k\|^2 - \gamma_1^2)^2] + B\sum_{i=1}^{N} E[((z_k^{(i)})^2 - \gamma_2^2)^2], \qquad (1)$$

where $\|\cdot\|(=\sqrt{z_k^T z_k})$ is Euclidean norm, $z_k^{(i)}$ is the i-th elememt of z_k, $\gamma_1^2 = E[\|a_k\|^4]/E[\|a_k\|^2]$, $\gamma_2^2 = E[(a_k^{(i)})^4]/E[(a_k^{(i)})^2]$, and A and B is positive constant parameters. To mimimize the first term makes the joint-distribution of z_k in the same shape as that of a_k. But ambiguity of arbitrary amount of rotation remains since the first term refers only to the information of radius of z_k. The second term contributes to adjustment of the rotation in the desired direction so that each pair of elements $z_k^{(i)}$ and $a_k^{(i)}$ has the same figure of distribution. In result, when the distribution of a_k is sub-Gaussian, to minimize the cost function J_1 guarantees blind equalization after remains the several ambiguities: the channel swapping, the polarity and time-shift among time series of each element of z_k. It should be noted that the time-shift ambiguity causes the troublesome problem. Consider the case where the time series of some elements of z_k shift in different ways. Then, it can be permitted if the time series of i-th element of z_k is detected in the time-shifted series $a_{k+m}^{(i)}$ of its own channel signal $a_k^{(i)}$. However, our algorithm has a risk of channel drop-out as the time serise of some elements of z_k are detected in the time-shifted series $a_{k+m}^{(i)}$, $a_{k+n}^{(i)}$, ... of another channel's time series $a_k^{(i)}$. One of possible ways to avoid such troubles is to employ a more general and large scale of joint-distribution matching including a number of time-shifted signals ..., z_{k-1}, z_k, z_{k+1}, ...

The proposed method is closely related to the minimum entropy deconvolution (MED) which derive equalizer output as far as possible from Gaussianity [4]. Note that the distribution of the received signal approximates monotonously to the Gaussian when the distortion of the channel increases. Therefore, the miminization of the distance between distributions of a_k and z_k means the maximization of the distance between the Gaussian and the dis tribution of z_k. Extending the MED theory to multi-dimensional cases, we have another cost function using the kurtosis such as

$$J_2 = A\frac{E[\|z_k\|^4]}{E[\|z_k\|^2]^2} + B\sum_{i=1}^{N}\frac{E[(z_k^{(i)})^4]}{E[(z_k^{(i)})^2]^2}. \qquad (2)$$

J_2 is similar to J_1 except that the power normalization function does not work due to non-dimensionality of J_2. J_2 can be applied to the super-Gaussian case of a_k, if the second term is maximized by setting $B < 0$. Fig.1 and 2 shows the simulation results of on-line algorithm for J_1 and off-line algorithm for J_2 in two-dimensional case, where a_k depends on the uniform distribution. For the illy conditioned channel response, if J_1 and J_2 lack the first terms, the phenomenon of channel drop-out easily occurs, since the distribution of each element of z_k is merely forced to that of a_k. It is seen in Fig.1 and 2 that each channel is successfully separated.

References

[1] Y.Sato, "A method of self-recovering equalization for multilevel amplitude modulation," IEEE Trans. Commum., pp.679-682, June 1975.
[2] A.Benveniste, M.Goursat, and G.Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," IEEE trans. Automat. Contr., vol.AC-25, no.3, pp385-399, June 1980.
[3] Y.Sato, H.Oda, and S.Hashimoto, "Blind suppression of time dependency and its extension to multi-dimensional equalization," IEEE Glabal Telecommunication Conference, Houston, Dec1-4, Vol.3 pp.1652-1656, 1986.
[4] D.Donoho, "On miminum entropy deconvolution," in Applied time series analysis II(edited by D. Findley), Academic Press, pp.565-608, 1981.

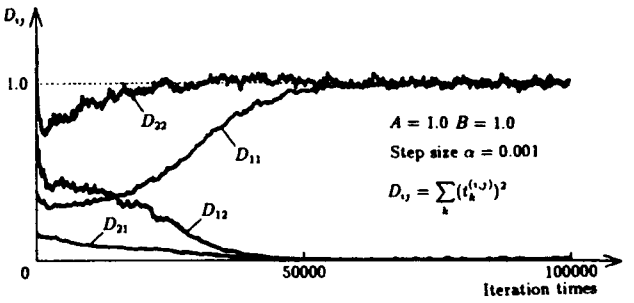


Fig.1 Time evolution of on-line algorithm for J_1

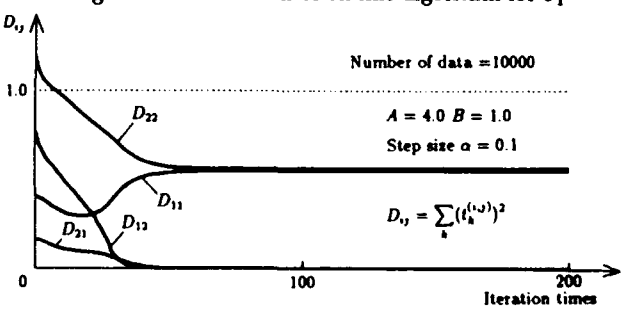


Fig.2 Time evolution of off-line algorithm for J_2

SAMPLING DESIGNS FOR ESTIMATION OF A RANDOM PROCESS

Yingcai Su
Department of Statistics
University of Arizona
Tucson, AZ 85721

Stamatis Cambanis
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599–3260

Summary

This paper deals with the following problem of estimating a random process from a finite number of observations, which arises in statistical communication theory and signal processing as well as in geology (Journel and Huijbregts, 1978) and environmental science (Christakos, 1992).

Suppose a random process $X(t)$, $t \in [0,1]$, is sampled at a finite number of appropriately designed points. On the basis of these observations, we want to estimate the values of the process at the unsampled points and we measure the performance by an integrated mean square error (IMSE).

The process can be modeled as

$$X(t) = m(t) + N(t), \quad t \in [0,1].$$

Here $m(t)$ is the nonrandom large-scale mean structure and we consider

(1) the case where $m(t)$ is known or, equivalently equals zero;

(2) the semiparametric (regression) model where the mean can be modeled as $m(t) = \beta_1 f_1(t) + \ldots + \beta_q f_q(t)$, where the β_i's are unknown coefficients and the f_i's are known (regression) functions; and

(3) the nonparametric case where the macroscopic mean structure $m(t)$ is unknown.

$N(t)$ is the small-scale random structure which models the temporal dependence and has zero mean and known covariance function $R(t,s) = \mathcal{E}N(t)N(s)$. The centered process $N(t)$ is assumed to have no quadratic mean derivative and the functions $m(t)$ and $f_i(t)$ are of comparable smoothness with the microscopic purely random part $N(t)$, specifically, $m(t)$ and $f_i(t)$ are of the form $\int_0^1 R(t,s)\psi(s)ds$.

There are three findings.

The main one is the specification of simple sampling designs which are asymptotically optimal as the sample size increases to infinity. This is done for a variety of estimators. First the best linear unbiased estimator (BLUE) of $X(t)$ is used in Cases (1) and (2) and a nearly BLUE in the nonparametric Case (3). The rate of convergence to zero of the IMSE is n^{-1}:

$$\lim_{n\to\infty} n\, IMSE_n = \int_0^1 \frac{c(t)}{h(t)}dt,$$

where $c(t)$ depends on the covariance function $R(t,s)$ and $h(t)$ is the density of the sampling design.

The second finding is that asymptotically the mean has no effect on the overall performance and can therefore be neglected. This quantifies the discussions in Journel and Rossi (1989) and Sacks et al. (1989, p. 415). However, an example of linear regression in Wiener noise shows that the mean function may cause some perturbation on the optimal sampling design points.

The third finding is that the very simple and nonparametric linear interpolation also leads to an asymptotically optimal performance!

If the centered $N(t)$ has exactly k ($k = 1,2,\ldots$) quadratic mean derivatives, the convergence rate of the IMSE is likely to be $n^{-(k+1)}$ but we do not investigate further this conjecture.

References

[1] G. Christakos, Random Field Models in Earth Sciences (Academic Press, 1992).

[2] A.G. Journel and C.J. Huijbregts, Mining Geostatistics (Academic Press, 1978).

[3] A.G. Journel and M.E. Rossi, When do we need a trend model in kriging?. *Math. Geology* **21** (1989) 715-739.

[4] J. Sacks, W.J. Welch, T.J. Mitchell and H.P. Wynn, Design and analysis of computer experiments, *Statist. Sci.* **4** (1989) 409-435.

[5] J. Sacks and D. Ylvisaker, Designs for regression problems with correlated errors, *Ann. Math. Statist.* **37** (1966) 66-89.

ON SAMPLING THEOREM, WAVELETS AND WAVELET TRANSFORMS *

Xiang-Gen Xia and Zhen Zhang, senior member, IEEE
Communication Sciences Institute, Dept. of EE-Systems
University of Southern California, Los Angeles, CA 90089-2565

Summary

The classical Shannon sampling theorem has many applications and generalizations. From wavelet transform point of view, it provides the sinc wavelets. Recently it has also been extended to general wavelet subspaces by Walter [1]. It says that if a signal $f(t)$ is in wavelet subspace $V_J(\phi)$, then

$$f(t) = \sum_n f(\frac{n}{2^J})\chi(2^J t - n), \qquad (1)$$

where the interpolant $\chi(t)$ has its Fourier transform

$$\hat{\chi}(\omega) = \frac{\hat{\phi}(\omega)}{\sum_n \hat{\phi}(\omega + 2n\pi)}, \qquad (2)$$

and $\hat{\phi}(\omega)$ is the Fourier transform of the scaling function $\phi(t)$ and $\sum_n \hat{\phi}(\omega + 2n\pi) \neq 0$ for any real ω. Aldroubi and Unser [2] considered the case where $\chi(t)$ is a scaling function. In particular, they called a scaling function satisfying

$$\phi(n) = \begin{cases} 1, & n = 0 \\ 0, & n = \pm 1, \pm 2, ..., \end{cases} \qquad (3)$$

a *cardinal scaling function*. In the following, we call the wavelets generated from cardinal scaling functions as *cardinal wavelets*. It is clear that, for a cardinal scaling function $\phi(t)$, the sampling theorem is

$$f(t) = \sum_n f(\frac{n}{2^J})\phi(2^J t - n), \ \forall f(t) \in V_J(\phi), \qquad (4)$$

Since $\phi(t)$ is a cardinal function if and only if $\sum_n \hat{\phi}(\omega + 2n\pi) \equiv 1$, Walter's sampling theorem implies that (3) is also necessary for (4) to be true. This concludes the following proposition.

Proposition 1. The sampling theroem (4) is true if and only if $\phi(t)$ is a cardinal scaling function. □

In this research, we further classify cardinal scaling functions which satisfy (3) and prove that a scaling function $\phi(t)$ with compact support is a cardinal scaling function if and only if $\phi(t)$ is the scaling function corresponding to the Haar wavelets, that is, $\phi(t) = \chi_{[0,1)}(t)$, where $\chi_{[a,b)}(t)$ is an indicator function which is 1 for $t \in [a, b)$ and 0 otherwise. We present a family of cardinal scaling functions which are generalizations of the Haar scaling function $\chi_{[0,1)}(t)$.

If $f(t)$ is not in $V_J(\phi)$, generally (1) is not true. When $f(t) \in V_{J+1}(\phi)$, Walter [1] estimated the error between $f(t)$ and

$$\tilde{f}(t) = \sum_n f(\frac{n}{2^J})\chi(2^J t - n),$$

where $\chi(t)$ satisfies (2). In this research, we consider cardinal wavelets, that is, the sampling theorem (4). We estimate the above error for $f(t)$ which are not necessarily in $V_{J+1}(\phi)$, when $\phi(t)$ is a cardinal scaling function.

As an application of the sampling theorems (1) and (4), efficient computations of wavelet series transform (WST) coefficients from uniform samples of a signal were considered by researchers [3-4], In particular, if $f(t)$ satisfies (1)[1] (or (4)) then the WST coefficients of $f(t)$ can be exactly obtained from $f(n/2^J)$ by using the Shensa algorithm (or the Mallat algorithm, a special case of the Shensa's). Since the Mallat algorithm is generally simpler than the Shensa's one, one might prefer the sampling theorem (4). For signals which are not necessarily in wavelet subspaces, usually error occurs when one uses the Mallat algorithm to compute the WST coefficients. In this research, we also present several numerical examples to compare the errors when the Haar wavelets, Daubechies D_4 and D_8 wavelets and cardinal wavelets are used. These examples indicate that the error for the cardinal wavelets is much smaller than the ones for other wavelets.

References

[1]. G.G.Walter, "A sampling theorem for wavelet subspace," IEEE Trans. on Information Theory, vol.38, pp.881-884, Mar. 1992.

[2]. A.Aldroubi and M.Unser, "Families of wavelet transforms in connection with Shannon's sampling theory and the Gabor transform," Wavelets: A Tutorial in Theory and Applications, ed. by C.K.Chui, pp509-528, Academic Press, New York, 1992.

[3]. O.Rioul and P.Duhamel, "Fast algorithms for discrete and continuous wavelet transform," IEEE Trans. on Information Theory, vol.38, pp569-586, Mar. 1992.

[4]. X.G.Xia, C.-C.J.Kuo and Z.Zhang, "On the accuracy of the computed wavelet coefficients," submitted.

*This research is supported in part by NSF under Grant NCR-8905052.

[1]The property (2) for $\chi(t)$ is not necessary.

Construction of Discrete Orthogonal Wavelet Bases

Chao Wei and Douglas Cochran

Telecommunications Research Center
Arizona State University
Tempe, Arizona 85287-7206

Abstract - *In general, two sequences formed by uniformly sampling two orthogonal signals will not be orthogonal. This paper presents families of discrete orthonormal wavelet bases for ℓ^2 that are obtained by sampling of certain dyadic orthonormal wavelet bases of L^2 over a bounded frequency band.*

1. INTRODUCTION

Numerous wavelet bases for $L^2(\mathbb{R})$ have been described in recent mathematical and engineering literature. Because of their tractability in applications, dyadic orthonormal wavelet bases have received considerable attention. Use of such bases in discrete-time settings generally involves sampling of the mother wavelet at uniform intervals which are power-of-two multiples of a fixed interval T. This raises the issue that the sample sequences obtained from orthogonal time-scale replicates of the mother wavelet may not be orthogonal.

This paper investigates the problem of obtaining (discrete-time) dyadic wavelet bases for ℓ^2 from (continuous-time) dyadic wavelet bases of L^2. In particular, a construction of S. Mallat is extended and the Whittaker-Kotel'nikov-Shannon (WKS) sampling theorem is applied to obtain a family of discrete orthonormal wavelet bases for ℓ^2.

2. BASIS CONSTRUCTION

In the frequency domain, define

$$\hat{W}^a(\omega) = \begin{cases} 1 & \text{if } \pi < |\omega| \le 2\pi \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and for $m \in \mathbb{N}$ and $n \in \mathbb{Z}$, define $\hat{W}^a_{m,n}(\omega) = 2^{m/2}\hat{W}^a(2^m\omega)e^{-i2^m n\omega}$ or $W^a_{m,n}(t) = 2^{-m/2}W^a(t/2^m - n)$. Then $\hat{W}^a_{m,n}$ with $m \in \mathbb{N}$ and $n \in \mathbb{Z}$ form an orthogonal basis of $L^2[-\pi,\pi]$. Moreover, each $W^a_{m,n}$ is bandlimited to $[-\pi,\pi]$ and hence by the WKS theorem may be represented by samples $\{W^a_{m,n}(k)\}_{k\in\mathbb{Z}}$ or $\{W_{m,n}(k)\}_{k\in\mathbb{Z}}$.

Theorem 1: The sample sequences $\{W_{m,n}(k)\}_{k\in\mathbb{Z}}$ are orthonormal; i.e., for arbitrary dyadic dilation indices $m \in \mathbb{N}$ and $m' \in \mathbb{N}$ and integer time shifts $n \in \mathbb{Z}$ and $n' \in \mathbb{Z}$,

$$\sum_{k\in\mathbb{Z}} W_{m,n}(k)W^*_{m',n'}(k) = \begin{cases} 0 & \text{if } (m,n) \ne (m',n') \\ 1 & \text{if } (m,n) = (m',n') \end{cases} \tag{2}$$

Proof: From above, $\{W^a_{m,n}\}$ is an orthogonal basis of L^2 and $\{W_{m,n}\}$ is generated by sampling the corresponding analog function with sampling interval $T = 1$. Note that dyadic dilation index m is restricted to be a natural number and the time shift index n to be an integer.

Let $\hat{W}_{m,n}$ be the discrete time Fourier transform (DTFT) of the discrete signal $W_{m,n}$

$$\hat{W}_{m,n}(\omega) = \sum_k W_{m,n}(k)e^{-i\omega k}$$

The relationship between $W^a_{m,n}$ and $W_{m,n}$ is

$$\hat{W}_{m,n}(\omega) = \sum_k \hat{W}^a_{m,n}(\omega - 2\pi k) \tag{3}$$

Since $\hat{W}^a_{m,n}(\omega) = 0$ when $|\omega| > \pi$,

$$\hat{W}_{m,n}(\omega) = \hat{W}^a_{m,n}(\omega) \text{ when } |\omega| \le \pi \tag{4}$$

By Parseval's theorem

$$\sum_k W_{m,n}(k)\, W^*_{m',n'}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{W}_{m,n}(\omega)\hat{W}^*_{m',n'}(\omega)\, d\omega$$

From equation (4),

$$\begin{aligned} \langle W_{m,n}, W_{m',n'}\rangle &= \sum_k W_{m,n}(k)W^*_{m',n'}(k) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{W}^a_{m,n}(\omega)\hat{W}^{a*}_{m',n'}(\omega)\, d\omega \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{W}^a_{m,n}(\omega)\hat{W}^{a*}_{m',n'}(\omega)\, d\omega \\ &= \langle W^a_{m,n}, W^a_{m',n'}\rangle \end{aligned}$$

Orthonormality of the analog functions $\{W^a_{m,n}\}$ thus implies orthonormality of the sample sequences $\{W_{m,n}\}$. ∎

Theorem 2: The sequences $\{W_{m,n}(k)\}_{k\in\mathbb{Z}}$ span ℓ^2. Thus they comprise an orthonormal basis for the space of discrete-time finite energy signals.

Proof: For any discrete signal $f \in \ell^2$ whose DTFT $\hat{f}$ is periodic with period 2π, an analog signal f^a can be constructed by inverse Fourier transform of the frequency domain function defined by

$$\hat{f}^a(\omega) = \begin{cases} \hat{f}(\omega) & |\omega| \le \pi \\ 0 & \text{otherwise} \end{cases}$$

The function f^a can be expressed as a weighted sum of the orthogonal basis elements $W^a_{m,n}$. Thus

$$f^a(t) = \sum_m \sum_n a_{m,n}W^a_{m,n}(t)$$

and

$$f(k) = \sum_m \sum_n a_{m,n}W_{m,n}(k) \qquad \forall k \in \mathbb{Z} \tag{5}$$

Hence $\{W_{m,n}\}$ spans ℓ^2. ∎

The procedure just described may be applied to other frequency-domain L^2 wavelet basis constructions using bandlimited wavelets to yield other discrete orthogonal bases for ℓ^2.

Some additional properties of the above construction are listed below without proof. Let $f(k)$ be as in equation (5) and define

$$g(k) = \sum_m \sum_n b_{m,n}W_{m,n}(k) \quad \text{and} \quad h(k) = \sum_m \sum_n c_{m,n}W_{m,n}(k)$$

Then

- *Linearity:* If $h = f + g$ then $c_{m,n} = a_{m,n} + b_{m,n}$
- *Convolution:* If $h = f * g$ (i.e., $h(n) = \sum_k f(k)g(n-k)$) then $c_{m,n} = a_{m,n} * b_{m,n} = 2^{m/2}\sum_k a_{m,n-k} \cdot b_{m,k}$
- *Dyadic dilation:* $W_{m,0}(k) = \sqrt{2}W_{m+1,0}(2k)$

Reference

[1] S. G. Mallat, *Multiresolution Representations and Wavelets*. Ph.D. thesis, University of Pennsylvania, August 1988.

Time-Warped Bandlimited Signals: Sampling, Bandlimitedness, and Uniqueness of Representation

James J. Clark

Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Douglas Cochran

Department of Electrical Engineering
Arizona State University
Tempe, AZ 85287-5706

1. INTRODUCTION

The ability to reconstruct a complex-valued signal on $\mathbb{R}$ from a sequence of sample values $\{f(t_n)\}_{n\in\mathbb{Z}}$ is desirable in a variety of engineering applications. While this problem is ill-posed in general, many reconstruction formulas of the form

$$f(t) = \sum_{n\in\mathbb{Z}} f(t_n)\varphi_n(t) \tag{1}$$

have been obtained for various restricted classes of functions.

It was observed in [1] that such a formula for reconstruction of functions from a given class $\mathcal{C}$ extends directly to a reconstruction formula for functions formed by composition of any $f \in \mathcal{C}$ with an invertible function $\gamma : \mathbb{R} \to \mathbb{R}$. Application of a coordinate transformation such as γ to the domain of a signal is commonly called "time-warping" in signal processing literature. Consequently, signals of this type have become known as "time-warped" signals.

Among the most important formulas of the type (1) are connected with reconstruction of bandlimited signals; i.e., functions having the form

$$f(t) = \frac{1}{2\pi}\int_{-\Omega}^{\Omega} \hat{f}(\omega)e^{i\omega t}\,d\omega \tag{2}$$

where $\hat{f} \in L^2(\mathbb{R})$ and $0 < \Omega < \infty$. Motivated by their reconstructability from samples, this note presents some comments on the class $\mathcal{B}\circ\Gamma$ of time-warped bandlimited signals; i.e., functions of the form $f\circ\gamma$ with f belonging to the class $\mathcal{B}$ of bandlimited signals and $\gamma : \mathbb{R} \to \mathbb{R}$ belonging to a class Γ of continuous and invertible warping functions.

2. RESULTS

The perspective of Paley and Wiener [3] that it is natural to consider bandlimited functions on the complex domain is adopted in what follows. It thus becomes necessary to consider warping functions on $\mathbb{C}$ as well. Given a bandlimited function $f : \mathbb{R} \to \mathbb{C}$, denote by F the corresponding entire function with values defined by

$$F(z) = \frac{1}{2\pi}\int_{-\Omega}^{\Omega} \hat{f}(\omega)e^{i\omega z}\,d\omega$$

Similarly, given $h \in \mathcal{B}$, denote by H the associated entire function. Define $\mathcal{G}$ to be the collection of all continuous functions $G : \mathbb{C} \to \mathbb{C}$ with restrictions γ to $\mathbb{R}$ that are real-valued and bijective. If $G \in \mathcal{G}$ then the corresponding $\gamma \in \Gamma$ is well defined. Thus, given bandlimited functions F and H on the complex domain, finding a $G \in \mathcal{G}$ such that $H = F\circ G$ ensures that there is some $\gamma \in \Gamma$ such that $h = f\circ\gamma$. Given $\gamma \in \Gamma$ such that $h = f\circ\gamma$, however, there is no *a priori* guarantee that any $G \in \mathcal{G}$ exists with the property that $H = F\circ G$. In this sense, considering complex warping functions in G is more restrictive than considering real-valued warping functions in Γ.

Theorem 1: If $f \in \mathcal{B}$ is not identically zero and $G \in \mathcal{G}$, then $H = F\circ G$ is bandlimited if and only if G is affine.

It is clear that $H = F\circ G$ will be bandlimited if G is affine. The proof of the "only if" part of this theorem is based on the growth properties of the entire functions F and H. Specifically, it relies on the following results.

Lemma: Suppose $G \in \mathcal{G}$, $f \in \mathcal{B}_\Omega$ is not identically zero, and $H = F\circ G$ is bandlimited, then G is entire.

Theorem 2 (from [4]): If F and G are entire and the order of $F\circ G$ is finite, then either (i) G is a polynomial and the order of F is finite, or (ii) G is a non-polynomial function of finite order and the order of F is zero.

Theorem 3 (based on results from [4]): If $f \in \tilde{\mathcal{B}}$ is not identically zero and G is a polynomial of degree $n > 1$, then the order of $H = F\circ G$ is greater than one.

The proof of Theorem 1 proceeds as follows. Assuming H is bandlimited, Theorem 1 establishes G is entire. Theorem 2 may be applied to show that G is a polynomial. Theorem 3 implies that the degree of G is either zero or one. If G were constant then H would be constant. Since $h \in L^2$, it cannot be constant without being identically zero. Thus G is a polynomial of degree exactly one; i.e., $G(z) = az + b$ with $a \neq 0$. The condition that γ is real valued implies that a and b are real. Hence $\gamma(t) = at + b$ for real numbers a and b with $a \neq 0$.

3. DEMODULATION

Earlier work [2] has established that $\mathcal{B}\circ\Gamma$ contains all bandlimited functions and many nonbandlimited functions, but not all of $\mathbf{L}^2$. A remaining issue is that of ***demodulation***: given $h \in \mathcal{B}\circ\Gamma$, can it be decomposed into a bandlimited function f and a bijective monotone time warping function γ?

If $h \in \mathcal{B}\circ\mathcal{G}$, then there are necessarily many ways to express h as a composition $f\circ\gamma$. Given any $\alpha > 0$, for example, define functions f_1 and γ_1 by $f_1(t) = f(\alpha t)$ and $\gamma_1(t) = \gamma(t/\alpha)$. Then $f_1 \in \mathcal{B}$, $\gamma_1 \in \mathcal{G}$, and $f_1\circ\gamma_1 = f\circ\gamma = h$. This kind of representational ambiguity can be circumvented by stipulating that f have exactly unit bandwidth. In this case, the question of representational ambiguity may be addressed by a corollary to Theorem 1.

Corollary [of Theorem 1]: Suppose $h = f_1\circ\gamma_1 = f_2\circ\gamma_2$ with f_1 and f_2 having exactly unit bandwidth and $\gamma_1, \gamma_2 \in \mathcal{G}$. Then $f_1(t) = f_2(t-b)$ and $\gamma_1(t) = \gamma_2(t) + b$ for some real constant b and all $t \in \mathbb{R}$.

References

[1] J.J. Clark, M.R. Palmer, and P.D. Lawrence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33(4), pp. 1151-1165, October 1985.

[2] D. Cochran and J.J. Clark, "On the sampling and reconstruction of time-warped bandlimited signals," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1539-1541, April 1990.

[3] R.E.A.C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, AMS Colloquium Publications, vol. XIX, American Mathematical Society, 1934.

[4] G. Pólya, "On an integral function of an integral function", *Journal of the London Mathematical Society*, vol. 1, pp. 12-15, 1926.

A UNIFIED PARTITIONING AND FOLDING PROCEDURE FOR SYSTOLIC ALGORITHMS

Flavio Lorenzelli and Kung Yao
Electrical Engineering Department
University of California, Los Angeles, CA 90024-1594

In many signal processing applications related to estimation or Kalman filtering problems, efficient techniques such as QR decomposition, recursive least squares (RLS), etc., are widely employed. Real-time use of such techniques is made possible through the development of **systolic algorithms** (SAs) and their mapping onto modular parallel processor arrays. We present a **linear** mapping procedure for SAs based on integral matrix theory which includes partitioning, folding, and predefined design constraints. Both partitioning and folding introduce useful degrees of freedom in the design of the final array.

Mapping Procedure

In order to apply the integral matrix theory on SA mapping, the algorithm must be described in geometrical terms as a dependence graph (DG), i.e., a lattice embedded in a multidimensional integral space. Here, we propose a mapping procedure for SAs which can include partitioning (locally sequential-globally parallel, LSGP, as well as locally parallel-globally sequential, LPGS) and folding, and takes into account predefined design constraints. To partition an algorithm means to break it into components of smaller size. These components can be physically executed in parallel or in a sequence. Care must be taken to satisfy the constraints dictated by the partial ordering among computations, the locality of the data flow, etc.. By folding we denote the operation of displacing sections of the projected graph, in order to obtain a desired pattern. This kind of operation is usually highly non-linear at the physical array level, and is normally performed in a heuristic manner, after projection. Classical systolic mapping procedures are based on linear or affine transformations, which on the one hand make the design simple and manageable, by using well-understood tools, but on the other hand limit the range of manipulations that can be applied to the original algorithm. Nonlinearity can be introduced in the procedure in order to make efficient use of computing and memory elements. Both partitioning and folding can be included in a unified **linear** mapping procedure. Partitioning can be included as a natural extension of the general mapping procedure, whereas folding requires a preliminary transformation of the DG. The idea is to artificially increase the dimensionality of the DG of the original algorithm, by fragmenting one or more of the old dimensions. In so doing, folding can actually become possible, without having to introduce nonlinear allocating functions.

As mentioned above, the algorithm must first undergo a number of refinements (regularization, single assignment form, etc.), so that the associated DG has the desired properties of locality and shift invariance. The DG $\mathcal{G}$ can be seen as an integral lattice, i.e., a proper convex and bounded subset of $\mathbf{Z}^n$, $\mathbf{Z}$ being the set of relative integers. Each node of the graph represents a computation, whereas the directed arcs connecting the nodes (dependences) represent the dependence relationships between computations. The dependence matrix D collects the dependence vectors D_i as columns. In the sequel assume for simplicity's sake the DG to be parallelepipedal:

$$\mathcal{G} = \{(i_1, \ldots, i_n) \in \mathbf{Z}^n | L_j \leq i_j \leq U_j, j = 1, \ldots, n\},$$

for given lower and upper limits $L_j < U_j, j = 1, \ldots, n$.

Suppose we want to project the n-dimensional DG onto an $(n-p)$-dimensional array. The p projection vectors are collected in the $n \times p$ matrix U_-, the left submatrix of the unimodular matrix $U = [U_-, U_+]$. The $(n-p)$ columns of the matrix Σ_+, the right submatrix of $\Sigma = [\Sigma_-, \Sigma_+] = U^{-\mathsf{T}}$, by definition orthogonal to the p projection vectors, can be considered as a valid basis for the processor space, so that a given node $\mathbf{J} \in \mathcal{G}$ is projected onto the point $\mathbf{J}' = \Sigma_+^{\mathsf{T}}\mathbf{J}$ in the processor space. The scheduling is also defined by p vectors, stacked as columns of the matrix Λ_-. (Again, the matrix $\Lambda = [\Lambda_-, \Lambda_+]$ is the unimodular extension of Λ_-.) The matrix Λ_- must be chosen in such a way as to maintain the ordering of the operations (precedence constraint). Furthermore, no two computations may be mapped on the same cell at the same time (compatibility constraint). The first constraint can be formally stated as follows

$$\Lambda_-^{\mathsf{T}} D_i \geq \mathbf{0}_p, \quad \mathbf{1}_p^{\mathsf{T}} \Lambda_-^{\mathsf{T}} D_i \geq 1, \quad \forall D_i.$$

The compatibility constraint can be specified only when a scheduling function is chosen. A natural definition is the affine function $\vartheta(\mathbf{J}) = \nu^{\mathsf{T}}\Lambda_-^{\mathsf{T}}\mathbf{J} + \zeta(\mathbf{J})$, where the components of the $p \times 1$ vector ν and the affine constant $\zeta(\cdot)$ are suitably chosen. The compatibility constraint requires then that

$$\vartheta(\mathbf{J}_1) \neq \vartheta(\mathbf{J}_2), \quad \text{if} \quad \mathbf{J}_1 \neq \mathbf{J}_2 \quad \text{and} \quad \Sigma_+^{\mathsf{T}}\mathbf{J}_1 = \Sigma_+^{\mathsf{T}}\mathbf{J}_2.$$

These equations immediately become $\mathbf{J}_1 = \mathbf{J}_2 + U_- k$, for some p-vector $k \neq \mathbf{0}$,

$$\nu^{\mathsf{T}}\Lambda_-^{\mathsf{T}} U_- k \neq \zeta(\mathbf{J}_1) - \zeta(\mathbf{J}_2), \quad k \neq \mathbf{0}.$$

These constraints, in addition to the minimization of the overall computation time, can be used to choose the right values for ν and the constants $\zeta(\cdot)$.

The matrix $M = \Sigma_+^{\mathsf{T}} T_+$ is related to the partitioning of the algorithm, whenever its determinant has absolute value larger than unity. It can be proven that the columns of M define the shape of each component of the partitioned algorithm and that its determinant is related to the number of nodes constituting the component (its "volume"). This mathematical framework also allows the inclusion of design constraints in the mapping procedure, dealing with the interconnection pattern, I/O port location, etc.. These requirements affect the choice of both the projection matrix U and the scheduling matrix Λ.

Folding can be included in the procedure by first artificially increasing the dimensionality of the DG. One way to achieve this is to fragment some or all dimensions of the original DG by, e.g., limiting the number of nodes along any given coordinate axis to a prespecified amount (say, n_ℓ nodes along direction ℓ). A natural way to do this is to split index i_ℓ into the $\left\lceil \frac{U_\ell - L_\ell + 1}{n_\ell} \right\rceil$ indices $i_{\ell\nu}$ so that $i_{\ell\nu} = i_\ell \bmod n_\ell, \quad \nu = \left\lfloor \frac{i_\ell}{n_\ell} \right\rfloor$. The DG is now composed of a number of different regions which can be mapped with some degree of independence. One needs to take care of the data movement between the newly formed regions, since it can result in non-local interconnections. If global links are not permitted, this poses a constraint on the projection vector. Besides, the scheduling function must allow a correct timing of the data transition between regions. The modification of the DG has to be done in a preliminary step, so that the mapping procedure can be applied in the normal fashion starting from this higher dimensional graph.

A Coding Theorem for Low-Rate Transform Codes

Daniel F. Lyons
MITRE Corporation
7525 Colshire Drive
McLean, Virginia 22102-3481

David L. Neuhoff
Department of Electrical Engineering
and Computer Science
University of Michigan
Ann Arbor, Michigan 48109

Abstract[1]

Transform codes are used to study low-rate quantization of stationary Gaussian sources. The transform decorrelates the source samples and then scalar quantization is applied to the vector of transform coefficients. Two bit allocations are considered: the first permits only zero or one bit to be allocated to each transform coefficient (i.e., the scalar quantizers have only one or two levels), and the second is an optimal bit allocation. For the transform codes with the "0-1" bit allocation, a closed-form, parametric expression is derived for the asymptotic (with dimension) rate vs. distortion performance. This expression is compared to the rate-distortion function, as well as to the performance of transform codes with optimal bit allocations. The principal result is that there is a critical rate, determined by the power spectral density, below which (and only below which) 0-1 allocations are optimal. This is a unique result in that it determines optimal theoretical performance for an important class of vector quantizers at low rates. Quantitative results are presented for Gauss-Markov sources.

Summary

Whereas a well understood body of theory exists for the analysis of high-rate quantization systems (see for example [1] or [2, ch. 5]), a general theory for analyzing and designing low-rate quantization systems is not yet available. In this paper we analyze two classes of low-rate transform codes. We examine rates less than (and frequently much less than) 1 bit/sample and allow the quantizer dimension (or block length) to become very large so that asymptotic methods may be applied. We consider only discrete-time, stationary Gaussian sources and mean-squared error as a fidelity criterion.

Transform-based source coding systems are examples where there is a need to design simple, low-rate quantizers for "lower energy" transform coefficients. We are able to show that for low rates, the Karhunen-Loeve transform is the optimal transform among the class of orthogonal transforms. Hence the coefficients are also Gaussian with variances that are the eigenvalues of the source covariance matrix. As blocklength becomes large we can determine the asymptotic distribution of the coefficient variances via Szego's Theorem for Toeplitz forms [3, ch. 5].

We first propose a product code that scalar quantizes each of the transformed components at rate either 0 or 1 bit/sample; we refer to the resulting transform code as a 0-1 transform code. The 0-1 transform codes may be designed from rates 0 to 1 bit/sample. For asymptotically large block lengths we find the following parametric expressions for the rate and distortion of 0-1 transform codes,

$$R_\alpha = \frac{1}{2\pi}\int_{\{\omega:\phi(\omega)\geq\alpha\}} \mathrm{d}\omega \tag{1}$$

$$D(R_\alpha) = \frac{1}{2\pi}\int_{\{\omega:\phi(\omega)\geq\alpha\}} c(1)\phi(\omega)\mathrm{d}\omega + \frac{1}{2\pi}\int_{\{\omega:\phi(\omega)<\alpha\}} \phi(\omega)\mathrm{d}\omega \tag{2}$$

where $\phi()$ is the power spectral density of the source, $\inf_\omega \phi(\omega) \leq \alpha \leq \sup_\omega \phi(\omega)$, and $c(r)$ is the mean squared error of a 2^r-level Lloyd-Max quantizer for a unit-variance Gaussian source. Thus by varying the free parameter α, rates from 0 to 1 bit/sample can be achieved. Comparisons are made between the above expressions and the source rate-distortion function. In particular, we find that the distortion penalty above the general Gaussian distortion-rate function is the same as that of Lloyd-Max quantizers above the iid distortion-rate function.

The principal result of our work is as follows. For any Gaussian random process, there exists a critical rate r_0 below which 0-1 codes are the optimal transform codes. In particular,

$$r_0 = \frac{1}{2\pi}\int_{\{\omega:\phi(\omega)\geq\beta\phi_{max}\}} \mathrm{d}\omega$$

where $\phi_{max} = \text{ess sup}_\omega \phi(\omega)$, and

$$\beta = \frac{c(1)-c(2)}{1-c(1)}$$

Knowledge of the process power spectral density is, therefore, sufficient to determine this critical rate. Our specific result is a coding theorem which states that below r_0, 0-1 codes are optimal, and above r_0 there exist other transform codes strictly better than 0-1 codes. Since the asymptotic distortion versus rate performance of 0-1 codes has been derived in (1) and (2), the optimal theoretical performance (OPTA) for any transform code on a particular source has now been found for all rates less than r_0. This represents, to our knowledge, the first complete characterization of the theoretically achievable performance of an important class of quantizers at low rates.

We demonstrate our results for Gauss-Markov sources, and make performance comparisons to other transform codes. The implications of this theory to the design of practical, low-rate codes are discussed.

[1] This work supported under NSF grant NCR-9105647

References

[1] A. Gersho, "Asymptotically Optimal Block Quantization," *IEEE Trans. Inf. Thy.*, IT-25 (July, 1979), pp. 373-380.

[2] R.M. Gray, *Source Coding Theory*, Kluwer Academic Press, 1990.

[3] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*, Chelsea, 1984.

[4] A. Segall, "Bit Allocation and Encoding for Vector Sources," *IEEE Trans. Inf. Thy.*, IT-22 (March, 1976), pp. 162-168.

Boundedness and Consistency of Greedy Growing for Tree-structured Vector Quantizers

Andrew B. Nobel
Beckman Institute
University of Illinois
Urbana, Illinois 61801

and

Richard A. Olshen
Division of Biostatistics
Stanford University
Stanford, California 94305-5092

ABSTRACT The problem of designing a TSVQ from random data has received considerable recent attention. A key step in many methods of design is the application of a greedy growing algorithm to the empirical distribution of the data. In applications of interest to us these empirical distributions are of vectorial pixel intensities. Here we analyze the behavior of the greedy growing algorithm when it is applied to the *true* underlying distribution of the observations, and we show that quantizers produced from large data sets will be close to quantizers produced from the true distribution.

I. INTRODUCTION

The problem of designing a tree-structured vector quantizer (TSVQ) from a sequence of random observations has received considerable attention ([2],[3]). Recall that a TSVQ is completely specified by a binary tree T whose nodes are labeled by points in $\mathbb{R}^k$: denote the corresponding quantizer by Q_T. Let $X_1, X_2, \ldots$ be a stationary ergodic sequence of random vectors $X_i \in \mathbb{R}^k$ having distribution P. The *rate* of a tree T is the expected depth of T or, equivalently, the expected number of comparisons required to encode a random vector X_i. The design problem is as follows:

> Given $X_1, \ldots, X_n$ and a rate $R \geq 0$, find a tree T_n whose rate is not more than R and whose distortion $E\|X - Q_{T_n}(X)\|^2$ is small.

Here X is a random variable distributed as P, but which is independent of the process $\{X_i\}$, and $\|\cdot\|$ denotes the ordinary Euclidean norm on $\mathbb{R}^k$.

One approach to the design problem, based on [1], is to apply a greedy growing algorithm to the empirical distribution $\hat{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ of the data $X_1, \ldots, X_n$. The algorithm "grows" a TSVQ in a step-wise optimal fashion, terminating when it obtains a tree whose (empirical) rate is greater than R. More precisely, given a distribution H on $\mathbb{R}^k$, and a rate $R \geq 0$, the algorithm produces a nested sequence of trees, each of which corresponds to a successive hyperplane-based partitioning of $\mathbb{R}^k$. At each stage, the algorithm splits any terminal node v of the current tree whose corresponding cell V maximizes the ratio $\Delta D^*(V)/\Delta R(V)$ over all such cells. Here $\Delta D^*(V)$ is the greatest reduction in distortion (with respect to H) achievable by a hyperplane split of V, and $\Delta R(V) = H(V)$ is the increase in rate associated with splitting the node v. The children of v are labeled by the centroids of the regions created when V is split. The algorithm terminates when every split of the sort described would make the overall rate of the next tree greater than R. Use of the splitting criterion $\Delta D^*(V)/\Delta R(V)$ amounts to a steepest descent in the rate-distortion plane.

II. RESULTS

Throughout we assume that the distribution P of the random vectors X_i has bounded support, and that P has a density with respect to Lebesgue measure. Our first three results pertain to to the "true" distribution P, while the last one pertains to random data.

Lemma 1 *If $R < \infty$ then the greedy growing algorithm will terminate in a finite number of steps, producing a finite tree.*

Although the assumption that P has bounded support is restrictive, it is *necessary* to ensure termination of the greedy growing algorithm.

Proposition 1 *Let $H = exp(1)$ be the ordinary one-sided exponential. Then if R is larger than some fixed constant R_0, the greedy growing algorithm will not terminate.*

Note that the output of the greedy algorithm need not be unique. At some stage of the algorithm there may be a multiplicity of optimal splits, involving one or more nodes: the algorithm selects *one* of these, and subsequent splits are made accordingly. In this context we can strengthen Proposition 1.

Theorem 1 *For every $R \geq 0$ there is a constant $K < \infty$, depending on R, such that every tree produced by greedy growing has maximum depth less than K.*

Definition: Let $\mathcal{A}(H, R)$ be the collection of trees produced by the greedy algorithm acting on a fixed distribution H and any rate $R' \leq R$.

The following consistency result shows that quantizers produced by the algorithm when it is applied to the empirical distribution $\hat{P}_n$ will eventually be close to quantizers produced by the algorithm when it is applied to the true distribution P of the observations.

Theorem 2 *Fix $\epsilon > 0$ and for $n = 1, 2, \ldots$ let $T_n \in \mathcal{A}(\hat{P}_n, R)$. With probability one there exists a sequence $\tilde{T}_n \in \mathcal{A}(P, R)$ such that*

$$P\{|Q_{T_n} - Q_{\tilde{T}_n}| \geq \epsilon\} \to 0$$

as n tends to infinity.

References

[1] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees.* Wadsworth, Belmont, CA, 1984.

[2] A. Buzo, A.H. Gray Jr., R.M. Gray, and J.D. Markel. Speech coding based upon vector quantization. *IEEE Trans. Acoust. Speech Signal Process.*, 28:562–574, 1980.

[3] E.A. Riskin and R.M. Gray. A greedy tree growing algorithm for the design of variable rate vector quantizers. *IEEE Trans. Signal Process.*, 39:2500–2507, 1991.

SOURCE CLUSTERING FOR CODEBOOK COMPRESSION[1]

Wai-Yip Chan† and Allen Gersho‡

† Department of Electrical Engineering
McGill University
Montreal, Canada H3A 2A7

‡ Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

Clustering algorithms can be applied to the design of N codebooks to be shared by M sources, $1 < M < N$. We previously introduced a *constrained storage vector quantization* algorithm for this design problem. In this work, we extend the algorithm to additionally design simple parametric *expandor functions* to enhance codebook sharing efficiency. We apply the particular case of scaling expandor functions to the compression of tree structured vector quantization codebooks. By allowing different levels of the tree codebook to share a library of feature (residual) codebooks, we were able to achieve in our experiment a 4 : 1 reduction of storage without compromising rate-distortion performance. For very deep trees, an earlier design method which effects sharing only within each level of the tree is more effective.

Summary

Clustering is a popular means to codebook "training," i.e. to reduce a "training set" characterization of the probability distribution of a source to a smaller set of representative vectors. The most common clustering algorithm for codebook design is perhaps the generalized Lloyd algorithm (GLA) [1], also known as the K-means algorithm. Recently, we introduced what could be considered as a generalization of the GLA, called the *constrained storage vector quantization* (CSVQ) algorithm [2], for clustering a set of N sources to *share* M codebooks, $1 < M < N$. The CSVQ algorithm was introduced to control the storage complexity for one *feature* at a time in *generalized product codes* [3]. The algorithm has been applied to improve the performance of multistage VQ (MSVQ) [2] and restrict the storage complexity of tree structured VQ (TSVQ) [4]. Chou [5] listed various other applications of GLA clustering.

In [6], Lindsay *et al.* described various ways of restricting the orientation of the partitioning hyperplanes of a binary TSVQ codebook to achieve storage reduction. Alternately, the method we described in [4] requires the TSVQ codebook to be stored in an equivalent trellis of *residual* feature codebooks [3]. In other words, TSVQ is equivalent to MSVQ except that TSVQ has path-dependent residual codebooks. The trellis is a consequence of codebook sharing by the conditional residual sources in one stage; without sharing, the trellis becomes a tree. The sharing is instrumented by applying the CSVQ algorithm to grow the tree one level at a time. Linear growth, as opposed to exponential growth, in storage complexity with rate has been achieved while incurring virtually no rate-distortion penalty [4]. While this method was used to design binary TSVQ codebooks up to many levels (say 26), for codebooks of moderate depth, it is possible to achieve sharing of the feature codebooks over the entire tree rather than just one level at a time.

Suppose there are N sources sharing M codebooks, $1 < M < N$. Associated with source $i \in \{1, \ldots, N\}$, is a *codebook pointer* μ_i, and a parametric *expandor function* $\mathbf{h}_i$, with parameters $p_{i,1}, \ldots, p_{i,J}$. The code vectors $\mathbf{c}$ for encoding the i-th source are obtained from the codebook whose address is given by the value of μ_i. Each code vector $\mathbf{v}$ from the codebook is transformed by the expandor function, *viz.* $\mathbf{c} = \mathbf{h}_i(\mathbf{v}, p_{i,1}, \ldots, p_{i,J})$. Examples of simple expandor transformations are scaling, translation, and rotation, whose parameters are respectively a scalar gain, a vector offset, and a unitary matrix. In the cases of scaling and translation, it is fairly straightforward to modify the CSVQ algorithm to jointly design the shared codebooks, the pointers, and the expandor parameters [7]. In particular, the relatively simple gain-companding CSVQ enables the sharing of residual codebooks across all or a subset of levels of a tree. The shared codebooks are stored in a library and the tree nodes are populated with codebook pointers and gain parameters. For an 11-level TSVQ of a source of high-fidelity audio transform coefficient vectors [4], we have obtained a compression ratio of 4 : 1 with virtually no rate-distortion penalty. This ratio is relatively modest in comparision with the orders of magnitude of storage reduction obtained for very deep trees grown level by level with CSVQ [4]. Nevertheless, higher compression ratio could be attained if rotation is also incorporated into the companding, as suggested by the asymptotic results of Lee *et al.* [8]. However, the required rotation matrices increase both the storage and processing requirement so that the overall complexity performance tradeoff is not necessarily improved. This problem will be further explored.

[1] This work was supported in part by the the National Science Foundation, the State of California MICRO program, Rockwell International Corporation, Compression Labs, Inc., Hughes Aircraft Company, Eastman Kodak Company, and the Natural Sciences and Engineering Research Council of Canada.

References

[1] Y. Linde, A. Buzo, and R.M. Gray, "An Algorithm For Vector Quantizer Design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, January 1980.

[2] W.Y. Chan and A. Gersho, "Constrained Storage Quantization of Multiple Vector Sources by Codebook Sharing," *IEEE Trans. Commun.*, vol. COM-38, no. 12, pp. 11–13, Jan. 1991.

[3] W.Y. Chan, "The Design of Generalized Product Code Vector Quantizers," *Proc. Int. Conf. Acoust., Sp., & Sig. Proc.*, pp. III-389–392, San Francisco, March 1992.

[4] W.Y. Chan and A. Gersho, "Constrained-Storage Vector Quantization in High Fidelity Audio Transform Coding," *Proc. Int. Conf. Acoust., Sp., & Sig. Proc.*, pp. 3597–3600, Toronto, May 1991.

[5] P.A. Chou, "Code Clustering for Weighted Universal VQ and Other Applications," *Proc. IEEE Int. Sym. Info. Th.*, pp. 253, Budapest, June 1991.

[6] R.A. Lindsay and D.E. Abercrombie, "Restricted Boundary Vector Quantization," *Proc. Data Compression Conference*, pp. 159–165, Snowbird, Utah, April 1991.

[7] W.Y. Chan and A. Gersho, "Generalized Product Code Vector Quantization," in preparation.

[8] D.H. Lee, D.L. Neuhoff, and K.K. Paliwal, "Cell-Conditioned Multistage Vector Quantization," *Proc. Int. Conf. Acoust., Sp., & Sig. Proc.*, pp. 653–656, Toronto, May 1991.

SELF SYNCHRONISING T-CODES TO REPLACE HUFFMAN CODES

Gavin R Higgie, Electrical and Electronic Engineering, University of Auckland, Private Bag 92019, Auckland, New Zealand

Abstract - This paper describes recent work on the T-Codes, which are a new class of variable length codes with superlative self-synchronizing properties. The T-Code construction algorithm is outlined, and it is shown that in situations where codeword synchronization is important the T-Codes can be used instead of Huffman codes, giving excellent self-synchronizing properties without sacrificing coding efficiency.

Background

When corruption occurs in a stream of data which is coded with variable length codes, the decoder can lose track of where codeword boundaries are located in the data stream, and so the effect of the corruption can extend over a large number of received symbols. Variable length code sets can be chosen such that the receiver is able to determine the correct location of these codeword boundaries relatively quickly after a corruption, but this is usually done by choosing codes in which certain bit sequences occur only at the end of codewords. The receiver must look for these special sequences in the received data stream. In some cases it is possible to choose codes which will self-synchronize as a result of the normal decoding process, but these are relatively difficult to find.

T-Codes

The original discovery of the T-Codes was published in 1984 by Titchener [1]. This work gave an algorithm for generating families of T-Code code sets, and showed that all code sets generated in accordance with this algorithm are self-synchronizing by nature. These properties are not derived from having specific synchronizing bit sequences, but rather the synchronizing information is spread throughout the code as an inherent part of its construction. The construction of the codes is such that when codeword synchronization is lost as a result of a corruption, the receiver will automatically re-synchronize on a subsequent codeword boundary without any special measures being taken. This will happen even when the loss or corruption is not recognized as such. When a data loss or corruption is known to have occurred, an algorithm is available for the receiver to determine the point at which subsequent codeword synchronization is re-established.

T-Code Construction Algorithm

The T-Code construction algorithm is very simple. Code sets are constructed by *augmenting* lower level T-Code code sets, with the lowest level being the code set 0 and 1. The augmentation process consists of writing out a list with two copies of the lower level code set, and then sacrificing a codeword from the first half of the list and using it as a prefix for every codeword in the second half of the list. This produces a new code set which has nearly twice the number of codewords of the lower level code set. An example of this process is given in Table 1.

Level 0	Level 1 Prefix 0	Level 2 Prefix 01
0	--> -	-
1	1	1
	00	00
	01	--> -
		-
		011
		0100
		0101

Table 1
T-Code Construction Algorithm

T-Code Synchronization Properties

Titchener showed that every T-Code will have self-synchronizing properties, with typical synchronizing delays of a few codewords, but it was not until the work by Higgie [2] showed that these self synchronizing properties are available without loss of coding efficiency that the significance of the codes became evident. Any codeword from a code set can be used as the prefix to produce the augmented code set at the next level. The length of the prefix chosen affects the codeword length distribution of the next level code set, so by careful choice of prefix lengths it is possible to produce T-Codes which match the codeword length distribution required for efficiently coding any particular information source (i.e. effectively the same codeword length distribution as a Huffman code designed for the source).

Higgie's work [2] also showed that the T-Codes which give maximum efficiency for any particular information source generally include at least one which has an average synchronization delay of around 1.5 codewords.

Current Research Activity

The work reported in the paper by Higgie [2] used Monte simulation techniques to show that it is possible to choose an efficient and rapidly synchronizing T-Code for any particular application. These simulations also showed that not all T-Codes are equal in their synchronization performance and that the task of choosing the best T-Code for a particular application is not a trivial one. Attempts at justifying why some T-Codes are better than others have recently led to a new technique for theoretically determining the average synchronization delay of T-Codes when they are used efficiently. This technique offers several advantages over the previously used Monte Carlo techniques, and provides insight into how the T-Codes achieve their enviable synchronizing properties.

The theoretical technique is now being used in calculating a database of the T-Codes which have the best synchronizing performance. It is hoped that this database will be useful in enabling a user with a particular information source to choose a T-Code which will be as efficient as a Huffman code designed for the source, but with an average synchronization delay of about 1.5 codewords.

Current research is also focusing on the use of T-Codes in FAX machines and in the JPEG image compression standard, particularly with respect to transmitting images in these formats over mobile radio channels. This is only one of many potential application areas, as T-Codes can be used to advantage in any situation where the probability of data corruption is high enough to make the use of non-synchronizing or poorly synchronizing Huffman codes difficult.

Conclusion

The T-Code generation algorithm has been demonstrated to provide variable length code sets which have both the desirable properties of coding efficiency and rapid self-synchronization. For any particular information source, properly chosen code sets can typically offer average synchronization delays of 1.5 codewords without sacrificing coding efficiency compared to that obtained with a Huffman code designed for the source. This means that it is now possible to use variable length codes in applications where the probability of corruption is high and the problems of codeword synchronization have previous excluded their use.

References

[1] Titchener, M.R.: (1984) 'Digital encoding by means of new T-Codes to provide improved data synchronization and message integrity', Proc IEE, Pt E, Computers & Digital Techniques, Vol 131, (4), pp151-153

[2] Higgie, G.R. and Williamson, A.G.,: (1990) 'Properties of Low Augmentation Level T-codes', Proc. IEE, Pt E, Computers & Digital Techniques., Vol 137 pp129-132

Kieffer's Sample Converses for Source Coding

En-hui Yang
Department of Mathematics, Nankai University
Tianjin 300071, P R China

Abstract—New proofs of recent Kieffer's sample converses for source coding are given using a sample path covering idea originated by Ornstein and Weiss and modified by Shields together with Birkhoff's ergodic theorem.

Throughout the paper, we fix a measurable space $(A, \mathcal{A})$ as our source alphabet and a measurable space $(\hat{A}, \hat{\mathcal{A}})$ as our reproducing alphabet. For our purposes, a source μ is a stationary, ergodic process $\{X_n\}$ taking values in the alphabet A. If $x = (x_i)$ is a finite or infinite sequence from A or $\hat{A}$, let $x_m^n = (x_m, x_{m+1}, \cdots, x_n)$ and, for simplicity, write x_1^n as x^n. Let $\{\rho_n | n = 1, 2, \cdots\}$ be a fixed fidelity criterion in which each ρ_n is a measurable function from $A^n \times \hat{A}^n \to [0, +\infty)$. Recently, Kieffer[1] proved the following two sample converses for source coding.

Theorem 1 *Assume that the fidelity criterion $\{\rho_n | n = 1, 2, \cdots\}$ is subadditive, i.e., $(n+m)\rho_{n+m}((x_1, x_2), (y_1, y_2)) \le n\rho_n(x_1, y_1) + m\rho_m(x_2, y_2)$, and that there exists a $y^* \in \hat{A}$ for which $E\rho_1(X_1, y^*) < +\infty$. Let $R > 0$. Then for any sequence $\{B_n\}$ in which B_n is an nth-order block code with rate no greater than R, we have*

$$\liminf_{n\to\infty} \rho(B_n) \ge D(R), \quad a.e.$$

where $\rho(B_n) = \min_{y \in B_n} \rho_n(X^n, y)$ and $D(R)$ is the distortion rate function of the source μ relative to the fidelity criterion $\{\rho_n | n = 1, 2, \cdots\}$.

Theorem 2 *Assume that $\{\rho_n | n = 1, 2, \cdots\}$ satisfies $\rho_{n+m}((x_1, x_2), (y_1, y_2)) \le \max\{\rho_n(x_1, y_1), \rho_m(x_2, y_2)\}$ and that for each $D > 0$ there exist a countable subset $S_0 = \{y_i\} \subset \hat{A}$ and a countable measurable partition $\{F_i\}$ of A such that $\rho_1(x, y_i) \le D$, $x \in F_i$, for each $y_i \in S_0$, and $-\sum_i \Pr\{X_1 \in F_i\} \log \Pr\{X_1 \in F_i\} < +\infty$. Then for any sequence $\{C_n\}$ in which C_n is an nth-order variable rate code that operates at the distortion level D, we have*

$$\liminf_{n\to\infty} r(C_n) \ge R(D), \quad a.e.$$

where $r(C_n)$ is the sample rate function of C_n and $R(D)$ is the (operational) rate distortion function of μ relative to $\{\rho_n | n = 1, 2, \cdots\}$.

Theorems 1 and 2 are called the sample converse for source coding at a fixed rate level and the sample converse for source coding at a fixed distortion level, respectively. They are the first general sample converses in source coding theory. For the case of fixed distortion level, Barron[2] and Shields[3] proved Theorem 2 for the special case in which $A = \hat{A}$ is finite and $\rho_n \equiv 0$, and Ornstein and Shields[4] showed Theorem 2 for the special case in which $A = \hat{A}$ is finite and $\{\rho_n\}$ is the Hamming fidelity criterion.

Both proofs of Theorems 1 and 2 given in [1] involve to a great extent a powerful new ergodic theorem[5]. We present here new proofs of Theorems 1 and 2 that use only some simple code construction techniques together with Birkhoff's ergodic theorem and the sample path covering idea originated by Ornstein and Weiss[6] and modified by Shields[7]. In fact, the trick lies in how to use subtly the sample path covering idea. Since both new proofs of Theorems 1 and 2 are almost the same, in what follows, we only give the sketch of proof of Theorem 1.

The Sketch of Proof of Theorem 1: First note that using the code construction technique outlined in [8], we can deduce from the tradition block source coding theorem the following result: there exists a sequence $\{\bar{B}_n\}$, where $\bar{B}_n$ is an nth-order block code with rate no greater than R, such that

$$\limsup_{n\to\infty} \rho(\bar{B}_n) \le D(R), \quad a.e..$$

Suppose Theorem 1 is not true. Then there exist a $\epsilon > 0$, a $\gamma > 0$, and a sequence $\{B_n\}$, where B_n is an nth-order block code having rate R or less, such that

$$\Pr\{\cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{x \in A^\infty | \rho(B_n)(x^n) < D(R) - \epsilon\}\} > \gamma. \qquad (1)$$

Let $B_0 = \{y^*\}$ be a one order block code with $\rho(B_0) = \rho_1(X_1, y^*)$. In order to lead to a contradiction, we distinguish two cases: (1) for any positive real number a, $\Pr\{\rho(B_0) \ge a\} > 0$; (2) there exists a real a such that $\Pr\{\rho(B_0) \ge a\} = 0$. Since it is easier to prove Case (2) than Case (1), we confine ourself to Case (1). Let a be a positive real number to be specified later. Let $\delta = \Pr\{\rho(B_0) \ge a\}$. Obviously, $a\delta \to 0$ as $a \to +\infty$. Fix a positive integer $m (2/m < \delta)$ such that if $D_1 = \{x \in A^\infty | \rho(\bar{B}_m)(x^m) \le D(R) + \delta\}$, then $\Pr\{D_1\} > 1 - \delta$. Let $D = D_1 \cap \{x \in A^\infty | \rho(B_0) < a\}$ (using D instead of D_1 is a trick). In view of (1), fix $M > m/\delta$ and choose $N > M$ so large that if $C = \cup_{n=M}^{N} \{x \in A^\infty | \rho(B_n) < D(R) - \epsilon\}$, then $\Pr\{C\} > \gamma$. For sufficiently large n, define

$$G_n = \left\{ x \in A^\infty | (n-N+1)^{-1} \sum_{i=0}^{n-N} J_D(T^i x) > 1 - 2\delta \right.$$
$$\left. \& (n-N+1)^{-1} \sum_{i=0}^{n-N} J_C(T^i x) > \gamma \right\},$$

where T denotes the shift on A^∞ defined by $(Tx)_i = x_{i+1}$, and $J_D(x)$ and $J_C(x)$ are the indicator functions of D and C, respectively.

We next associate with each $x \in G_n$ a partition $\{I_i\}_1^{t(x)}$ of $[1, n]$ into consecutive sub-intervals. Assume $I_1, \cdots, I_{i-1}$ have been defined and $\cup_{j=1}^{i-1} I_j = [1, u-1]$. The $I_i (i \ge 1)$ is defined according to the following procedure:

- S1 If $T^{u-1}x \notin D \cup C$ or $u > n - N + 1$, then put $I_i = [u, u]$.
- S2 Otherwise, test the membership of $T^{u-1}x$ in C. If $T^{u-1}x \in C$, put $I_i = [u, v]$, where v is the least positive integer such that $\rho(B_{v-u+1})(x_u^v) < D(R) - \epsilon$.
- S3 Otherwise, test whether there exists $1 \le j < m$ such that $T^{u+j-1}x \in C$. If exists, put $I_i = [u, u]$; if not, put $I_i = [u, u+m-1]$.

The total number of all these partitions can be upper bounded by $2^{nH(\delta)}$. For each partition $\{I_i\}_1^{t(x)}$, construct an nth-order block code $B[\{I_i\}_1^{t(x)}] = B(I_1) \times \cdots \times B(I_{t(x)})$, where $B(I_i) = B_0$ if $|I_i| = 1$, $\bar{B}_m$ if $|I_i| = m$, and $B_{|I_i|}$ otherwise. Combining all these block codes together with B_0^n, we get a new nth-order block code B_n^* with rate less than $R + H(\delta) + 1/n$. From the above construction, we can deduce that for any $x \in G_n$,

$$\rho(B_n^*)(x^n) \le \delta a + D(R) + \delta - \gamma(\epsilon + \delta)/2 + n^{-1} \sum_{i=0}^{n-N} (1 - J_D(T^i x))\rho(B_0)(T^i x) + n^{-1} \sum_{i=n-N+1}^{n-1} \rho(B_0)(T^i x).$$

Since $D(R)$ is convex and nonincreasing, finally we can obtain

$$D(R) \le s(H(\delta) + 1/n) + \delta a + D(R) + \delta - \gamma(\epsilon + \delta)/2 + (NE\rho(B_0))/n + \int (1 - J_D(x))\rho(B_0)(x)\, d\mu + \int_{A^n - G_n} n^{-1} \sum_{i=0}^{n-1} \rho(B_0)(T^i x)\, d\mu$$

where $-s (s \ge 0)$ is the right derivative of $D(\cdot)$ at R. Letting $n \to \infty$ and then letting $a \to \infty$ lead to a contradiction.

REFERENCES

[1] J. C. Kieffer, *IEEE Trans. Inform. Theory*, vol. 37, pp. 263–268, 1991.
[2] A. R. Barron, "Logically smooth density estimation," Ph. D. thesis, Stanford University, Stanford, CA, 1985.
[3] P. C. Shields, *IEEE Trans. Inform. Theory*, vol. 37, pp. 1645–1647, 1991.
[4] D. S. Ornstein and P. C. Shields, *Annals of Prob.*, vol. 18, pp. 441–452, 1990.
[5] J. C. Kieffer, *Proceedings of the Conf. on A. E. Convergence in Prob. and Ergodic Theory*, Academic Press, pp. 151–166, 1991.
[6] D. S. Ornstein and B. Weiss, *Israel J. Math.*, vol .44, pp. 53–60, 1983.
[7] P. C. Shields, *IEEE Trans. Inform. Theory*, vol. 33, PP. 263–266, 1987.
[8] J. C. Kieffer, *IEEE Trans. Inform. Theory*, vol. 24, pp. 674–682, 1978.

Channel Error Control of Variable-Length Transform Coding

Ning Guo
Fraunhofer-Institute for Integrated Circuits
INEL, Wetterkreuz 13, 8520 Erlangen, Germany

Some Preliminary Descriptions

Recently, some video/audio compression algorithms of variable-length transform coding are suggested. However, a common problem associated with the use of variable-length source codes is that channel errors may cause the loss of synchronization, which leads to extended errors in the decoded text. In this paper, a trial-and-error algorithm with cross-length-checks has been proposed for the partial correction of these error propagations without recourse to error-correcting codes.

A block of N transform coefficients are divided into m groups of length n. Over these codeword lengths $cl_{i,j}$ (i=1,...,m; j=1,...,n), two sets of parity checks are defined as the volume check sums VCS_j ($j = 1, \cdots, n$) by

$$VCS_j = \sum_{k=1}^{m} cl_{k,j}, \quad (\text{mod } L)$$

and as the cross check sums CCS_i ($i = 1, \ldots, m+n-1$) by

$$CCS_i = \sum_{k=\max(1,i-n+1)}^{\min(i,m)} cl_{k,i-k+1}, \quad (\text{mod } L)$$

where L is an integer larger than the maximum codeword length of the variable-length code used.

Error Propagation Detection Procedure

The length in bits of coefficient groups are transmitted to the receiver to detect error propagations. After decoding n coefficients, the receiver is reset to the initial state so as to resynchronize the decoding procedure when error propagations occur in the past coefficient group. At the end of decoding each coefficient group, the test variable L_i ($i = 1, \cdots, m$) of the group length are evaluated by

$$L_i = \begin{cases} 1, & \text{if the total number of bits read for the i-th group does not match the transmitted group length;} \\ 0, & \text{else.} \end{cases}$$

L_i=1 means that channel errors have occurred in the i-th coefficient group and caused error propagations.

Indicating the Suspected Causal Codewords

The reason for the occurrence of an error propagation is that the first codeword in the error propagation is changed by erroneous bits into an another codeword of different length, and the following text can be decoded as a number of coefficients, which are different from the originals. In the view of this, if we can find these causal codewords and determine their length, we can correct the error propagations by resynchronizing the decoding procedure. For this reason, two sets of test variables, the volume test variables V_i ($i=1,\cdots,n$) and the cross test variables C_i ($i=1,\cdots,n$), are calculated by

$$V_i = \begin{cases} 1, & \text{if the } VCS_i \text{ calculated at the receiver does not match that received;} \\ 0, & \text{else;} \end{cases}$$

and

$$C_i = \begin{cases} 1, & \text{if the } CCS_i \text{ calculated at the receiver does not match that received;} \\ 0, & \text{else;} \end{cases}$$

with

$$V_{min} = \min (i: V_i = 1) \text{ and } C_{min} = \min (j: C_j = 1)$$

It is affirmative that all the codewords which are the V_{min}-th or (C_{min}-i+1)-th codeword and at the same time in the groups with L_i=1 are regarded as the suspected causal codewords. They can be divided into two sets and their index sets represented by

$$S_1 = \{ (i , V_{min}) : L_i = 1 \text{ and } i \geq \alpha \},$$

and

$$S_2 = \{ (i , C_{min} - i+1) : L_i = 1 \text{ and } i < \alpha \},$$

where α can be obtained by

$$\alpha = C_{min} - V_{min} + 1.$$

Determining the length of Suspected Causal Codewords

In order to resynchronize the decoding procedure, the length of the suspected causal codewords must be evaluated by using the received VCS's, CCS's. For $cl_{i,j}$ in S_1, the codeword length $L_v(i)$ is

$$L_v(i) = VCS_{V_{min}} - \sum_{\substack{k=1 \\ k \neq i}}^{m} cl_{k,V_{min}} \quad (\text{mod } L),$$

and for $cl_{i,j}$ in S_2, $L_c(i)$ is

$$L_c(i) = CCS_{C_{min}} - \sum_{\substack{k=\min(1,\, C_{min}-n+1) \\ k \neq i}}^{\min(C_{min},\, m)} cl_{k,C_{min}-k+1} \quad (\text{mod } L).$$

Main Procedure

The suspected causal codewords are searched with the check sums VCS's and CCS's. For one of them, the length $L_v(i)$ or $L_c(i)$ is evaluated, the decoding of the associated coefficient group is resynchronized and performed once again. The length check L_i is used to determine the success of the correction. This trial-and-error procedure is iteratively performed for all suspected causal codewords.

INTEGRATED INDEX ASSIGNMENT, SOURCE AND CHANNEL CODING

Petter Knagenhjelm
Department of Information Theory
Chalmers University of Technology

A sample recursive method of co-optimizing source and channel coding in digital transmission of analog signals is presented. The procedure is an iterative approach addressing both the adjustment of the reproduction vectors according to channel errors, as well as the problem of index assignment.

1. Introduction

Vector quantization (VQ) plays an important role as a source encoder in digital transmission of analog signals. The difficulty of designing high dimensional VQs is a severe obstacle for practical usage. Besides search and storage difficulties, the two main problems in the design are *how* to distribute the reconstruction vectors over the source-space, and *how* to choose the code words, or indices, so that the effect of channel errors on the performance is minimized. Traditionally, the two problems are treated separately, but there is a current trend to regard them as one and minimize the distortion at the receiver using zero-redundant VQs, i.e., the additional bits normally incorporated for error protection are employed to refine the quantizer without explicit error protection. A VQ trained with the LBG algorithm can be very sensitive to channel errors due to its tendency for random ordering of the indices. Index ordering, known as the Index Assignment (IA) problem is an important part of the VQ design. Unfortunately, finding the optimal IA belongs to the class of NP-complete problems.

IA is discussed for scalar quantizers in [1] More recent work is[2] or [3] and [4] where the IA is a post-process to the VQ design. In [5] the IA is incorporated in the LBG algorithm and thereby is a co-optimization of the source- and of the channel-coding.

One difficulty with Channel Optimized VQ (COVQ) is that the channel error probability is a *design parameter* in the optimization. In a real transmission situation, this parameter is difficult to estimate. It may even vary in time, making the design according to a specific value rather academic. More important is how *robust* the design is to a mismatch between the actual error probability, q, and the design parameter ε.

2. The method

The iterative approach suggested in this paper is thoroughly described in [6] and is only recapitulated here. Let $\alpha_i(\mathbf{x})$ denote the total squared error distortion associated with choosing the i:th of the $M = 2^k$ reconstruction vectors, $\mathbf{y}_i$, given an observation $\mathbf{x}$, i.e.,

$$\alpha_i(\mathbf{x}) = \sum_{j=0}^{M-1} \|\mathbf{x} - \mathbf{y}_j\|^2 \cdot p_{j|i} \qquad (1)$$

The total distortion to be minimized can be written

$$D = \sum_{i=0}^{M-1} E\left[\alpha_i(\mathbf{X}) \middle| \mathbf{X} \in K_i\right] \cdot P_i \qquad (2)$$

where the optimal partitioning of the signal space gives the regions

$$K_i = \left\{\mathbf{x} \in \mathbf{R}^d : \alpha_i(\mathbf{x}) \le \alpha_j(\mathbf{x}) \; \forall \; j\right\} \qquad (3)$$

where $p_{j|i}$ is the probability of receiving the index j given that i was sent and where P_i is the probability of $\mathbf{X} \in K_i$.

The iterative algorithm adjusts all reproduction vectors after each observation X_n. The reproduction vector causing minimum expected distortion at the receiver is voted winner and is denoted $\mathbf{y}_I$.

$$I = \arg\min_i \{\alpha_i(X_n)\} \qquad (4)$$

The individual adjustments, or step sizes γ'_j, depend on the probability of interchanging, due to channel errors, the code word j with the winner I.

$$\gamma'_j = f(t) \cdot p_{j|I} \qquad (5)$$

where $f(t)$ is an annealing function with $f(T) = 0$, T being the predetermined training time.

The sample vector updating formulas, in conjunction with steepest descent, become

$$\mathbf{y}'^{+1}_j = \mathbf{y}'_j + \gamma'_j \cdot \left(X_n - \mathbf{y}'_j\right) \quad \forall \; j \qquad (6)$$

3. The structure of a robust quantizer

A VQ optimized according to (3) with $\varepsilon > 0$, becomes more conservative, in the meaning that the reproduction vectors are shifted towards the center of mass of the information source, tha a VQ trained without bit errors as can be seen e.g. in fig 1b. and 1c. In this figure, code vectors are connected with a line if their code words are Hamming-1 neighbors. Fig 1a) and c) show the VQs designed with $\varepsilon = 0.00$ and $\varepsilon = 0.05$ respectively. The structural ordering in c) compared to a) is striking and is decisive for the robustness if $q \neq \varepsilon$ (i.e. design mismatch). Figure 1b) shows the same VQ as in a) but with an ordering of the indices. Fig 1d) shows the inherent robustness of a good IA to design mismatch

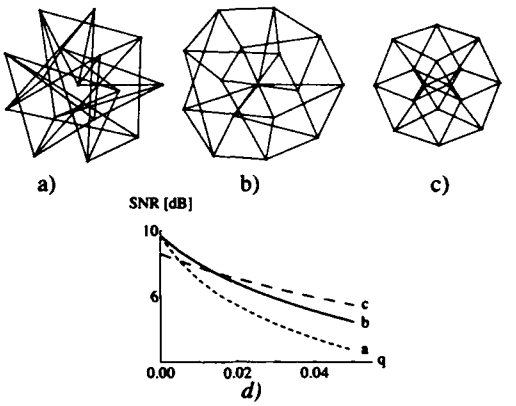


Figure 1 a-d) Examples of the robustness to error mismatch for three VQs. The VQs in a) and b) have identical reconstruction vectors, trained with ε=0, but an IA procedure is added to the design in b). Hamming-1 neighbors are connected with a line. The VQ in c) is trained with ε=0.05. d) shows the performance of a-c), for a BSC when the channel error varies from q=0 to q=0.05.

4. Experiments

The capability of the algorithm is thoroughly investigated using a number of Gauss-Markov sources in 2,3,4 and 8 dimensions. The following table presents a few of the results, comparable to the tests in [5], at the rate 1 bit/sample.

Table 1. Experimental results obtained using 2 and 4 dimensional Gauss-Markov sources with correlation factors of 0.0 and 0.9. The training/evaluation sets are large ($2 \cdot 10^6$ / $1 \cdot 10^5$) and the channel is binary symmetric.

N	Bit error	Corr=0.0	Corr=0.9
2	0.00	4.40	7.91
	0.05	3.15	4.71
	0.10	2.27	3.34
4	0.00	4.66	10.2
	0.05	3.15	6.05
	0.10	2.28	4.52

5. Conclusions

The optimization method presented in this paper has proved to yield structured solutions, close to linear mappings of the hyper cube, which offers both a good VQ **and** an accurate IA. The design parameter can during training, be decreased to zero and thereby obtaining a robust VQ due to the incorporated IA. The results obtained, points out that the algorithm works favourably compared with others.

6. References

[1]N. Rydbeck and C.-E. Sundberg, "Analysis of Digital Errors in Nonlinear PCM Systems." *IEEE Trans. Commun.*, vol. Com-24, no. 1, pp. 59-65, January 1976.

[2]J.R.B.D. Marka and N.S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multidimensional quantizer.", *Proc. IEEE Int. Comn. Conf.*, Seattle, WA, June 1987, pp. 1128-1132.

[3]N. Farvardin, "A Study of Vector Quantization for Noisy Channels." *IEEE Trans. Inform. Theory*, vol. IT-36, no. 4, pp. 799-809, July 1990.

[4]K. Zeger and A. Gersho, "Pseudo-Gray Coding." *IEEE Trans. Commun.*, vol. Com-38, no. 12, pp. 2147-2158, December 1990.

[5]N. Farvardin and V. Vaishampayan, "On the Performance and Complexity of Channel-Optimized Vector Quantizers." *IEEE Trans. Inform. Theory*, vol. IT-37, no. 1, pp. 155-160, January 1991.

[6]P. Knagenhjelm, "A Recursive Design Method for Robust Vector Quantization", *Proc. International Conference on Signal Processing Applications and Technology*, Boston, November 1992.

A New Deterministic Codebook Structure for CELP Speech Coding

Yu-Hung Kao*, John S. Baras†
University of Maryland
College Park, MD 20742

Abstract

Low bit rate, high quality speech coding is a vital part in voice telecommunication systems. The introduction of CELP (1984, Codebook Excited Linear Prediction) speech coding provides a feasible way to compress speech data to 4.8 kbps with high quality. However, the formidable computational complexity required for real-time processing has prevented its wide application. Using our codebook, we reduce the computational complexity of codebook search, which originally accounts for 2/3 of the computational complexity, to almost nothing; while preserving the same good speech quality. This tremendous reduction in computational complexity is achieved by replacing the traditional **stochastic codebook** by an artificially constructed **deterministic codebook**. After a careful study of the minimization of vector quantization distortions, we found that although "randomness" is usually observed in speech residuals; it is not necessary to use a noise-like stochastic codebook to encode the speech residuals. As long as the code vectors were distributed uniformly over a sphere, very small VQ errors can be achieved. The most significant advantage of using this deterministic codebook is extremely fast codebook search. After this reduction, we have an algorithm about 5 MIPS. It can be handled by even inexpensive DSP chips, while maintaining the same high quality. Besides extremely simple encoding and decoding schemes, this codebook also provides optimal error tolerance property and it doesn't require codebook storage. We hope our contribution can finally make CELP speech coding a widely applicable technology.

*Texas Instruments
†Martin Marietta Chair in Systems Engineering

LOSSLESS COMPRESSION ALGORITHMS FOR HIGH FIDELITY AUDIO COMPRESSION

Talal Shamoon and Chris Heegard
School of Electrical Engineering
Cornell University
Ithaca, New York 14853

1.1 *Summary*

Real-time algorithms for the compression of high-fidelity audio are presented. The goal of these algorithms is to provide a compact, high fidelity, digital representation for an input stream of audio samples. We are developing an adaptive transform coding system that consists of five specialized functional blocks: An octave-based subband decomposition signal transformer [1,2], a bank of adaptive quantizers assisted by a bit allocator, and a lossless compressor coupled with a buffer.

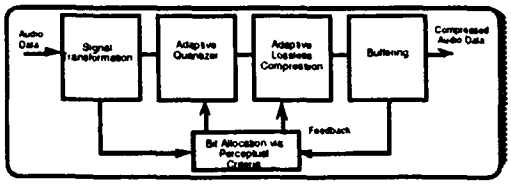


Figure 1. High-Fidelity Audio Compression System

This paper concerns fast adaptive algorithms for the lossless compression stage. While the techniques described herein can be applied to any type of data, we specifically seek efficient compression of digital audio. We also give results from entropy estimation experiments for a digital audio source.

A variable-to-variable length algorithm for lossless compression is presented. This algorithm consists of two stages: A variable-to-fixed length coder that is based on tree-building algorithms in the style of Lempel-Ziv, followed by a fixed-to-variable length arithmetic coder [3]. The first stage parses source symbols into a fixed number of frequently occurring strings. This *dictionary* of strings varies over time, and an adaptive procedure that tracks the source's behavior is outlined in the talk. An arithmetic coder takes the labels of these strings, and their associated frequencies of occurrence from the dictionary coder, and generating a compressed, variable length, representation. We compare the performance of this strategy against that of a variable-to-fixed coder working alone, and that of a fixed-to-variable working alone. While variable-to-fixed coders are effective at exploiting correlation between source outputs, they are limited in practice by limitations on data structure size (finite table length.) Limitations on compression caused by this phenomenon are obviated by the introduction of the arithmetic coder. In a sense, the structure that we propose resembles the paradigm of a statistical coder (the arithmetic coder) coupled with an adaptive source modeler. Similar structures that use Dynamic Markov Modeling for the source modeler have been proposed by [3,4]. These schemes are hampered by the fact that the adaptation algorithms for developing the source model can cause it to grow unmanageably. While the variable-to-fixed length front-end of our system implicitly models the source, the data structure evolves to accommodate source variations in a bounded fashion. Results of our technique applied to digital audio requantized via ADPCM quantizer banks discussed above are provided.

Further, we provide results of entropy estimation experiments performed on a digital audio source. These were performed to better understand the behavior of these sources. Our estimator is based on an extension of the techniques proposed in [5]. This work forms the basis for designing the variable-to-fixed coder described above. The entropy estimates are compared to entropy estimates calculated using marginal probabilities from the same source output.

In short, we illustrate an efficient compression strategy that allows us to pick up extra compression gain in our audio compression system. This strategy is adaptive and dynamically changes to respond to idiosyncracies in the non-stationary audio source that we are tracking.

1.2 *References*

[1] Talal Shamoon and Chris Heegard, "Audio Compression via Wavelets and Multiresolution Analysis," *Proceedings of the 25th Annual Conference on Information Sciences and Systems*, pp. 902-906, March 1991.

[2] Chris Heegard and Talal Shamoon, "High Fidelity Audio Compression: Fractional-Band Wavelets," *Proceedings of 1992 ICASSP*, pp. II-201 - II-203, March 1992.

[3] Timothy Bell, John Cleary, and Ian Whitten, *Text Compression* Englewood Cliffs, NJ, Prentice-Hall, 1990.

[4] G.V. Cormack and R.N.S Horspool "Data Compression Using Dynamic Markov Modeling," *Computer Journal*, 30 (6), pp541-550

[5] Frederick Jelinek and Kenneth Schneider "On Variable-to-Length-Block Coding," *IEEE Transactions on Information Theory*, vol IT-18, no.6 pp765-774, November 1972.

[This work was supported by NSF grants NCR-8903931 and NCR-9207331]

CONSTELLATIONS DESIGNED FOR THE RAYLEIGH FADING CHANNEL

X. Giraud - K. Boullé - J.C. Belfiore
TELECOM PARIS - 46, rue Barrault - 75634 Paris cedex 13 - FRANCE

The error probability of the usual linear modulations (M-PSK or M-QAM) in the Rayleigh fading channel (RFC) varies as the inverse of the signal to noise ratio. To increase the slope of the error curve, a diversity technique or an error correcting code combined with interleaving can be added. The diversity systems are spectral efficiency costly in case of frequency diversity, or involve additional complexity if multiple antennas are used. Trellis Coded Modulations (TCM) are an efficient way of achieving good performance without spectral efficiency loss. However, the construction of well suited TCM codes becomes a very difficult task when M-QAM modulation (M>16) schemes are used. Here, we consider the design of constellations matched to the RFC. We search for n dimensional ($n \geq 2$) lattices which can provide a diversity of order n in the RFC, without the addition of diversity techniques or TCM.

The main features of our approach are very much the same as for the AWGN channel; we first determine a metric measuring channel symbols insulation as Euclidean distance does in the AWGN channel. A careful appreciation of what a constellation matched to a channel means leads us to define a theoretical frame for the lattice coding problem : instead of searching packing lattices, we look for "admissible" lattices with respect to a body specific of the considered channel, a concept derived from the geometry of numbers . At high SNR, the distance function of the RFC simplifies and becomes $d(\mathbf{x},\mathbf{y}) = |N(\mathbf{x}-\mathbf{y})|$ where $N(\mathbf{x}) = \prod_{i=1}^{n} x_i$; we address the lattice coding problem in that case. The corresponding body is homothetic to $S = \{\mathbf{x} \setminus d(\mathbf{x},\mathbf{0}) \leq 1\}$ and we look for S-admissible lattices. The distance function shows that an S- admissible lattice should not possess two vectors that have the same value in any component; this feature should provide an n^{th} order diversity. Under certain conditions N is the algebraic norm of some real number field K of degree n; in that case, the embedding of the ring of algebraic integers of K in $\mathbf{R}^n$ provides an S-admissible lattice. Hence, number field theory enable us to define a procedure to find dense n dimensional lattices matched to the RFC : 1) find a totally real algebraic number field K of degree n with a small absolute discriminant; 2) determine an integer basis of K. The densest n dimensional lattices using this technique are known when $2 \leq n \leq 8$. Hence, we have ready a family of n dimensional lattices which provide a n^{th}-order diversity in the RFC. For a normalised rate of $\rho = 2$ bits/dim, the gain is in the range of 10 to 15 dB, at a symbol error rate of 10^{-3}, compared to 16-QAM. Besides, ρ can be easily increased as any subset of an S-admissible lattice match the RFC. Finally we address detection; the maximum likelihood decoder selects the channel symbol minimising the channel metric. A detection algorithm is presented which provides the same performance as the exhaustive search.

MULTILEVEL TRELLIS MPSK MODULATION CODES FOR THE RAYLEIGH FADING CHANNEL[1]

Jiantian Wu
Simon Fraser University
Burnaby, B.C., Canada V5A 1S6

Shu Lin
University of Hawaii at Manoa
Honolulu, Hawaii 96822, U.S.A.

The error performance of a trellis modulation (or TCM) code over the Rayleigh fading channel depends strongly on the minimum symbol and product distances of the code[1-2]. Both these distances should be as large as possible, and they play different roles in determining the error performance of the code. At low SNR, the minimum product distance is more important; whereas at high SNR, the minimum symbol distance becomes more important. Apart from these two distance parameters, the path multiplicity (or error coefficient) is also an important factor in determining the error performance of a code at low SNR, and it should be kept as small as possible.

The multilevel coding method devised by Imai and Hirakawa[3] is a powerful technique for constructing bandwidth efficient modulation codes from Hamming distance component codes in conjunction with proper bits-to-signal mapping through set partitioning[4]. This method has been used to construct both trellis and block modulation codes for the AWGN channel. In this paper, the multilevel coding method is used to construct multilevel multi-dimensional trellis MPSK modulation codes for the Rayleigh fading channel. This method allowed us to coordinate all the important parameters of a code such that no single parameter severely degrades the performance of the code. A specific construction method is proposed. In this method, the minimum symbol and product distances of a multilevel trellis MPSK code are expressed in terms of the minimum Hamming distances of its component codes and the intra-set distances of the signal constellation and its subspaces. In the construction of a code, all the factors which affect the code performance and its decoding complexity are considered. Good codes have been constructed. The error performances of some of these codes based on both one-stage optimum decoding and multi-stage suboptimum decoding have been simulated. The simulation results show that these codes achieve good error performance with small decoding complexity.

As an example, consider the two-level coding scheme in which the first component code is a two-state four-dimensional binary trellis code and the second component code is a 4-state eight-dimensional QPSK trellis code. To construct the first component code, the single-parity-check (4, 3, 2) code is partitioned into 4 cosets by the repetition (4, 1, 4) code as follows: $A_0 = \{0000, 1111\}, A_1 = \{1100, 0011\}, A_2 = \{1010, 0101\}$, and $A_3 = \{0110, 1001\}$. The inter-set distance between A_i and A_j $(i \neq j)$ is 2, and the intra-set distance of A_i $(i = 0, 1, 2, 3)$ is 4. Using a two-state trellis code with trellis structure as shown in Figure 1, the resultant code has minimum Hamming distance 4. The second component code is formed as follows: Use a one-to-one mapping from QPSK signal set to $GF(4)$. A four-dimensional set partition chain can be obtained by using extended Reed-Solomon codes. Let $RS(n, d)$ denote a Reed-Solomon code with block length n and minimum Hamming distance d. Form a set partition chain, $RS(4,1)/RS(4,2)/\ RS(4,3)/RS(4,4)/\{0\}$, where 0 is the all-zero vector. A linear, partial-unit-memory four-state trellis code over $GF(4)$ can be obtained as shown in Figure 2 where D denote a buffer with an unit-time delay. This code has Hamming distance 3. Combining the above two trellis codes, we obtain an 8-state eight-dimensional trellis 8-PSK code with minimum symbol distance 3, minimum product distance 8, and information rate 2 bits/symbol. Each state transition has 32 parallel branches. Figure 3 shows the simulation results on bit error performance of the code. The performance of the code is better than the 8-state Ungerboeck code at $E_b/N_o > 13$ dB. The normalized branch complexity of this code is only half that of the 8-state Ungerboek code. In Figure 3, we also include the simulation results of Divsalar and Simon's 4-state four-dimensional code with R=2.0 bits/symbol[2]. It turns out that the performance of Divsalar and Simon's code is worse than that of 8-state Ungerboeck's code although its symbol distance is not less than that of Ungerboeck code. This is because the Ungerboeck code has a better distance spectrum.

[1]This research was supported by NSF Grants NCR-9115400, BCS-9020435, and NASA Grant NAG 5-931

REFERENCE

1. D. D. Divsalar and M. D. Simon, " The Design of Trellis Coded MPSK for Fading Channels: Performance Criteria", **IEEE Trans. Commun**, Vol. COM-36, pp. 1004-1012, Sept., 1988.
2. D. D. Divsalar and M. D. Simon, " The Design of Trellis Coded MPSK for Fading Channels: Set Partitioning of Optimum Code Design", **IEEE Trans. Commun**, Vol. COM-36, pp. 1013-1022, Sept., 1988.
3. H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error Correcting Codes," **IEEE Trans. on Information Theory**, Vol. IT-23, N0. 3, pp. 371-376, May 1977.
4. G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," **IEEE Trans. on Information Theory**, Vol. IT-28, No. 1, pp. 55-67, January 1982.

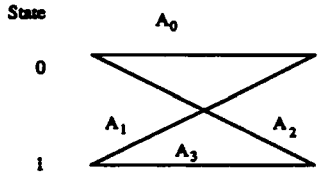


Figure 1. The trellis structure of a 2-state rate-1/2 trellis code

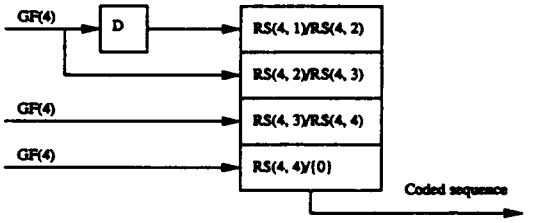


Figure 2. An encoder of a 4-state convolutional code over GF(4)

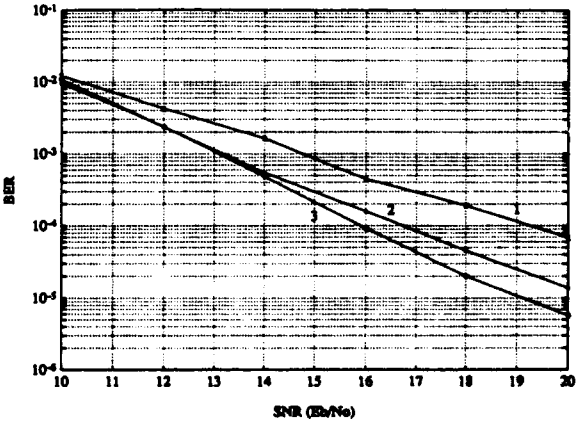


Figure 3 The simulation results of the BER performance: Curves 1, 2, and 3 are the performances of the 4-state Divsalar and Simon's code, the 8-state Ungerboeck code, and the example code respectively.

PERMUTED MODULATION AND CODED MODULATION FOR INTERLEAVED SLOW FADING CHANNELS

François GAGNON

Dept. of Electrical Engineering, École de technologie supérieure
4750, Henri-Julien, Montréal (Québec), Canada H2T 2C8

Abstract

A new scheme is presented to improve the error performance of modulation and trellis coded modulation on slow fading channels. This technique consists in permuting coordinates of multidimensional constellations on interleaved channels. This permutation insures that each coordinate of the signal is faded independently.
Permutation alone does not insure good improvements, the choice of a particular rotation of the signal set is also critical to obtain performance gains. A number of uncoded and coded modulation schemes have been found to be improved with permutation. Theoretical and simulation results show that this simple permutation provides gains of up to 4 dB with uncoded modulation. For trellis coded modulation, gains of 5 dB were achieved for a 64 state rate 3/4 convolutional code and 16 QAM modulation.

Summary

In this paper, we present a scheme that improves the error performances for slow fading channels. This technique is used with interleaving, it consists in permuting each coordinate of the transmitted signals. This permutation may be viewed as an interleaving at the level of individual coordinates. It improves the error performances by transmitting signals such that they are not completely deteriorated by fades.

Permutation of coordinates may be used with or without coding. It is simply a different kind of interleaving that provides up to 5 dB gains as compared to usual interleaving. This technique has been explored both theoretically and with computer simulations. It is shown that particular rotations of the signal set give good performances. Hence some care must be taken when choosing a coded or uncoded modulation technique when a permutation of coordinates is used.

I. Permutation of Coordinates for Uncoded Modulation

The technique under study is most easily described as an interleaving of the individual coordinates of subsequent signals to be transmitted. If, for example, we have signals (X_1,Y_1), (X_2, Y_2), (X_3, Y_3), (X_4, Y_4), taken in a constellation of two dimensions. After interleaving, the signals actually being sent may be (X_1, X_2), (X_3, X_4), (Y_1, Y_2), (Y_3, Y_4). Once received, the new signals are corrupted by fades and noise, the coordinates are then reordered before the usual decisions are taken. If the interleaving is such that each coordinates of a pair (X_i, Y_i) are faded independently, the error performance may be significantly improved. If fades are independent on each coordinate, the average error performance is improved since averaging is carried out with two independent variables instead of one. In this paper, two-dimensional constellations are used throughout but the scheme is easily applicable to other multidimensional constellations.

The usefulness of permutation is now presented by deriving bounds on the error performance with and without permutation. For high average signal-to-noise ratios, γ_o, the error probabilities of 16 QAM signals transmitted over a Rayleigh fading channel are respectively approximated as :

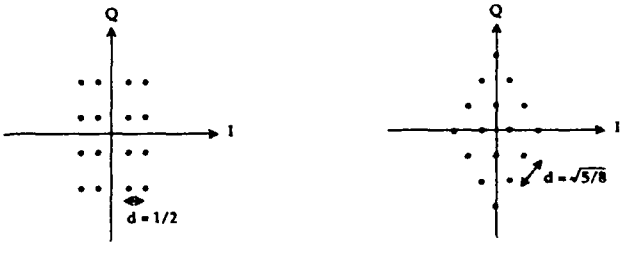
$$\bar{P}e \approx \frac{3}{2}\left(1-\sqrt{\frac{\gamma_o}{10+\gamma_o}}\right) \quad (1) \qquad \bar{P}e \approx \frac{9}{8}\left(1-\sqrt{\frac{\gamma_o}{5+\gamma_o}}\right) \quad (2)$$

where eq.(1) holds without permutation and eq.(2) is used for indepedent fades on each coordinate (i.e. permuted coordinates) and a 45° rotation of the signal set, as in Figure 1 b). As shown on Figure 1, this rotation improves the error probabilities by increasing the distances between received faded signals. By comparing both equations, we find that more than 4dB is gained with a permutation of coodinates. The same technique may be used for 8 PSK modulation, and the energy gain provided by the permutation is now greater than 3 dB. It is to be noted that a rotated 16 QAM signal set, once permuted gives a 49 QAM transmitted signal. Hence, the performance improvement may also be viewed as a consequence of an increase in the complexity of the signal set.

II. TCM with Permutation of Coordinates

With Trellis Coded Modulation (TCM), the same technique may be used to improve performances. Many 64 state TCM schemes [1] were simulated for fully interleaved slow Rayleigh fading channels. Two results are of particular interest, at a BER of 10^{-4}, the permutation of coordinates for the rate 2/3, 8 PSK, TCM provides an additional 1.8 dB gain and for the rate 3/4, 16 QAM, TCM a 5 dB improvement is obtained. For the rate 3/4, 16 QAM TCM scheme the best gain is obtained without rotation. Note that without coding, a rotation of the signal set was necessary to improve the average intersignal distance. It is quite surprising that such a rotation is not needed with coding. This may be explained by an inappropriate mapping with rotated signals, but other mappings were simulated without improving these performances. The 16 QAM performances are particularly noteworthy, since permutation without rotation does not change the transmitted signal constellation. Hence by simply interleaving the coordinates, improvements of 5 dB are possible, without changing other aspects of the transmission link.

[1] G. Ungerboeck, "Channel Coding with Multilevel/ Phase Signals", IEEE Trans. on Inf. Th., IT-28, January 1982.



a) Without Rotation b) With a 45° Rotation

Figure 1 : Received Constellation Without Noise when Only One Component is Affected by a Fade. The Amplitude of the I Component is Reduced by 1/2.

Unified Analysis on Performance Limits of Coded Multilevel DPSK in Rayleigh Fading Channels

Tadashi MATSUMOTO, Member, IEEE
and
Fumiyuki Adachi, Senior Member, IEEE

R&D Department of NTT Mobile Communications Network Inc.,
1-2356, Take, Yokosuka, 238, Japan

Abstract

This paper analyzes performance limits of coded multilevel differential PSK (MDPSK) in frequency selective Rayleigh fading channels. It is assumed either that interleaving degree is large enough, or that there is a sufficient bandwidth for frequency hopping, to randomize the burst errors produced by fading. The channel cutoff rate of MDPSK is calculated based on the "Gaussian metric"; AWGN, co-channel interference and multipath channel delay spread are taken into account. For practical, reliable communications over cellular mobile radio systems employing coded MDPSK, the three optimal information bit rates that achieve

1)minimum required average signal energy per information bit-to-noise power spectral density ratio (E_b/N_0),

2)maximum tolerable rms delay spread τ_{rms}, and

3)maximum spectrum efficiency

are determined from the channel cutoff rate. It is shown that without fading frequency selectivity, the optimal information bit rate (=information bits /MDPSK symbol) which minimizes the required average E_b/N_0 is around 0.25 information bits /symbol for 2DPSK, and 0.4 bits /symbol for other MDPSK schemes with $M \geqq 4$.

In frequency selective fading, the lower the rate of codes for error correction, the higher the channel symbol rate for a given information bit rate $1/T_b$, and the transmission performance becomes more sensitive to the fading frequency selectivity. It is shown that a larger τ_{rms}/T_b value can be tolerated with larger values of M. The optimal code rate for 32DPSK is around 0.3 (1.5 information bits /symbol), and the maximum τ_{rms}/T_b value is 1.5.

In co-channel interference environments, it is obvious that a larger error correction capability reduces required average signal-to-interference power ratio (SIR). Therefore, the same frequency can be used in closer cells when lower rate codes are used. This increases the system spectrum efficiency. However, the lower rate codes require larger transmission bandwidth, and this decreases the efficiency.

In the analysis of the spectrum efficiency of cellular mobile radio systems employing coded MDPSK, the service area is defined as the area in which practical, reliable communications are possible. It is shown that for a given channel bandwidth, the spectrum efficiencies are maximized when the information bit rate is around 0.5 information bits /symbol for 2DPSK, 1 bit /symbol for 4DPSK, and 1.4 bits /symbol for other MDPSK schemes. For a given information bit rate, spectrum efficiency is increased with larger M values. The optimal code rate for 32DPSK is around 0.3, and the maximal spectrum efficiency is 2.6 times as large as that for 1 bit-per-symbol coded 4DPSK.

PERFORMANCE OF JOINT EQUALIZATION AND TRELLIS-CODED MODULATION ON MULTIPATH FADING CHANNELS*

Mao-Ching Chiu *Chi-chao Chao*
Department of Electrical Engineering
National Tsing Hua University
Hsinchu, Taiwan 30043, R.O.C.

Abstract

In this paper an upper bound on bit error probability of a maximum-likelihood sequence estimation (MLSE) equalizer for trellis-coded modulation (TCM) systems with diversity reception is derived for multipath Rayleigh fading channels. Analytical and simulation results show that our bound is good for all cases considered especially when diversity reception is used.

Summary

Sequence estimation and trellis-coded modulation are effective means to combat channel impairments such as intersymbol interference (ISI) and channel noise. It is known that MLSE joint with decoding in the maximum-likelihood sense is the optimum way to detect TCM signals on ISI channels. On frequency-selective fading channels, the performance of MLSE for TCM systems were analyzed in [1] [2] [3] under the assumption that fading is so slow that the channel remains fixed during an entire error event. In reality, the fade speed of the channel is closely related to the autocorrelation function of the time-variant channel impulse response. In this paper, a general upper bound on bit error probability of MLSE for TCM systems on multipath Rayleigh fading channels is derived.

We assume that diversity reception is available, so the whole channel is modeled as D independent fading channels corrupted by i.i.d. complex white Gaussian noise. Let $g_{k,i}^d$, $0 \le i \le L$, $1 \le d \le D$, denote the ith tap coefficient of the dth diversity branch at time k in the equivalent discrete-time channel model.. Let x_k be the output of the TCM encoder at time k and y_k^d be the corresponding output of the channel at the dth diversity branch. Then

$$y_k^d = \sum_{i=0}^{L} x_{k-i} g_{k,i}^d + \eta_k^d,$$

where $\{\eta_k^d\}$ are i.i.d. zero-mean complex Gaussian random variables with variance $\sigma_{\eta_k^d}^2 = (1/2)\mathrm{E}\{|\eta_k^d|^2\} = N_0$. Since the channel is assumed to be Rayleigh faded, tap coefficients $\{g_{k,i}^d\}$ are modeled as independent (in terms of indices i and d) zero-mean complex Gaussian random variables (but which are correlated in index k.)

Let $\mathbf{v} = \{v_k\}$ be the transmitted information sequence and $\mathbf{v} \oplus \mathbf{e} = \{v_k \oplus e_k\}$ be the information sequence at the receiver output. Since $\{x_k\}$ is the transmitted signal sequence of the information sequence $\{v_k\}$, let $\{x_k + \varepsilon_k\}$ denote the corresponding signal sequence of $\{v_k \oplus e_k\}$. By employing the union-bounding technique, the average bit error probability for an MLSE receiver can be bounded by

$$P_b \le \frac{1}{n} \sum_{\mathbf{e} \in \mathbf{E}} W_b(\mathbf{e}) \sum_{\mathbf{v}} \mathrm{P}\{\mathbf{v}\} \mathrm{P}\{\Gamma(\mathbf{v} \oplus \mathbf{e}) \ge \Gamma(\mathbf{v}) | \mathbf{v}\},$$

where $\mathbf{E}$ is the set of all error events, $W_b(\mathbf{e})$ is the number of bit errors of the error sequence $\mathbf{e}$, $\mathrm{P}\{\mathbf{v}\}$ is the prior probability of the transmitted information sequence $\mathbf{v}$, and $\Gamma(\mathbf{v})$ is the path metric of $\mathbf{v}$. The pairwise error probability should be averaged over the fading characteristic, which gives

$$\mathrm{P}\{\Gamma(\mathbf{v} \oplus \mathbf{e}) \ge \Gamma(\mathbf{v}) | \mathbf{v}\} = \int \mathrm{P}\{\Gamma(\mathbf{v} \oplus \mathbf{e}) \ge \Gamma(\mathbf{v}) | \mathbf{v}, \mathbf{g}\} p(\mathbf{g}^1) \cdots p(\mathbf{g}^D)\, d\mathbf{g}^1 \cdots d\mathbf{g}^D.$$

We assume perfect channel estimate and the same fading characteristics for all diversity branches in our derivation. Let $g_{k,i} = g_{k,i\,R} + j g_{k,i\,I}$, where the superscript d is ignored because of our assumption, and $\varepsilon_k = \varepsilon_{k\,R} + j\varepsilon_{k\,I}$. For an error event of length N, let the covariance matrix of the random sequence $(g_{1,i\,R}, g_{1,i\,I}, \cdots, g_{N,i\,R}, g_{N,i\,I})$ be $\mathbf{R}_i$. Consider the matrix

$$\mathbf{\Lambda}' = \sum_{i=n}^{L} \mathbf{A}_i^T \mathbf{R}_i \mathbf{A}_i,$$

where

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{i,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{A}_{i,N} \end{bmatrix},$$

and

$$\mathbf{A}_{i,k} = \begin{bmatrix} \varepsilon_{k-i\,R} & \varepsilon_{k-i\,I} \\ -\varepsilon_{k-i\,I} & \varepsilon_{k-i\,R} \end{bmatrix}.$$

Since in general the matrix $\mathbf{\Lambda}'$ may be singular, let, without loss of generality, the first M rows are independent and $\mathbf{\Lambda}'_k = \sum_{m=1}^{M} \alpha_{k,m} \mathbf{\Lambda}'_m$, for $k = M+1, \ldots, 2N$, where $\mathbf{\Lambda}'_k$ is the kth row of $\mathbf{\Lambda}'$. Let $\mathbf{\Lambda}$ be the resulting $M \times M$ submatrix of $\mathbf{\Lambda}'$ by deleting dependent rows and columns. We can show that the pairwise error probability can be Chernoff-bounded by

$$\mathrm{P}\{\Gamma(\mathbf{v} \oplus \mathbf{e}) \ge \Gamma(\mathbf{v}) | \mathbf{v}\} \le \frac{1}{2} \left[\frac{\beta_M}{2^{M/2} |\mathbf{\Lambda}|^{1/2}} \right]^D,$$

where β_M can be computed from the following recursive relation:

Initialization:

$F_{i,j}^{(1)} = \frac{1}{2}(\mathbf{\Lambda}^{-1})_{i,j} + \frac{1}{4N_0}\delta_{ij} + \frac{1}{4N_0}\sum_{k=M+1}^{2N} \alpha_{k,i}\alpha_{k,j}$;

$\beta_1 = \frac{1}{\sqrt{F_{1,1}^{(1)}}}$;

For $(k = 2, 3, \ldots, M)$ {

  For $(i = k, \ldots, M;\ j = k, \ldots, M)$ {

    $F_{i,j}^{(k)} = F_{i,j}^{(k-1)} - \frac{F_{k-1,k}^{(k-1)} F_{k-1,j}^{(k-1)}}{F_{k-1,k-1}^{(k-1)}}$;

  }

  $\beta_k = \beta_{k-1} \frac{1}{\sqrt{F_{k,k}^{(k)}}}$;

}

The recursive formula is simple and can be easily carried out in a computer. The final upper bound on bit error probability is computed from the system's error-state diagram and by a stack algorithm similar to that in [1].

In the following an example is given for an 4-state 8-PSK TCM scheme on a two-tap fading channel. The autocorrelation function of tap coefficients is modeled as

$$\frac{1}{2}\mathrm{E}\{g_{k,i}^* g_{k+m,j}\} = \begin{cases} 0, & i \ne j, \\ \sigma_i^2 J_0(2\pi f_i m T), & i = j. \end{cases}$$

Analytical and simulation results are shown in Figure 1, where four normalized fade rates $f_D T = 0.005, 0.03, 0.05, 0.08$ are considered. From the results shown, our bound is good for all the cases considered especially when diversity reception is available. It accurately predicts the tendency of the performance curves.

References

[1] W. H. Sheen, "Performance analysis of sequence estimation techniques for intersymbol interference channel," Ph. D. dissertation, Georgia Institute of Technology, Atlanta, 1991.

[2] W.-H. Sheen and G. L. Stüber, "MLSE equalization and decoding for multipath-fading channels," *IEEE Trans. Commun.*, vol. COM-39, pp. 1455–1464, Oct. 1991.

[3] W.-H. Sheen and G. L. Stüber, "Performance analysis of trellis codes over ISI channels," *IEEE Trans. Commun.*, to be published.

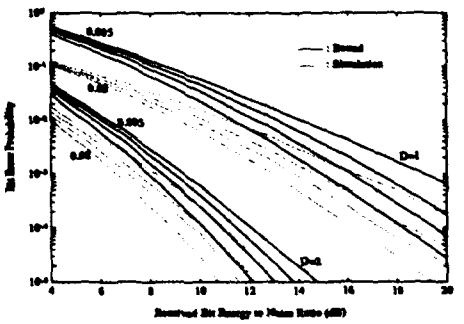


Figure 1: Analytical and simulation results for a 16-state 8-PSK MLSE on multipath Rayleigh fading channels with $f_D T = 0.005, 0.03, 0.05, 0.08$.

This work was supported by the National Science Council of Republic of China under Grant No. NSC-81-0404-E007-015.

ERROR RATE ANALYSIS OF TRELLIS-CODED MODULATION AND OPTIMUM CODE SEARCH FOR IMPULSIVE NOISE CHANNEL

Haruo OGIWARA and Hiroki IRIE
Nagaoka University of Technology
1603-1,Kamitomioka,Nagaoka-shi, 940-21 Japan.
Email ogiwara@voscc.nagaokaut.ac.jp

1. Introduction

In the maximum-likelihood decoding under a non-Gaussian noise, the decoding region is bounded by complex curves instead of a perpendicular bisector corresponding to the Gaussian noise. Therefore, the error rate is not evaluated by the Euclidean distance. The Bhattacharyya distance is adopted since it can evaluate the error performance for a noise with an arbitrary distribution.

Upper bound formulae of a bit error rate and an event error rate are obtained based on the error-weight-profile. Using the bound, optimum code is searched.

2. System Model

Figure 2 shows the system model discussed here. An Information bit stream is fed to a linear convolutional encoder of rate 2/3 and mapped to the 8-AM signal of equal signal-spacing by natural mapping. The received signal, disturbed by a non-Gaussian noise, is fed to a viterbi-decoder for maximum-likelihood decoding. The decoder is assumed to know the probability density of the noise.

3. Upper bound of error rate

The Bhattacharyya distance, BD(A,B), between signal point A and B is given by

$$BD(A,B)=-\ln\left[\int_{\text{whole space}} \{Pn(x-x_A)\ Pn(x-x_B)\}^{1/2}\ dx\right]$$

where Pn(x) is the probability density of the receiving noise. The Union bound of event or bit error rate is estimated based on the error-weight-profile method[1] by using the Bhattacharyya distance instead of the squared Euclidean distance.

4. Optimum code search

To determine the optimum code for an impulsive noise channel, the upper bound of the bit error rate is calculated for each code having an encoder with given shift-register length. The best code is selected as that having the minimum upper bound of the bit error rate.

To lighten the computation burden, a suboptimum search is also attempted. In the suboptimum search, the candidate codes having the maximum free Bhattacharyya distance codes is searched at the first stage. In the next stage, the upper bound of bit error rate is calculated among the candidate codes and the suboptimum code is determined.

5. Results of code search

Figure 2 shows the probability density of the impulsive noise, modeled from an observation in digital subscriber loops. For the noise, the optimum or suboptimum code is searched for among codes having Massey type encoders with a shift-register of up to 4 bits. Figure 3 shows the relation between the upper bound and the simulation result. Figure 4 shows the result of BER. By using the suboptimum code with a 4-bit encoder, a coding gain of 20dB is obtained at the bit error rate 10^{-5}. It is 11dB more than that obtained by Ungerboeck's code. The detailed result is reported[2].

[1] E.Zehavi and J.Wolf,IEEE Trans.IT,IT-33, pp196-202(1987). [2] H.Ogiwara and H.Irie, IEICE Trans.,E75-A,pp1063-1070(1992).

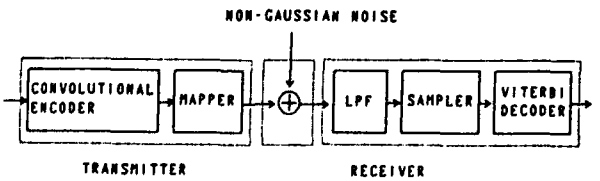


Fig. 1 System model.

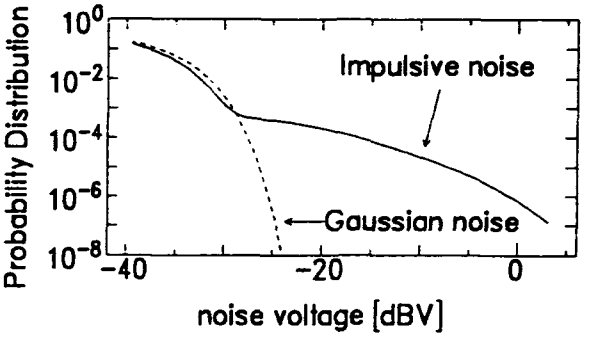


Fig. 2 Probability distribution of an impulsive noise and the Gaussian noise of the same variance.

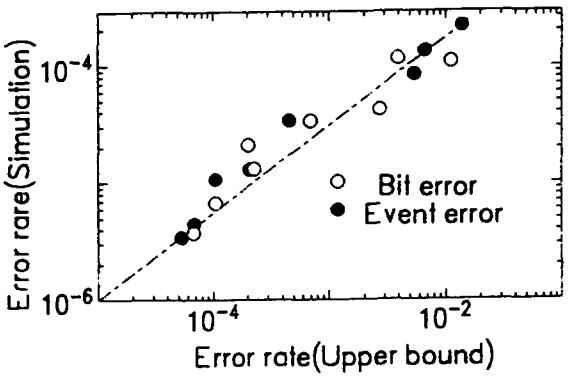


Fig. 3 Relation between the upper bound and simulation results.

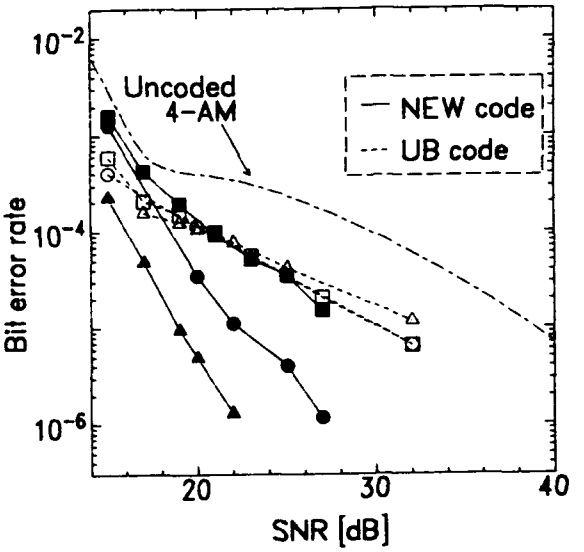


Fig. 4 Bit error rate by simulation of optimum (encoder shift register length $\nu=2,3$) or suboptimum ($\nu=4$) codes and bit error rate without coding. "UB code" is the code found by Ungerboeck. □, ■ : $\nu=2$, ○, ● : $\nu=3$, △, ▲ : $\nu=4$.

TRELLIS CODED MODULATION FOR DIGITAL MICROWAVE RADIO

Tomoko Kodama Matsushima[+] Hirokazu Tanaka[++]

+ Research and Development Center, Toshiba Corporation
1, Komukai Toshiba-cho, Saiwai-ku, Kawasaki 210 Japan

++ Communication Systems and Technology Lab., Toshiba Corporation
3-1-1, Asahigaoka, Hino-city, Tokyo 191 Japan

In this paper, we propose a modified symbol-rate-increased TCM for digital microwave radio. The original symbol-rate-increased TCM accomplishes coding redundancy through bandwidth expansion, instead of through signal point expansion, in order to obtain a greater coding gain than the Ungerboeck-type TCM. A drawback of this scheme is that it requires the bandwidth to be expanded by a fixed factor m/(m-1) for a 2^mQAM system.

The proposed scheme permits the setting of the bandwidth expansion ratio to an arbitrary value smaller than m/(m-1). The simulation results clarified that the proposed scheme can set lower bandwidth expansion ratio than the symbol-rate-increased TCM with only slightly reduced coding gain.

Error Control Schemes for DMR Systems

Quadrature Amplitude Modulation(QAM) is widely used in Digital Microwave Radio (DMR) systems. Error control is indispensable for realizing highly reliable multilevel QAM systems, particularly when the number of constellation points is large. Two principal error control schemes for DMR systems are block coding and Trellis Coded Modulation (TCM). Block coding, such as BCH code or Reed-Solomon code, can be easily adopted to DMR[1], because it allows the bandwidth expansion ratio to be set to an arbitrary value. Though block coding is utilized in many commercial systems, the coding gain is rather small(2-3 dB at BER=10^{-5}).

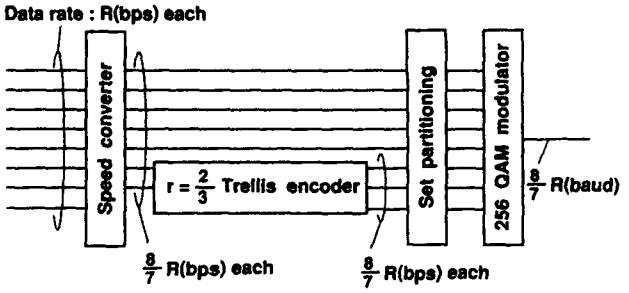


Fig. 1 Symbol-rate-increased TC-256 QAM

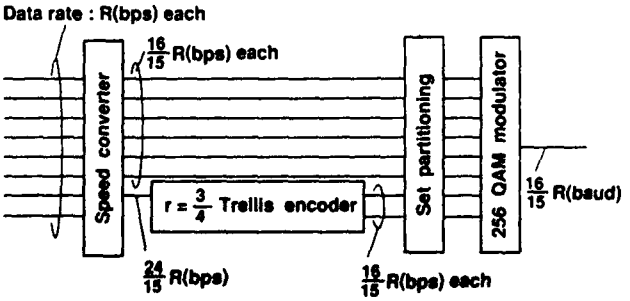


Fig. 2 Proposed 256 QAM scheme

Saito proposed a new scheme, named symbol-rate-increased TCM, for DMR systems[2]. It accomplishes coding redundancy through bandwidth expansion, instead of through signal-set expansion, and acheives a remarkable coding gain (greater than 5 dB at BER=10^{-5}). A drawback of this scheme is that it requires the bandwidth to be expanded by a fixed factor m/(m-1) for 2^mQAM. For example, the bandwidth expansion ratio for the symbol-rate-increased TC-256QAM (Figure 1) is 8/7.

Modified Symbol-Rate-Increased TCM Scheme

The scheme proposed in this paper permits the setting of the bandwidth expansion ratio to an arbitrary value smaller than m/(m-1). In the scheme, an m-bit parallel input data sequence is converted into m_1-bit and m_2-bit parallel sequences with different data rates, and then the m_1-bit sequence is encoded by a trellis encoder whose coding rate is r. Then the overall bandwidth expansion ratio is $m/\{(m-m_2)r+m_2\}$. It is smaller than m/(m-1), if r is greater than $(m-m_2-1)/(m-m_2)$. As an example, we have designed a scheme for 256QAM whose bandwidth expansion ratio is 16/15 (Figure 2).

Bit Error Rate Performance

Figure 3 shows the bit error rate performance for three different schemes: the symbol-rate-increased TC-256QAM, the proposed 256QAM scheme with a bandwidth expansion ratio of 16/15, and a practical 256QAM scheme with a (255,239) BCH code. The number of the encoder states is 32. The coding gain for the symbol-rate-increased TCM and that for the proposed scheme are 5.1 dB and 4.3 dB, respectively, at a BER of 10^{-5}. Though the figure for the proposed scheme is slightly smaller than that obtainable with the original symbol-rate-increased TCM, it shows that the proposed scheme can attain a remarkable improvement over a BCH coded 256QAM with the same bandwidth expansion ratio.

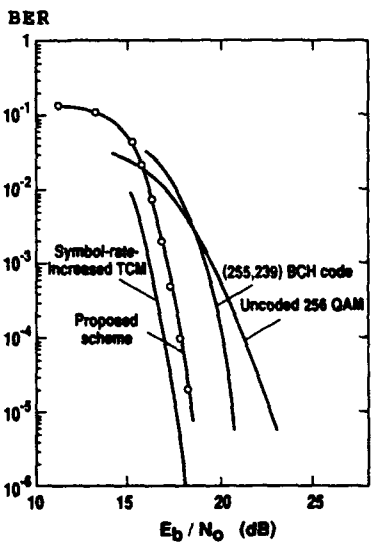


Fig. 3 BR error rate performance

References

[1]T.Kodama et al., "Error control schemes for differentially encoded multilevel QAM transmission systems",Electron. and commun. in Japan,Part 3,Vol. 74, No.1, 1991.

[2]Y.Saito,"Trellis coded modulation for multi-state QAM", Trans. of the IEICE, Vol. E73, No.10,1990.

OPTIMAL MULTI-h PHASE CODES FOR PARTIAL RESPONSE CONTINUOUS PHASE MODULATION

Rongqiang Mao and John P. Fonseka

School of Engineering and Computer Science
The University of Texas at Dallas, EC 33
Richardson, Texas 75080–0688

ABSTRACT

Multi-h partial response CPM signals are analyzed at preselected number of states in contrast to the standard method of analyzing them at preselected number of modulation indices. Optimal multi-h phase codes which produce the highest minimum Euclidean distance are presented according to the number of states. Three orders of multi-h signaling; 2-h, 3-h, and 4-h signaling are considered in the optimal code search.

1. INTRODUCTION

In the literature multi-h phase codes have been extensively considered with continuous phase modulation (CPM) with regard to their performance, spectral properties, detection etc [1-4]. It is known that the performance of CPM signaling can be improved by combining with multi-h phase codes while maintaining the properties inherent to the phase continuity of the signals [1-4].

A k-h CPM signal changes its modulation index cyclically at the end of every interval over k values ($h_0, h_1, \ldots, h_{k-1}$). Binary k-h partial response signals considered in this study generally take the form

$$x(t) = \sqrt{\frac{2E_b}{T}} \cos\left[\omega_c t + 2\pi a_n h_n \int_{nT}^{t} g(\alpha - nT)d\alpha + \phi_0\right]; \quad nT \le t \le (n+1)T. \tag{1}$$

The error rate performance of any wideband CPM signaling system is usually expressed in terms of the minimum Euclidean distance of the signals. In fact, the asymptotic error probability variation at high signal to noise ratio is approximately given by [1],

$$P_e \approx Q\left(\sqrt{d_{min}^2 E_b/N_0}\right). \tag{2}$$

It is seen from (2) that the performance of a CPM system can be improved by increasing the minimum Euclidean distance. The minimum Euclidean distance can generally be increased by increasing the constraint length of the signals. The total number of states N' is the product of the number of phase states N and the number of symbol states.

The complexity of a multi-h signaling scheme mainly depends on the number of modulation indices (or the order of signaling) k, the number of states N', and the receiver path memory length N_R [1, 2]. Among them the number of states can be considered as the most significant factor as it is necessary to compute $2N'$ number of metrics at the end of every interval for trellis decoding [1]. Considering the factors which determine the complexity, it is more appropriate to analyze multi-h signals at preselected values of N' in situations where the allowed complexity is limited. Since partial response signals generally have better spectral properties than full response signals, the results presented here are more beneficial than those in [2] especially when designing bandlimited systems.

2. EVALUATION OF OPTIMAL MULTI-h CODES

The evaluation of the optimal multi-h codes for both 2REC and 2RC signals at any given value of N' (or N) consists of the evaluation of the optimal number of modulation indices, k and determination of the set of k optimal modulation indices ($h_0, h_1, \ldots, h_{k-1}$), to maximize the minimum Euclidean distance. For any selected value of k, the average modulation index which is defined as the mean value of set of modulation indices can take only discrete values which are determined by the selection of the set of integers of the numerator of modulation indices.

Since the maximization of the constraint length does not necessarily maximize the minimum Euclidean distance, the direct maximization of the minimum Euclidean distance has to be carried out over all possible values of k to evaluate the optimal codes. In order to reduce the complexity of calculations and of the resulting signaling schemes, only 2-h, 3-h and 4-h signals with k=2, k=3 and k=4 respectively, are considered in the optimal code search. It is important to note that even though the set of possible $h_{avg.}$ values on the range $0 < h_{avg.} < 1$ differs from one value to the other, the best value of k and the corresponding $h_{avg.}$ value along with the optimal combination of modulation indices which produces the highest minimum Euclidean distance can be determined in range $0 < h_{avg.} < 1$.

At any given value of k, the best combination of modulation indices at each possible $h_{avg.}$ value is numerically found by searching over all possible combinations of modulation indices. Since the primary objective is to maximize the minimum Euclidean distance, all special combinations of modulation indices are also considered. These special combinations include the ones with any number of equal modulation indices including the constant h signals. Further, in all of the minimum Euclidean distance calculations the minimization is carried out over all cyclic shifts of the modulation index pattern in order take into account of the paths that originate at the beginning of all signaling intervals.

References

[1] J.B. Anderson, T. Aulin and C.E. Sundberg, Digital Phase Modulation. New York: Plenum Press, 1986.

[2] J.P. Fonseka, "Optimal multi-h phase codes for full response continuous phase signaling," in Proc. IEEE Inter. Telecom. Symposium, Rio de Janerio, Brazil, pp. 22.1.1–22.1.5, Nov. 1990.

[3] W. Holubowicz, "Optimum parameter combinations for multi-h phase codes," IEEE Trans. on Commun., Vol. COM-38, pp. 1929–1931, Nov. 1990.

[4] J.P. Fonseka and R. Mao, "Multi-h phase codes for continuous phase modulation," Electron. Lett. vol. 28, No. 16, pp. 1495–1497, July 1992.

A Demonstration of a Robust Occam-Based Learner

Timothy D. Ross and Michael J. Noviskey
Wright Lab, WL/AART, WPAFB, OH 45433-6543
Mark L. Axtell
Veda Inc., 5200 Springfield Pk, Dayton, OH 45431
Michael A. Breen
Tennessee Tech., Math Dept., Box 5054, Cookeville, TN 38505

Machine learning is often modeled as the process of extrapolating samples of a function. This extrapolation requires both samples and "inductive bias." Bias towards low complexity, as in Occam's Razor, is particularly important. Kolmogorov complexity was developed to formalize this process [5]. Important theoretical results have also been developed using more abstract measures of complexity [4]. Therefore, there is a strong theoretical basis for Occam-based learning. Kolmogorov complexity is a general measure, which would allow learning of many different kinds of functions (i.e. robust learning); however, it has been proven that its exact computation is not tractable. There have been some tractable measures of complexity used in actual implementations of Occam-based learning [3], such as the Abductory Inference Mechanism (AIM). However, these measures of complexity are relatively narrow, which implies non-robust learning. The challenge is to develop *robust and* tractable measures of complexity.

One approach to this challenge is called pattern theory [6], where we think of robust complexity determination as the problem of finding a pattern. Pattern theory uses Decomposed Function Cardinality (DFC), proposed by Y. S. Abu-Mostafa as a general measure of complexity [1, p.128]. We demonstrate that this measure of complexity allows robust learning, yet is sufficiently tractable to support the learning of *non*-trivial functions. We develop support for the generality of the measure both theoretically and experimentally. Generality is supported theoretically by proving its relationship to the conventional measures of circuit complexity, time complexity and program length. Generality is supported experimentally by using a decomposition program (referred to as AFD) derived from the work of R. L. Ashenhurst [2] and others.

The experimental work includes the measurement of the DFC of a large variety of functions, determining the correlation of DFC with more specialized measures within their domain of application, and machine learning experiments. The DFC of over 800 non-randomly generated functions was measured, including many *kinds* of functions (numeric, symbolic, chaotic, string-based, graph-based, images and files). Roughly 98 percent of the non-randomly generated functions had low DFC (versus less than 1 percent for random functions). The 2 percent that did not decompose were the more complex of the non-randomly generated functions rather than some class of low complexity that AFD could not deal with. It is important to note that when AFD says the DFC is low, which it did some 800 times, it also provides an algorithm (or a description of the pattern found). AFD found *the* classical algorithms for a number of functions.

The correlation coefficient between DFC and a ranking of the complexity of images by people was 0.8. The correlation between DFC and the compression factor of two commercial data compression programs was about 0.9. The correlation between DFC and the Lyapunov exponent for logistic functions was 0.9. These high correlations show that the single measure, DFC, reflects the essential structure in each, very different, situation. A fourth correlation experiment found no correlation between people's ability to recognize concepts and the DFC of those concepts. This lack of correlation may be a result of the narrow range of DFC's involved (5% of its full range).

In learning experiments, AFD did as well as a back-propogation trained Neural Network (NN) on problems well-suited to NN's. However, on other problems such as parity, AFD learned a 256 point function from 50 samples whereas the NN required all 256 points. The findings were similar for the AIM program. In both cases, the extrapolations of the NN and AIM were not robust while AFD consistently learned functions of low complexity with few samples.

The experiments to date have been limited to small (less than 10 variables) binary functions. The results on these small, but non-trivial, functions have consistently pointed to a promising ability to find many different kinds of patterns. Therefore, we believe these results are the first demonstration of robust Occam-based learning and help join an important body of theoretical results with practical machine learning.

References

[1] Yaser S. Abu-Mostafa, editor. *Complexity in Information Theory.* Springer-Verlag, New York, 1988.

[2] Robert L. Ashenhurst. The decomposition of switching functions. In *Proceedings of the International Symposium on the Theory of Switching*, April 1957.

[3] A. R. Barron and R. L. Barron. Statistical learning networks: a unifying view. In *1988 Symposium on the Interface: Statistics and Computing Science*, page 12, 1988.

[4] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Occam's razor. *Information Processing Letters*, 377–380, October 1987.

[5] Ming Li and Paul M. B. Vitányi. Two decades of applied Kolmogorov complexity. In *Proceedings Structure in Complexity Theory*, pages 80–101, IEEE, 1988.

[6] Timothy D. Ross, Michael J. Noviskey, Timothy N. Taylor, and David A. Gadd. *Pattern Theory: An Engineering Paradigm for Algorithm Design.* Final Technical Report WL-TR-91-1060, Wright Laboratory, USAF, WL/AART, WPAFB, OH 45433-6543, August 1991.

Nonparametric Regression–Based Method for Neural Network Training*

Terrence L. Fine and Jen–Lun Yuan
School of Electrical Engineering
ETC 388
Cornell University
Ithaca, NY 14853

Summary for 1993 Inter. Symp. on Information Theory

Artificial neural network training algorithms, based upon gradient search minimizations of cost/loss functions when the training set contains many high–dimensional inputs, are time–consuming and can only address the issue of an appropriate network architecture (network topology) either through repeated training of different networks or the addition of penalty functions so as to control the problem of 'overfitting' that results from 'overtraining', a counterpart to the classical statistical estimation issue of the balance between bias and variance (e.g., Geman, et al. [1992]). We propose combining ideas drawn from the nonparametric regression approaches of projection pursuit (PP) (Huber [1985]), the recent idea of sliced inverse regression (SIR) (Li [1991]), backfitting (Hastie and Tibshirani [1990]), and the design of scalar smoothers via Gaussian kernel smoothing (Hastie and Tibshirani [1990]), to decompose the neural network design/training process into one involving gradient methods only at the final stages where we need only fit scalar–valued functions of scalar inputs.

From PP we take the class of regression functions

$$y = \sum_{i=1}^{m} g_i(\underline{w}_i' \underline{x}),$$

where the regressor vector (input) is $\underline{x}$, the response (output) variable is y, and we need to select the 'ridge functions' $\{g_i\}$ and the projection directions $\{\underline{w}_i\}$ as well as the number of terms m. This model can be embedded in a neural network where we may use several nodes to approximate each of the ridge functions. Our goal is to improve the efficiency of fitting this model to training data by introducing a fairly direct method for extracting the projection directions. We need then only rely upon time–consuming iterative methods to approximate the scalar–input ridge functions by scalar–input network node functions.

Our experience with backpropagation (BP) methods applied to forecasting electric load time series (Yuan and Fine [1992]) suggests that often the resulting neural network has several sigmoidal nodes operating about the origin where they are essentially linear; e.g., for the standard logistic node $\sigma(z) = [1+e^{-z}]^{-1}$ this could be the region $|z| \leq 1$. Hence, we select initially g_1 to be linear. SIR provides a direct calculation method for calculating the weight vector $\underline{w}$ when the 'true' model is

$$y = g(\underline{w}'\underline{x} + \tau) + \epsilon,$$

for noise ϵ independent of input $\underline{x}$ and $\underline{x}$ elliptically symmetrically distributed. While this is not likely to be our situation, we can use the SIR algorithm to provide reasonable projection directions for the PP regression function and improve them through a process of iterative backfitting. Each ridge function g_i is trained to model the residual $\hat{y}^i$ resulting from the difference of the actual response y and the output of the other $m-1$ ridge functions. The number m is selected so that the estimated residual error is below a pre-assigned threshold. Given that we are selecting g_i, we first use SIR to calculate the direction $\underline{w}_i$ and, keeping the direction fixed, we use Gaussian kernel functions to smooth the resulting scatter plot $\{(\underline{x}_j, \hat{y}_j^i)\}$ between the scalar response residual $\hat{y}_j^i$ to the jth input vector. At the conclusion of the iterative backfitting process, in which we cycle repeatedly through the m terms in the regression equation, we then use backpropagation to approximate the smoothed terms in the PP equation by small neural network subnets that now have scalar inputs and outputs.

Concerned by the hypothesis of elliptically symmetrically distributed $\underline{x}$ required by SIR, we have developed a related method that uses a new projection index motivated by the continuity of the unknown regression function g. As in SIR, we slice the output values into H intervals and estimate projection directions based on the groups $\{I_h, h = 1, H\}$ of input values that share output values in the same interval. For example, we find a single projection direction

$$\underline{w} = \text{argmin}_{\{\underline{\beta}:||\underline{\beta}||=1\}} \sum_{h=1}^{H} \sum_{\{\underline{x}_i, \underline{x}_j \in I_h\}} \underline{\beta}^T (\underline{x}_i - \underline{x}_j)(\underline{x}_i - \underline{x}_j)^T \underline{\beta}.$$

We have applied these nonparametric regression–based training processes to the construction of a neural network to forecast daily extremes (daily minimum, evening peak, morning peak) of demand for electric power, using data supplied by a midwestern utility, and will compare our results with those we have achieved previously (Yuan and Fine [1992]) using backpropagation–based methods.

References

Geman, S., E. Bienenstock, R. Doursat [1992], Neural networks and the bias/variance dilemma, *Neural Computation*, **4**, 1-58.

Hastie, T., R. Tibshirani [1990], *Generalized Additive Models*, Chapman and Hall.

Huber, P. [1985], Projection pursuit, *Annals of Statistics*, **13**, 435-475.

Li, K.–C. [1991], Sliced inverse regression for dimension reduction, *Jour. Amer. Statistical Assn.*, **86**, 316–342.

Yuan, J.-L., T. Fine [1992], Forecasting demand for electric power using autoregressive neural networks, *Proc. Conf. on Information Sciences and Systems*, Princeton, NJ.

* **Prepared with partial support from NSF Grant No. ECS-9017493**

A Measure of Relative Entropy between Individual Sequences with Application to Universal Classification

Jacob Ziv and *Neri Merhav*

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL

Abstract

A new notion of empirical informational divergence between two individual sequences is introduced. If the two sequences are independent realizations of two stationary Markov processes, the empirical relative entropy converges to the true divergence almost surely. This new empirical divergence is based on a version of the Lempel-Ziv data compression algorithm.

A simple universal classification algorithm for individual sequences into a finite number of classes which is based on the empirical divergence, is introduced. It discriminates between the classes whenever they are distinguishable by some finite-memory classifier, for almost every given training sets and almost any test sequence from these classes. It is universal in the sense of being independent of the unknown sources.

Summary

Suppose one observes a sequence $\mathbf{x} = (x_1, \ldots, x_n)$ emitted from an unknown l-th order stationary Markov process $p(\cdot)$ over a finite-alphabet $\mathbb{A}$ with $|\mathbb{A}|=A$ letters, and wishes to estimate the nth order entropy, or equivalently $-n^{-1}\log p(\cdot)$. While the straightforward approach of calculating the l-th order conditional empirical distribution is computationally prohibitedly complex for large l and is impossible if l is unknown, it has been shown in [1],[2] that the Lempel -Ziv (LZ) codeword length for $\mathbf{x}$ divided by the length n, is a computationally efficient, reliable estimate of the entropy, and hence also of $-n^{-1}\log p(\cdot)$.

More precisely, let $p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|s_{i-1})$ where $s_i = (x_{i-l+1}, x_{i-l+2}, \ldots, x_i)$ for $i \geq \ell$ and where $s_i = (s_0, x_1, x_2, \ldots, x_i)$ for $i < \ell$, s_0 being the initial state; $s_i \in \mathbb{A}^\ell$, and $\mathbb{A}^\ell$ is the set of all length ℓ vectors with components in $\mathbb{A}$.

Let $c(\mathbf{x})$ denote the number of phrases in $\mathbf{x}$ resulting from the incremental parsing of $\mathbf{x}$ [1], i.e., sequential parsing of $\mathbf{x}$ into distinct phrases such that each phrase is the shortest string which is not a previously parsed phrase. Then, the LZ codelength for $\mathbf{x}$ can be approximated by $c(\mathbf{x})\log c(\mathbf{x})$ and $\lim_{n\to\infty}\ n^{-1}\left[-\log p(\mathbf{x}) - c(\mathbf{x})\log c(\mathbf{x})\right] = 0$ almost surely. In fact, this property still holds as long as $p(\cdot)$ is more generally a stationary ergodic process.

Here we generalize this result to the case where there are two stationary ℓth order Markov sources, $p(\cdot)$ and $q(\cdot)$. Let $\mathbf{x}$ and $\mathbf{z}$ be realizations of $p(\cdot)$ and $q(\cdot)$, respectively. Given $\mathbf{x}$ and $\mathbf{z}$, we would like to estimate reliably $-n^{-1}\log p(\mathbf{z})$ and similarly, $-n^{-1}\log q(\mathbf{x})$. In particular, we seek an easily calculable function of $\mathbf{x}$ and $\mathbf{z}$, independent of ℓ, which discriminates between two unknown Markov sources $p(\cdot)$ and $q(\cdot)$. To this end, recall that the divergence $D(q\|p)$, defined as

$$D(q\|p) = \limsup_{n\to\infty} \frac{1}{n} \sum_{\mathbf{A}^n} {}_{(}\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} \,. \tag{1}$$

is intuitively interpreted as a measure of distance between $p(\cdot)$ and $q(\cdot)$ [2]. In the Markovian case considerded here, we have

$$D(q\|p) = \sum_{a\in\mathbb{A};\ s\in\mathbb{A}^\ell} q(a,s)\log\frac{q(a|s)}{p(a|s)} \,, \tag{2}$$

where $p(a|s)$ and $q(a|s)$ are the conditional probabilities of a letter $a \in \mathbb{A}$ given a state $s \in \mathbb{A}^\ell$ under $p(\cdot)$ and $q(\cdot)$, respectively. Let $q_{\mathbf{z}}(a_1, \ldots, a_{\ell+1}) = q_{\mathbf{z}}(a^{\ell+1})$ be the relative frequency of an $(\ell+1)$-length string $a^{\ell+1} \in \mathbb{A}^{\ell+1}$. Then, it is well known that

$$-\log p(\mathbf{z}) = nH_{\mathbf{z}} + nD(q_{\mathbf{z}}\|p) \tag{3}$$

where

$$H_{\mathbf{z}} = -\sum_{a\in\mathbb{A}}\sum_{s\in\mathbb{A}^\ell} q_{\mathbf{z}}(a,s)\log q_{\mathbf{z}}(a|s). \tag{4}$$

Analogously to the single source case, where $-n^{-1}\log q(\mathbf{z})$ is efficiently estimated by $n^{-1}c(\mathbf{z})\log c(\mathbf{z})$, we introduce an empirical quantity $Q(\mathbf{z}\|\mathbf{x})$ which will be shown to have the property,

$$\lim_{n\to\infty}\left[-\frac{1}{n}\log p(\mathbf{z}) - Q(\mathbf{z}\|\mathbf{x})\right] = 0$$

almost surely w.r.t the product measure $p \times q$, for every finite ℓ. Following (3), the function $Q(\mathbf{z}\|\mathbf{x})$ can be decomposed into two terms, the first of which is an estimate of the empirical entropy associated with $\mathbf{z}$, i.e., $n^{-1}c(\mathbf{z})\log c(\mathbf{z})$, and the second, denoted by $\Delta(\mathbf{z}\|\mathbf{x})$ is an estimate of the divergence between $q_{\mathbf{z}}(\cdot)$ and $p(\cdot)$ with the property, $\lim_{n\to\infty}[\Delta(\mathbf{z}\|\mathbf{x}) - D(q_{\mathbf{z}}\|p)] = 0$ almost surely with respect to the product measure $p \times q$, for every finite ℓ. In parallel to the fact that the entropy is estimated by *self* LZ incremental parsing of $\mathbf{z}$, here intuition suggests that $\Delta(\mathbf{z}\|\mathbf{x})$, which is an estimate of the *cross entropy* $D(q_{\mathbf{z}}\|p)$, will be associated with *cross parsing* of $\mathbf{z}$ with respect to $\mathbf{x}$.

Specifically, the cross parsing procedure of $\mathbf{z}$ w.r.t $\mathbf{x}$ works as follows. First, find the longest prefix of $\mathbf{z}$ that appears as a string in $\mathbf{x}$, i.e., the largest integer m such that $(z_1, z_2, \ldots, z_m) = (x_i, \ldots, x_{i+m-1})$ for some i. The string $(z_1, z_2, \ldots, z_m)$ is defined as the first phrase of $\mathbf{z}$ with respect to $\mathbf{x}$. If $m = 0$ (i.e., z_1 does not appear in $\mathbf{x}$), the first phrase of $\mathbf{z}$ with respect to $\mathbf{x}$ is z_1. Thus, the case $m = 0$ is treated as though $m = 1$. Next, start from z_{m+1} and find, in a similar manner, the longest prefix of $z_{m+1}, z_{m+2}, \ldots, z_n$, which appears in $\mathbf{x}$, and so on. The procedure is terminated once the entire vector $\mathbf{z}$ has been parsed with respect to $\mathbf{x}$. Let $c(\mathbf{z}|\mathbf{x})$ denote the number of phrases in $\mathbf{z}$ with respect to $\mathbf{x}$.

Intuitively, $\Delta(\mathbf{z}\|\mathbf{x})$ may serve as a reasonable discrimination function for universal classification of individual sequences. Indeed, in contrast to the probabilistic framework in which the classification problem is normally posed, we show that a classifier based on the comparison of $\Delta(\mathbf{z}\|\mathbf{x})$ to a threshold, results in an asymptotically optimal performance for almost every individual tested data sequence among all finite-memory classifiers that are trained by given training sequences from each class, have a rejection option, and that assign to each class a small as possible set of vectors so as to make the sources distinguishable. We assume that any competing finite memory classifier is *consistent* in the sense that if a test sequence, to be classified, appears in the training set it will be classified correctly. It should be pointed out that while the order of the optimal competing finite memory classifier is normally unknown, the discrimination procedure based on the above described cross-parsing, is independent of ℓ and computationally efficient.

References

[1] J. Ziv and A. Lempel, "Compression of Individual Sequences via Variable-Rate Coding," *IEEE Trans. Inf. Theory*, Vol. IT-24, No. 5, pp. 530-536, September 1978.

[2] T. M. Cover and J. A. Thomas *Elements of Information Theory*, Wiley, New York, 1991.

On Radial Basis Function Net and Kernel Regression: Approximation Ability, Convergence Rate and Receptive Field Size

Lei Xu[1,2], Adam Krzyżak[3] and Allan Yuille[2]
1. Dept. of Mathematics, Peking University, P.R.China
2. Harvard Robotics Laboratory, Harvard University, USA
3. Dept. of Computer Science, Concordia University, Canada H3G 1M8

Abstract

Radial Basis Function (RBF) network have been studied intensively ([1], [9], [8], [3], [2],[12], [13]). Besides its applications several theoretical results have been obtained. E.g., (1) RBF net can be naturally derived from *regularization theory* ([9]), (2) RBF net has universal approximation ability [4], (3) RBF net has also best approximation ability ([3],[5]).

In this paper, connections between RBF network and *Kernel Regression Estimator (KRE)* are built up. Recent theoretical results about KRE are used as tools to obtain the theoretical sults on RBF net in several aspects. First, the statistical consistencies of RBF nets are proved in various situations, which extend the current results on the approximation ability (e.g. universal approximation, ..., etc) of RBF net from deterministic case to more practical stochastic case. Second, the convergence rates of RBF net are provided in different sitautions, which is more useful than merely convergence of network approximation in the case of infinite number of hidden units. For example, if the mapping function to be learned is bounded and of α orders of smooth, the L_2 convengence rate of RBF net made of basis functions with a compact support is up-bounded by $O(n^{-\frac{2\alpha}{2\alpha+d}})$, which also means that for a given error bound ε_0^2, the number of hidden units is about the order of $O(|\varepsilon_0|^{-\frac{d}{2\alpha+d}})$. This gives us some quantative insights on the designing of RBF net. Third, the problem of selecting the appropriate size of the receptive field of radial basis function is investigated and how the selection of size is influenced by a number of factors is elaborated. These studies are new in the literature and quite useful for the further theoretical analysis of *RBF* as well as for guiding the design of RBF net in practice.

RBF net and KRE

We consider the normalized version of RBF net ([8])

$$f_n(x) = \frac{\sum_{i=1}^n w_i \phi([x-c_i]^t \Sigma^{-1}[x-c_i])}{\sum_{i=1}^n \phi([x-c_i]^t \Sigma^{-1}[x-c_i])} \quad (1)$$

where $\phi(r^2)$ is some prespecified basis function satisfying some mild condition. The most common one is Gaussian function $\phi(r^2) = e^{-r^2}$, but a number of alternatives can also be used([9]). c_i is called center vector which locates $\phi(r^2)$ centering around c_i. $w_i \in R^m$ is a weight vector corresponding to the center vector c_i. Σ is a $d \times d$ positive matrix which is usually chosen as $\Sigma = \sigma^2 I$ with σ called the size of receptive field of the basis function.

For a given fixed $\phi(r^2)$, in eq.(1) there are three sets of the parameters: (1) $w_i, i = 1, \cdots, n$, which are the weight vectors of the output layer of a RBF net, (2) the center vectors $c_i, i = 1, \cdots, n$ and (3) the size σ. The last two sets constitute the weights of the hidden layer of a RBF net. Theoretically, all the parameters can be determined based on a given sample set $(X_i, Y_i), i = 1, \cdots, N$ by minimizing the following total approximating error

$$\varepsilon^q_{RBF}(Y, f_n) = \frac{1}{N}\sum_{i=1}^N |Y_i - f_n(X_i)|^q \quad (2)$$

where $0 < q \leq \infty$ and the usual case is $q = 2$.

However, the minimization with respect to all the parameters simultaneously is a hard problem. It is usually assumed that σ and $c_i, i = 1, \cdots n$ are determined from the samples $\{X_i, i = 1, \cdots, N\}$ ([10], [8], [3], [2]). In this case, the minimization of $\varepsilon^q_{RBF}(Y, f_n)$ can be simplified considerably since now it is performed with respect to $w_i, i = 1, \cdots, n$. A special case is that $q = 2$ in which the solution is given by $W = YK^T(KK^T)^{-1}$ with $W = [w_1, \cdots w_n]$, $Y = [Y_1, \cdots, Y_N]$ and $K = [k_{ij}]_{n\times N}$, $k_{ij} = \phi_{ij}/\sum_{i=1}^n \phi_{ij}$, $\phi_{ij} = \phi([X_j - c_i]^t \Sigma^{-1}[X_j - c_i])$.

A special way of determining $c_i, i = 1, \cdots n$ is quite simple: a subset $\{X_i, i = 1, \cdots n\}$ is randomly selected among $\{X_i, i = 1, \cdots N\}$ and every selected sample is directly used as a center vector, i.e., $c_i = X_i, i = 1, \cdots, n$ and

$$f_n(x) = \frac{\sum_{i=1}^n w_i \phi([x-X_i]^t \Sigma^{-1}[x-X_i])}{\sum_{i=1}^n \phi([x-X_i]^t \Sigma^{-1}[x-X_i])} \quad (3)$$

In sequel, we call the RBF nets obtained by the minimization of all the parameters the idealistic type nets, and we call RBF nets given by eq.(3) Type-I nets. Let $(X, Y), (X_1, Y_1), \cdots, (X_n, Y_n)$ be independent identically distributed $R^d \times R^m$-valued random vectors. Let $R(x) = E\{Y|X = x\}$ be the regression function and μ be the probability measure of X. The kernel regression estimator of $R(x)$ is defined as follows:

$$f_n(x) = \frac{\sum_{i=1}^n K(\frac{x-X_i}{h})Y_i}{\sum_{i=1}^n K(\frac{x-X_i}{h})} \quad (4)$$

where, h is *smoothing parameter* and $K \geq 0$ is a μ integrable kernel on R^d. Estimator eq.(4) is studied in [6, 7]. The probabilistic neural network proposed by [11] is one type of direct extensions of Parzen Window estimator.

Connections between RBF net and KRE

Let $K(r^2) = \phi(r^2)$, $\Sigma = \sigma^2 I$, $h^2 = \sigma^2$, and $w_i = Y_i, i = 1, \cdots, n$, we see that eq.(4) is identical to eq.(3). That is, a spherically symmetrical kernel $K(r^2)$ is a type of radial basis function, the smoothing parameter h represents the size σ of the basis function's receptive field, and Y_i acts as an approximate solution of w_i. Thus, we can consider the KRE eg.(4) as a particular case of RBF net eq.(3). Assumption $\Sigma = \sigma^2 I$ is in fact commonly used in the existing studies on RBF nets ([1],[9], [8], [2]).

Furthermore, in parallel to eq.(2) we denote the total approximating error of KRE eq.(4) by

$$\varepsilon^q_{KRE}(Y, f_n) = \frac{1}{N}\sum_{i=1}^N |Y_i - f_n(X_i)|^q \quad (5)$$

and for eq.(2), we let $\varepsilon^q_{RBF-opt}(Y, f_n)$ denote the ε^q_{RBF} obtained by minimizing $w_i, c_i, i = 1, \cdots, n$ and Σ simultaneously, i.e., $\varepsilon^q_{RBF-opt}$ is the minimal error obtainable by a RBF net of the idealistic type. Let $\varepsilon^q_{RBF-Type-I}(Y, f_n)$ denote ε^q_{RBF} for a RBF net of Type-I defined by eq.(3).

Lemma *Let $K(r^2) = \phi(r^2)$, we have: (A) $\varepsilon^q_{RBF-opt}(Y, f_n) \leq \varepsilon^q_{KRE}(Y, f_n)$; $E\varepsilon^q_{RBF-opt}(Y, f_n) \leq E\varepsilon^q_{KRE}(Y, f_n)$; (B) $\varepsilon^q_{RBF-opt}(Y, f_n) \leq \varepsilon^q_{RBF-Type-I}(Y, f_n)$ $\varepsilon^q_{KRE}(Y, f_n)$; $E\varepsilon^q_{RBF-opt}(Y, f_n) \leq E\varepsilon^q_{RBF-Type-I}(Y, f_n) \leq E\varepsilon^q_{KRE}(Y, f_n)$, under the same receptive field specified by $\Sigma = h^2 I$; (C) $\varepsilon^q_{RBF-opt}(Y, f_n) \leq \varepsilon^q_{RBF-Type-II}(Y, f_n)$; $E\varepsilon^q_{RBF-opt}(Y, f_n) \leq E\varepsilon^q_{RBF-Type-II}(Y, f_n)$.*

references

[1] D.S.Broomhead and D.Lowe, Multivariable functional interpolation and adaptive networks, *Complex Systems 2*, pp321-323, 1988.

[2] S.Chen, C.F.N.Cowan and P.M.Grant, Orthogonal least squares learning algorithm for Radial basis function networks, *IEEE Trans. on Neural Networks 2*, 1991, pp302-309.

[3] F.Girosi and T.Poggio, Networks and the best approximation property, *M.I.T. AI Memo. No.1164*, MIT, 1989.

[4] E.J.Hartman, J.D.Keeler and J.M.Kowalski, Layered neural networks with Gaussian hidden units s universal approximations, *Neural Computation 2*, 1990, 210-215.

[5] K.Hornik, Approximation capabilities of multilayer feedforward networks, *NN 4*, 1991, 251-257.

[6] A. Krzyżak, The rates of convergence of kernel regression estimates and classification rules, *IEEE Trans. on Information Theory, 32*, pp668-679, 1986.

[7] A. Krzyżak, On exponential bounds on the Bayes risk of the kernel classification rule, *IEEE Trans. on Information Theory, 37*, pp490-498, 1991.

[8] J.Moody and J.Darken, Fast learning in networks of locally-tuned processing units, *Neural Computation 1* 1989, pp281-294.

[9] T.Poggio and F.Girosi, A Theory of networks for approximation and learning, *M.I.T. AI Memo. No.1140*, MIT, 1989.

[10] M.J.D.Powell, Radial basis functions for multivariable interpolation: a review, eds, J.C.Mason and M.G.Cox, *Algorithms for Approximation*, Clarendon Press, Oxford, 1987.

[11] D.F.Specht, Probabilistic neural networks, *Neural Networks 3*, 1990, 109-118.

[12] N.Weymaere and J. Martens, A fast robust learning algorithm for feedforward neural networks. *Neural Networks 4*, 1991, 361-369.

[13] L. Xu, Adam Krzyżak and E.Oja, Rival Penalized Competitive Learning for Clustering Analysis, RBF net and Curve Detection, *IEEE Trans. on Neural Networks*, 1992 (to appear).

ON THE FINITE SAMPLE PERFORMANCE OF THE NEAREST NEIGHBOR CLASSIFIER*

Demetri Psaltis[†], Robert R. Snapp[‡]
and
Santosh S. Venkatesh[§]

ABSTRACT

The finite sample performance of a nearest neighbor classifier is analyzed for a two-class pattern recognition problem. An exact integral expression is derived for the m-sample risk R_m given that a reference m-sample of labeled points, drawn independently from Euclidean n-space according to a fixed probability distribution, is available to the classifier. For a family of smooth distributions characterized by asymptotic exapansions in general form, it is shown that the m-sample risk R_m has a complete asymptotic series expansion $R_m \sim R_\infty + \sum_{k=1}^{\infty} c_k m^{-k/n}$ $(m \to \infty)$ where R_∞ denotes the nearest neighbor risk in the infinite-sample limit. Improvements in convergence rate are shown under stronger smoothness assumptions, and in particular, $R_m = R_\infty + O(m^{-2/n})$ if the class-conditional probability densities have uniformly bounded third derivatives on their probability one support. This analysis thus provides further analytic validation of Bellman's curse of dimensionality. Numerical simulations corroborating the formal results are included, and extensions of the theory discussed. The analysis also contains a novel application of Laplace's asymptotic method of integration to a multidimensional integral where the integrand attains its maximum on a continuum of points.

*The work reported here was supported in part by the Air Force Office of Scientific Research under grant AFOSR 89-0523 to Santosh S. Venkatesh.

[†]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125; *electronic mail*: psaltis@sunoptics.caltech.edu

[‡]CS/EE Department, University of Vermont, Burlington, VT 05405; *electronic mail*: snapp@uvm.edu

[§]Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104; *electronic mail*: venkatesh@ee.upenn.edu

The relative value of labeled and unlabeled samples in pattern recognition *

Vittorio Castelli
Stanford University

Thomas M. Cover
Stanford University

Abstract

We attempt to discover the role and relative value of labeled and unlabeled samples in reducing the probability of error of the classification of a sample based on the previous observation of labeled and unlabeled data. We assume that the underlying densities belong to a regular family that generates identifiable mixtures.

The unlabeled observations, under the above conditions, carry information about the statistical model and therefore can be effectively used to construct a decision rule. When the training set contains an infinite number of unlabeled samples, the first labeled observation reduces the probability of error to within a factor of two of the Bayes risk. Moreover subsequent labeled samples yield exponential convergence of the probability of classification error to the Bayes risk. We argue that labeled samples are exponentially more valuable than unlabeled samples and identify the exponent as the Bhatthacharyya distance.

Summary

Assume we sample from two populations, $\theta = 1$ and $\theta = 2$, with prior probabilities η and $\overline{\eta} = 1 - \eta$. Let observations from population 1 be distributed according to density $f_1(x)$, with respect to some measure μ, and observations from population 2 according to $f_2(x)$. We observe l independent samples together with their classifications, $\{(X_1, \theta_1), .., (X_l, \theta_l)\}$, where the θ_i are Bernoulli(η) and the X_i are i.i.d. $\sim f_{\theta_i}(x)$, and we observe u unlabeled samples $\{X'_1, .., X'_u\}$. The totality constitutes the training set.

Let X be a sample, similarly drawn, which we wish to classify with minimum probability of error. Let $R(l, u)$ be the probability of error of a given decision rule when the training set is composed of l labeled and u unlabeled samples.

If $f_1(x)$, $f_2(x)$ and η are known, the likelihood ratio test

$$\text{Decide} \quad \hat{\theta}(X) = \begin{cases} 1 & \text{if} \quad \dfrac{\eta f_1(X)}{(1-\eta) f_2(X)} \geq 1 \\ 2 & \text{if} \quad \dfrac{\eta f_1(X)}{(1-\eta) f_2(X)} < 1, \end{cases}$$

minimizes the probability of error (Bayes risk) which is equal to

$$R^* = \Pr\left\{\hat{\theta} \neq \theta\right\} = E_\mu\left[\min\left(\eta f_1(x), (1-\eta) f_2(x)\right)\right].$$

If $f_1(x)$ and $f_2(x)$ belong to a regular family $\mathcal{F}$, the distributions and the prior probabilities may be estimated from the labeled data. If an infinite number of labeled samples is available, the risk is given by $R(\infty, u) = R^*$ for any number u of unlabeled samples.

The distributions and prior probabilities can also be estimated using the unlabeled observations, under the additional hypothesis that the family of mixtures [1] generated by $\mathcal{F}$ is identifiable [2]. For example, $\mathcal{F}$ can be the family of Gaussian distributions with mean μ and variance σ^2. Then any mixture of the form $\eta\varphi(\mu_1, \sigma_1^2) + \overline{\eta}\varphi(\mu_2, \sigma_2^2)$ can be uniquely decomposed into its component densities.

Thus $u = \infty$ yields the information that the underlying distributions are either $(\eta f(x), \overline{\eta} g(x))$ or $(\overline{\eta} g(x), \eta f(x))$, but no information is available on whether $f_1(x) = f(x)$ or $f_1(x) = g(x)$. Thus for $l = 0$ labeled samples,

$$R(0, u) = \frac{1}{2} \qquad \text{for all } u.$$

Labeled data are therefore needed. The first labeled sample helps enormously.

Theorem. *When the training set contains an infinite number of unlabeled samples, the first labeled observation yields a probability of error*

$$R(1, \infty) = 2R^*(1 - R^*)$$

for the classification of a new sample.

The expected probability of classification error is thus reduced to within a factor two of the Bayes risk.

Theorem. *When the number of unlabeled samples is infinite, the risk converges to the Bayes risk exponentially fast, i.e.*

$$R(l, \infty) = R^* + O\left(e^{-l\alpha}\right)$$

where $\alpha = -\log\left(\int 2\sqrt{\eta\overline{\eta}}\,\sqrt{f_1(x) f_2(x)}\, d\mu(x)\right)$.

Labeled samples can reduce the risk exponentially fast, but unlabeled samples reduce the risk only polynomially fast. Under smoothness conditions on the family $\mathcal{F}$, similar to those that allow efficient estimation of parameters [3], there exists a procedure such that

$$R(l, u) = R^* + O(1/u) + O(e^{-l\alpha}).$$

Roughly speaking, labeled samples are exponentially more valuable than unlabeled samples.

References

[1] Teicher, Henry. "On the mixtures of distributions" *Ann. Math. Statist.* 1960, **32** 244-248.

[2] Teicher, Henry. "Identifiability of finite mixtures" *Ann. Math. Statist.* 1963, **34** 1265-1269.

[3] Lehmann E.L. *Theory of point estimation* 1983, John Wiley and Sons, New York.

[4] Cover T.M., Thomas J.A. *Elements of Information Theory*. 1991, John Wiley and Sons, New York.

[5] Duda R.O., HartP.E. *Pattern Classification and Scene Analysis*. 1973, John Wiley and Sons, New York.

[6] Andrews H.C. *Introduction to Mathematical Techniques in Pattern Recognition*. 1972, John Wiley and Sons, New York.

*This work was partially supported by NFS Grant NCR-8914538-02.

Vittorio Castelli (vittorio@isl.stanford.edu) and Thomas M. Cover (cover@isl.stanford.edu) are with the Information Systems Laboratory, Department of Electrical Engineering, 4055 Stanford University, Stanford, California 94305.

On the posterior probability estimate of the error rate of nonparametric classification rules

Gábor Lugosi
Department of Mathematics,
Technical University of Budapest
1521 Stoczek u. 2, Budapest, Hungary

and

Miroslaw Pawlak
Department of Electrical and Computer Engineering
The University of Manitoba
Winnipeg, Manitoba R3T 2N2, Canada

ABSTRACT We address the problem of estimating the error probability of nonparametric classification rules. Instead of the well known *counting-type estimators* we propose a so called *posterior* probability estimator, which plugs a nonparametric estimate of the a posteriori probabilities into an algebraic expression of the error probability. We explore the properties of the plug-in estimator. Unlike the standard estimators, the variance of our estimator is shown to have some remarkable distribution-free properties for the $k-$nearest neighbor, kernel and histogram rules. We pay special attention of histogram classification rules, and show the consistency of the estimate in this case. Investigating the bias of the estimate we also obtain rate-of-convergence results under mild conditions on the distribution.

I. INTRODUCTION

Let the random variable pair (X,Y) take its values from $\mathcal{R}^d \times \{0,1\}$. It is well known that the *decision rule* $g : \mathcal{R}^d \to \{0,1\}$ which minimizes the *error probability* $\Pr\{g(X) \neq Y\}$ is given by the Bayes decision:

$$g^*(x) = \begin{cases} 0 & \text{if } P_0(x) \geq P_1(x) \\ 1 & \text{otherwise} \end{cases}$$

($x \in \mathcal{R}^d$), where the $P_i(x) = \Pr\{Y = i | X = x\}$ $i = 0,1$ are the *a posteriori* probabilities. In practice the $P_i(x)$ are not known, but a *training sample* of n independent identically distributed (i.i.d.) random variable pairs

$$\xi_n = ((X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n))$$

is given, where the (X_i,Y_i) have the same distribution as that of (X,Y), and ξ_n is independent from (X,Y). Most nonparametric classification rules can be formulated as

$$g_n(x) = g_n(x,\xi_n) = \begin{cases} 0 & \text{if } P_0^{(n)}(x) \geq P_1^{(n)}(x) \\ 1 & \text{otherwise,} \end{cases} \qquad (1)$$

where $P_i^{(n)}(x) = P_i^{(n)}(x,\xi_n)$ is an estimate of $P_i(x)$ from the sample ξ_n. The error probability of the rule g_n is given by

$$L_n = \Pr\{g_n(X) \neq Y | \xi_n\} = 1 - E(P_{g_n(X)}(X)|\xi_n), \qquad (2)$$

Examples of this kind of classification rules are histogram, $k-$nearest neighbor and kernel rules.

It is always a crucial question to estimate the error probability of the classification rule. The most standard methods are based on counting the number of errors on the training data. Such estimates are the resubstitution, holdout and deleted (leave-one-out) estimates. These estimates have usually small bias, but their variance can be undesirably large. Another possible strategy is plugging an estimate of the a posteriori probabilities into the expression (2) of the error probability. In the case of nonparametric rules of form (1) it is natural to use the estimates that define the classification rule.

The formal definition of the estimator that we investigate is the following: Let $\xi_{n,j}$ denote the training sequence with (X_j,Y_j) deleted and $P_i^{(n-1)}(x,\xi_{n,j})$ the estimate of the a posteriori probability $P_i(x)$ from the data $\xi_{n,j}$ Then our estimate for the error probability is

$$L_n^{(P)} = 1 - \frac{1}{n}\sum_{j=1}^{n} P_{g_{n-1}(X_j,\xi_{n,j})}^{(n-1)}(X_j,\xi_{n,j}) \qquad (3)$$

The most remarkable property of the estime is summarized in the following result:

Theorem 1 *For any histogram, kernel and $k-$nearest neighbor classification rule, for all n and $t > 0$*

$$\Pr\{|L_n^{(P)} - EL_n^{(P)}| \geq t\} \leq 2e^{-cnt^2}$$

and

$$Var(L_n^{(P)}) \leq \frac{c}{8n}$$

regardless of the distribution of (X,Y), where the constant c depends on the dimension only.

The proof is based on McDiarmid's extension of Azuma's martingale inequality and geometrical considerations. The upper theorem suggests that using the proposed estimate could be favourable in practice as compared to counting-type estimates. To get more insight to the properties we also investigated the bias of the estimate. Here we list some of the results that we obtained for the case of the cubic histogram classification rule with cube-size $h > 0$. The first of them shows, that under the usual conditions on h the estimate is strongly consistent for all distributions.

Theorem 2 *For the cubic histogram rule for all distributions*

$$L_n^{(P)} - EL_{n-1} \to 0$$

with probability one whenever $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$.

The next one is an interesting property.

Theorem 3 *For the histogram rule $EL_n^{(P)} \leq EL_n$, that is, $L_n^{(P)}$ is always optimistically biased.*

The next rate-of-convergence results provide more insight into the behavior of the estimate:

Theorem 4 *If the distribution of X is of compact support, then for the histogram rule there exists a constant C such that for all n*

$$E|L_n^{(P)} - L_n| \leq \frac{C}{h^d\sqrt{n}}.$$

If in addition the support of the distribution is convex and satisfies $\mu(A) \geq mh^d$ for all measurable A for some $m > 0$, then

$$E|L_n^{(P)} - L_n| \leq \frac{C_1}{\sqrt{nh^d}}$$

for some C_1. If $P_0(x)$ is uniformly Lipschitz, that is, for some L $|P_0(x) - P_0(y)| \leq L\|x - y\|$ for any x, y , then

$$E|L_n^{(P)} - L_n| \leq Lh.$$

A BAYESIAN APPROACH FOR CLASSICIFICATION OF CONTINUOUS-TIME MARKOV SOURCES

Erdal Panayırcı

Istanbul Technical University
Faculty of Electrical and Electronics Engineering
Maslak 80626,Istanbul
Turkey

Abstract: *A Bayesian approach for classification of Markov source is developed and studied. Each of M sources is described by a continuous-time, discrete-state Markov chain All states and times of transitions between states can be observed perfectly but the transition rate matrices which establish the parameters of the sources are not known* a priori. *A Bayesian training algorithm using a fixed amount of memory digests the training samples that consist of a member function from each chain. This leads to an iterative computationally simple classification algorithm.*

Extended Summary

The general Bayesian classification problem can be described as follows. Let $\theta_s \in \Theta$ be the parameter set of the $s-th$ source $1 \leq s \leq M$, where Θ is the parameter space. We consider the unknown parameter set $\{\theta_s\}_{s=1}^{M}$ as both independent of each other and of the active source, and identically distributed random variables each governed by a *prior* probability density function(PDF) $\mu(\theta_s), s = 1,2,\ldots,M$.Let $\mathbf{y}_s = (y_{s1}, y_{s2}, \ldots, y_{sm_s}), s = 1,2,\ldots,M$, be a training sequence from the $s-th$ source.Let $\mathbf{x} = (x_1, x_2, \ldots, x_n), i = 1,2,\ldots,n$ be a test sequence to be classified,produced by one of the M sources, henceforth called the *active source*. The index of the active source is unknown and considered to be a discrete random variable ω taking values $\{1,2,\ldots,M\}$.

The classification problem is that of identifying the active source upon observing $\mathbf{x}$ and $Y = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$. The Bayes decision rule $\delta(\mathbf{x}|Y)$ which is optimal in the sense of minimizing the error probability of classification can be defined as

$$\begin{aligned} &\delta(\mathbf{x}|Y) = s \text{ if} \\ &\{p(H_\ell)p(\mathbf{x}|\mathbf{y}_\ell, H_\ell)\} \\ &\text{is maximum for } \ell = s \in \{1,2,\ldots,M\} \end{aligned} \tag{1}$$

Note that the decision rule $\delta(.)$ is a partition of the observation space U^n of all possible test sequences $\mathbf{x}$ into M disjoint regions $R_1, R_2, \ldots, R_m$ whose union equals U_n. Therefore the conditional error probability $p_\delta(e|Y)$ associated with a decision rule $\delta \equiv \delta(\mathbf{x}|Y)$ is defined as

$$p_\delta(e|Y) = \sum_{i=1}^{M} P(H_i) \sum_{\mathbf{x} \notin R_i} p(\mathbf{x}|\mathbf{y}_i, H_i) \tag{2}$$

where $\bar{R}_i$ is the complement of R_i, $p(H_i)$ is the prior probability of the $i-th$ source and, $p(\mathbf{x}|\mathbf{y}_i, H_i)$ is the conditional probability density of $\mathbf{x}$ given both the training sequence $\mathbf{y}_i$ and the $\mathbf{x}$ are generated from the $i-th$ source.

The posterior densities $p(\mathbf{x}|\mathbf{y}_s, H_s)$ in (2) are difficult to compute in general and depends on the prior densities of the unknown parameters which are usually unavailable. Recently, Merhav and Jiv [1] developed a suboptimal Bayesian test statistic for classification of discrete- time,discrete-state Markov sources whose transition probabilities are not known explicitly. They showed that the test does not require knowledge of prior densities and achieves,within a constant factor, the minimum error probability in Bayesian sense.Unfortunately, the main assumption of the approach that the prior density $\mu(.)$ must be bounded,*i.e.* for all $\theta \in \Theta$,

$$0 < \mu_{min} \leq \mu(\theta) \leq \mu_{max} < \infty$$

does not allow to apply this method to those problems whose parameter set have infinite support.

The study reported in this paper defines and solves a classification problem based on a Bayesian approach involving information sources each described by a continuous-time Markov chain,whose sample functions can be observed directly but whose parameters are not known *apriori*. As opposed to the approaches presented in [1], the posterior densities in (2) are computed exactly by choosing appropriate prior density functions for θ. We show that a *Naturel Conjugate prior density* exists for the problem investigated here and that the posterior PDF's have the same functional forms as the prior PDF's. Thus the classification analysis can be done by operating solely on the parameters of the prior densities updated by the training sequences. Naturel conjugate PDF's form a rich class of distributions, giving the classifier considerable flexibility in choosing a prior density for θ by setting suitable values for the prior parameters. The decision making and training algorithm derived in this way are optimal in Bayesian sense, as well as computationally simple, recursive and require a fixed amount of computer storage regardless of features in $\mathbf{x}$ and Y.

References

[1]N.Merhav and J.Jiv,"A Bayesian approach for classification of Markov sources,*IEEE Trans.Infom.Theory*, vol.37,no.4,July 1991.

A Computer Algebra Algorithm for the Adjoint Divisor

D. Polemi, M. Hassner, O. Moreno, and C.J. Williamson

Abstract. Using the algorithm in [1] for desingularizing a singular plane curve, we describe a polynomial time algorithm, which can be used for computing the adjoint divisor, finding the genus or adding points on the Jacobian of the curve. We also do a complexity analysis of the mentioned algorithm. The algorithm can be implemented in the IBM computer algebra system SCRATCHPAD.

Summary. We denote by $\mathbf{k} = \mathbf{F}_q$ a finite field of $q = p^e$ elements where p is a prime number, and by $\overline{\mathbf{k}}$ the algebraic closure of $\mathbf{k}$. Consider a projective plane singular curve C of degree d, defined over $\mathbf{k}$, specified by a polynomial $f(x,y) \in \mathbf{k}[x,y]$ which is irreducible in $\overline{\mathbf{k}}[x,y]$. By definition C is the variety $C = \{(a,b) \in \overline{\mathbf{k}}^2 : \quad f(a,b) = 0\}$. The affine coordinate ring $\mathbf{k}[C]$ of C is the quotient ring $\mathbf{k}[x,y]/I$ where I is the principal ideal generated by f. We denote the field of rational functions of C by $\mathbf{k}(C)$. The next is an outline of a construction for the adjoint divisor.

Step A. Construct an affine non-singular model $\tilde{C}$ (smooth curve) of the curve C.

For this step we use the algorithm in [1] which can be outlined as follows:

1. Compute an integral basis $\{u_0, ..., u_n\}$ of the integral closure $\overline{\mathbf{k}[C]}$ of $\mathbf{k}[C]$ considered as a $\mathbf{k}[X]$-module. ($\mathbf{k}[X]$ is the ring of polynomials of one variable.)

2. Introduce new variables $\{X, Y_1, ..., Y_n\}$ which correspond to the basis $\{u_0, ..., u_n\}$. Then the non-singular model is $\tilde{C} = \{a \in \overline{\mathbf{k}}^{n+1} : \quad f_{ij}(a) = 0 \forall i,j\}$, where $1 \le i,j \le n$, $f_{ij} \in \mathbf{k}[X, Y_1, ..., Y_n]$ are polynomials of the form $f_{ij} = Y_iY_j - \sum_{\alpha=0}^{n} c_{ij}^{\alpha}(X)Y_\alpha$ and $c_{ij}^{\alpha} \in \mathbf{k}[X]$. The coordinate ring of $\tilde{C}$ is $\mathbf{k}[\tilde{C}] = \mathbf{k}[X, Y_1, ..., Y_n]/(f_{ij})$. There is a natural map $\pi : \tilde{C} \to C$ inducing the map $\pi^* : \mathbf{k}(C) \to \mathbf{k}(\tilde{C})$ on the field of rational functions of the curves [1](p.9).

Step B. Compute the points $P_{k,l} \in \pi^{-1}(P_k)$ lying over the singular points $P_k = (\alpha_k, \beta_k) \in C$.

In particular, find the solutions of the systems of equations $f_{ij}(\alpha_k, Y_1, ..., Y_n) = 0$ where $i = 1, ...n$, $j = 1, ...n$.

Step C. Compute the *adjoint divisor* of the curve C.

i. Form the differential $\omega = \pi^*\left(\frac{dx}{\partial f/\partial y}\right)$ in the field of differentials $\Omega(\tilde{C})$.

ii. Compute the *local uniformizing parameter* (l.u.p.) around each point $P_{k,l} \in \pi^{-1}(P_k)$ by forming the matrix $\|\frac{\partial f_{ij}}{\partial Y_j}\|$, $1 \le i,j \le n$. If $\frac{\partial f_{ij}}{\partial Y_i}(P_{k,l}) = 0$ for some i, then Y_i is the l.u.p. around $P_{k,l}$.

iii. Express the variables X, $Y_1, ...Y_n$ in terms of the l.u.p. Y_i, and estimate the *orders*, $\mathrm{ord}_{P_{k,l}}X$, $\mathrm{ord}_{P_{k,l}}Y_j$ $j = 1, ...n$. Furthermore, compute the orders, $\mathrm{ord}_{P_{k,l}}\omega$, of ω at $P_{k,l}$.

iv. The *adjoint divisor of the curve C at P_k* is $\Delta_{P_k} = \sum_{k,l} -\mathrm{ord}_{P_{k,l}}\omega\, P_{k,l}$, of degree $\delta_k = \sum_{k,l} -\mathrm{ord}_{P_{k,l}}\omega$. The *adjoint divisor of C* is $\Delta = \sum_k \Delta_{P_k}$, with degree $\delta = \sum_k \delta_{P_k}$, furthermore $\delta = 2\dim_{\mathbf{k}} \overline{\mathbf{k}[C]}/\mathbf{k}[C]$.

Step D. Find the genus of the curve C.

Using the computation of δ in Step C, the genus of C is $g = \frac{(d-1)(d-2)}{2} - \frac{1}{2}\delta$. The time complexity of the above algorithm is of the order $O(q^2r^2d^6\log^3 q)$, where r is the multiplicity of C at its worst singular point; it is better than other algorithms since the desingularization technique in Step A does not require field extensions.

Example. Let $C : f(x,y) = x^2 + xy + y^4 = 0$ over $\mathbf{k} = \mathbf{F}_2$ with an ordinary singular point $P_1 = [0,0,1]$ of multiplicity $r = 2$, and $\mathbf{k}[C] = \mathbf{k}[x,y]/(x^2 + xy + y^4)$. An integral basis of $\overline{\mathbf{k}[C]}$ is $\{1, y, y^2, y^3/x\}$ and $\tilde{C} = \{a \in \overline{\mathbf{k}}^4 : f_{ij}(a) = 0\}$, where $f_{11} = Y_1^2 - Y_2$, $f_{12} = Y_1Y_2 - xY_3$, $f_{13} = Y_1Y_3 - (x + Y_1)$, $f_{22} = Y_2^2 - x^2 - xY_1$, $f_{23} = Y_2Y_3 - Y_2 - xY_1$, $f_{33} = Y_3^2 - Y^2 - Y_3$ and $\pi^{-1}(P_1) = \{P_{11} = (0,0,0,0), P_{12} = (0,0,0,1)\}$. The l.u.p. around P_{11} and P_{12} is Y_1. We rewrite $Y_2 = Y_1^2$, $Y_3 = Y_3^2 - Y_1^2$, and $x = Y_1Y_3 - Y_1$. Thus $\mathrm{ord}_{P_{11}}Y_2 = 2$, $\mathrm{ord}_{P_{12}}Y_2 = 2$, $\mathrm{ord}_{P_{11}}Y_3 = 2$ and $\mathrm{ord}_{P_{12}}Y_3 = 2$, $\mathrm{ord}_{P_{11}}x = 1$ and $\mathrm{ord}_{P_{12}}x = 1$. Let $\omega = dx/x$, then $\mathrm{ord}_{P_{11}}\omega = -1$ and $\mathrm{ord}_{P_{12}}\omega = -1$. Hence $\Delta = P_{11} + P_{12}$ and $g = 1$.

Following similar methods as in Huang and Ierardi [3](§5), adding points on the Jacobian of a plane singular curve can also be done using our algorithm.

We can generalize the algorithm of [1], which is for desingularizing plane curves, to the case of an arbitrary curve, using similar methods as in [4]. In this way we can generalize our algorithms above. Furthermore, since the teqniques of [4] depend heavily on the theorem of the primitive element, we can also obtain an effective method to generalize our results in the case of an arbitrary curve, using [2].

References

[1] M. Bronstein, M. Hassner, A. Vasquez, and C.J. Williamson. Computer algebra algorithms for the construction of error correcting codes on algebraic curves. *IEEE Proceedings on Information Theory*, June 1991.

[2] B. Buchberger, G.E. Collins, and R. Loos. *Computer Algebra, Symbolic and Algebraic Computation*. Springer-Verlag, New York, 1982.

[3] M.D. Huang and D.J. Ierardi. Efficient algorithm for the Riemann-Roch theorem and for addition in the jacobian of a curve. *IEEE Symposium on the Foundations of Computer Science*, pages 678–687, 1991.

[4] D. Polemi, C. Moreno, and O. Moreno. A construction of a.g. Goppa codes from singular curves. ***submitted for publication***, 1992.

Codes over Gaussian Integers

Klaus Huber, FI 17d
Deutsche Bundespost Telekom
Research Institute
P.O.Box 10 00 03
6100 Darmstadt
Germany

1 Introduction

In this contribution we give an algebraic approach for coding over two-dimensional signal space for QAM-like constellations. The two main points are an isomorphic mapping of fields $GF(p)$, $p \equiv 1 \bmod 4$ onto a subset of the Gaussian integers, and a new two-dimensional modular distance called Mannheim distance.

2 $\mathcal{G}_\pi$, Mannheim Distance, and Error Correcting Codes

Gaussian integers are those complex numbers which have integers as real and imaginary parts (for Gaussian integers see e.g. [2], pp.182-187). Primes of the form $p \equiv 1 \bmod 4$ can be written in exactly one way as sum of two squares. Hence such primes p are the product of two conjugate complex Gaussian integers: $p = a^2+b^2 = \pi \cdot \pi^*$ where $\pi = a+i \cdot b$ and $*$ denotes complex conjugation $\pi^* = a - i \cdot b$. Let $[.]$ denote rounding to the closest integer and define rounding of a complex number by $[x+iy] = [x]+i[y]$. Then the modulo function

$$\mu(g) = g \bmod \pi = \gamma = g - [\frac{g \cdot \pi^*}{\pi \cdot \pi^*}] \cdot \pi$$

maps $GF(p) \rightarrow \mathcal{G}_\pi$ where $\mathcal{G}_\pi$ denotes the residue class of the Gaussian integers modulo π. In figure 1 $\mathcal{G}_{3+2i}$ is displayed.

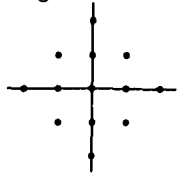


Figure 1 : $\mathcal{G}_{3+i\cdot 2}$

Clearly the inverse mapping then immediately follows as $\mu^{-1}(\gamma) = g = \gamma v \pi^* + \gamma^* u \pi \bmod p$, where $1 = u\pi + v\pi^*$.

To profit algebraically from the representation of $GF(p)$ as Gaussian integers, we introduce a two-dimensional modular distance which we call Mannheim distance. The Mannheim distance $d_M(\alpha,\beta)$ between two Gaussian integers α and β is defined as $d_M(\alpha,\beta) = \mathrm{Re}\{\gamma\} + \mathrm{Im}\{\gamma\}$, where $\gamma = (\beta - \alpha) \bmod \pi$, i.e. the Mannheim-distance is the well-known Manhattan-distance modulo a two-dimensional grid. (The streets/avenues of Manhattan and Mannheim form a rectangular grid.) In a straightforward way we define the Mannheim weight of $\gamma \in \mathcal{G}_\pi$ as $w_M(\gamma) = d_M(\gamma,0)$, and the Mannheim weight of a vector $\mathbf{r} = (r_0, r_1, \ldots, r_{n-1})$ over $\mathcal{G}_\pi$ as $w_M(\mathbf{r}) = \sum w_M(r_j)$. Similar to the usual Hamming distance codes we characterize linear Mannheim error correcting codes by the triple $[n,k,d_M]$ where n is the length, k the dimension, and

$$d_M = \min\{w_M(\mathbf{c}) | \mathbf{c} \neq \mathbf{0}, \mathbf{c} \in C\}$$

the minimum Mannheim distance of the code. We start with the design of perfect icyclic One Mannheim Error Correcting (OMEC) codes which are able to correct errors of Mannheim weigth one. Then Mannheim error correcting codes having $d_M \geq 3$ are designed and decoders working in a similar way as Berlekamp's negacyclic codes for the Lee distance (see [1], pp.207-217). The codes are 90-degree rotationally invariant and are very easy to encode and decode. Synchronization is also very easy. The coding gains which can be achieved are considerable, in particular if simple code concatenations are considered. For primes $p \equiv 3 \bmod 4$ we can use $\mathcal{G}_{ip}$ which is isomorphic to $GF(p^2)$.

To give the flavour of the ideas, consider the perfect $[n, n-1, 3]$ OMEC-codes defined by the parity check matrix

$$\mathbf{H} = \left(\alpha^0, \alpha^1, \alpha^2, \ldots, \alpha^{\frac{(p-1)}{4}-1}\right),$$

where α is a primitive element of $\mathcal{G}_\pi$. Hence $\alpha^{\frac{p-1}{4}} \in \{\pm i\}$, and all errors of Mannheim weight ≤ 1 will produce different syndromes. Decoding is straightforward: Take the received vector $\mathbf{r} = \mathbf{c} + \mathbf{e}$ and compute the syndrome $(s) = \mathbf{H} \cdot \mathbf{r}^\top$. The location of an error having $w_M(\mathbf{e}) = 1$ is then given by $l = \log_\alpha s \bmod \frac{(p-1)}{4}$ and its value by $s \cdot \alpha^{-l}$.

Example *Let $p = 13$, $\pi = 3 + i \cdot 2$, and $\alpha = 1 + i$, then*

$$\mathbf{H} = (1, 1+i, 2i)$$

Let us assume that at the receiving end we get the vector $\mathbf{r} = (1+i, i, -1+i)$, then $s = \mathbf{H} \cdot \mathbf{r}^\top = -2 = \alpha^{11}$ and we find that at position $2 \equiv 11 \bmod 3$ we have an error value of $s \cdot \alpha^{-2} = i \Rightarrow \mathbf{e} = (0,0,i)$, $\Rightarrow \mathbf{c} = \mathbf{r} - \mathbf{e} = (1+i, i, -1)$.

References

[1] E.R.Berlekamp, "Algebraic Coding Theory", Aegean Park Press 1984.

[2] G.H.Hardy, E.M.Wright, "An introduction to the theory of numbers", fifth edition, Oxford 1979.

CONSTRUCTION OF LINEAR BLOCK CODES OVER GROUPS

Ezio Biglieri *Michele Elia*

Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy).

Let $\mathcal{G}$ be a finite group with a multiplicative operation and identity element e. A *block code* $\mathcal{C}$ of length n over $\mathcal{G}$ is any non-empty subset of the n-fold direct product $\mathcal{G}^n$, i.e., of the set of all the n-tuples of group elements. We assume the group order $|\mathcal{G}|$ to be finite. The *dimension* of a code $\mathcal{C}$ is $k = \log_{|\mathcal{G}|}|\mathcal{C}|$ symbols per block, where $|\mathcal{C}|$ is the code size, bounded above by $|\mathcal{G}|^n$. The code *rate* is $r = k/n$. The *Hamming distance* between two code words is the number of positions in which they differ. Let $\mathcal{I}$ denote the index set of the n-tuples of $\mathcal{C}$. An *information set* of $\mathcal{C}$ [1] is any index subset $\mathcal{J} \subseteq \mathcal{I}$ of size $|\mathcal{J}| = k$ such that every k-tuple of elements of $\mathcal{G}$ occurs in $\mathcal{J}$ precisely once as the code words run through $\mathcal{C}$. Codes exist without an information set.

The direct product of $\mathcal{G}$ by itself n times, say $\mathcal{G}^n$, forms a group. A *linear* block code over $\mathcal{G}$ is a subset of $\mathcal{G}^n$ that forms a group, i.e., is a subgroup of $\mathcal{G}^n$. This paper is devoted to the description of such subgroups.

In algebraic coding theory, the "classical" construction of linear codes concatenates a k-tuple of information symbols with $n-k$ *check symbols* chosen so as to satisfy certain linear *parity check equations*. We show that this construction can be mimicked to generate linear codes over groups.

Definition 1. *An (n,k) systematic block code $\mathcal{C}$ with block length n and dimension k over a group $\mathcal{G}$ is a subgroup of $\mathcal{G}^n$ with order $|\mathcal{G}|^k$ formed by the n-tuples*

$$(x_1,\ x_2,\ \ldots,\ x_k,\ y_1,\ \ldots,\ y_{n-k}) \tag{1}$$

with $y_i = \Phi_i(x_1,\ x_2,\ \ldots,\ x_k)$ where Φ_i are $(n-k)$ maps of $\mathcal{G}^k$ into $\mathcal{G}$.

For linearity, the maps Φ_i must be homomorphisms:

Proposition 1. *The (n,k) systematic code with code words (1) is linear if and only if the maps Φ_i are homomorphisms of $\mathcal{G}^k$ into $\mathcal{G}$.*

A more compact definition of a linear code can now be provided. We denote by $X^{(m)}$ the elements of $\mathcal{G}^m$.

Definition 2. *A linear (n,k) systematic block code $\mathcal{C}$ over a group $\mathcal{G}$ is the image of an endomorphism Ψ of $\mathcal{G}^n$:*

$$\Psi(X^{(k)} \mid Y^{(n-k)}) = (X^{(k)} \mid [\Phi(X)]^{(n-k)})$$

where Φ is a homomorphism of $\mathcal{G}^k$ into $\mathcal{G}^{n-k}$.

The following proposition shows that the actual algebraic structure of a linear code does not carry information about the properties of the code itself.

Proposition 2. *All linear (n,k) systematic codes over the same group $\mathcal{G}$ are isomorphic.*

We are especially interested in linear codes that cannot be obtained by concatenating shorter codes. A group $\mathcal{G}$ is called *indecomposable* [2, p. 121] if $\mathcal{G} \neq \{e\}$, and if $\mathcal{G} = \mathcal{H} \times \mathcal{K}$ implies either $\mathcal{H} = \{e\}$ or $\mathcal{K} = \{e\}$. Consequently, we define linear indecomposable codes as follows.

Definition 3. *A linear (n,k) systematic block code is called indecomposable if it is an indecomposable group.*

The code words of a decomposable code can be written (possibly after reordering its components) as the concatenation of two words, each one with its information set and such that the parity check symbols depend only on one word or the other:

$$(x_1,\ \ldots,\ x_{k_1},\ y_1,\ \ldots,\ y_{\ell_1} \mid x_{k_1+1},\ \ldots,\ x_k,\ z_1,\ \ldots,\ z_{\ell_2})$$

with $n = k + \ell_1 + \ell_2$, and each y_i depends only on $x_1, \ldots x_{k_1}$ while each z_j depends only on $x_{k_1+1}, \ldots x_k$. If this is the case, we write $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$.

From the point of view of Hamming distance, linear codes over non-abelian groups $\mathcal{G}$ are bad. In fact, we prove that under certain mild conditions any (n,k) systematic linear code over $\mathcal{G}$ is decomposable into k repetition codes, and consequently we have the following upper bound to its minimum Hamming distance:

$$d \leq \left\lfloor \frac{n}{k} \right\rfloor. \tag{2}$$

Bound (2) improves upon a previous result by Forney [1], who proved that $d \leq n - 2k + 2$ for $k \leq n/2$ and $d = 1$ for $k > n/2$.

By examining in more detail the construction of linear codes over abelian groups, we prove that they can be characterized by a *parity-check matrix* $\mathbf{H}$ that describes the parity-check symbols $y_1, \cdots, y_{n-k}$ in (1) by expressing them as powers of the generators of $\mathcal{G}$.

The simplest case is that of a cyclic $\mathcal{G}$:

Theorem 1. *Consider the (n,k) linear code $\mathcal{C}$ over the cyclic group $\mathbf{Z}_t$ of order t. The parity check matrix $\mathbf{H}$ is an $(n-k) \times n$ echelon matrix over the ring $\mathcal{Z}_t$. The minimum Hamming distance d of the code is equal to the minimum number of linearly dependent columns of $\mathbf{H}$.*

Example 1. Let $n = 5$ and $k = 2$. A (5,2,3) code over $\mathbf{Z}_2$ is defined by the parity check matrix over $\mathcal{Z}_2$

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

It is easy to check that the minimum number of linearly dependent columns is 3. Thus, the minimum Hamming distance of the code is 3. If g denotes a generator of $\mathbf{Z}_2$, the code words have the form

$$(g^{x_1}, g^{x_2} \mid g^{x_1+x_2}, g^{x_2}, g^{x_1}).$$

A similar theorem can be proved for general abelian groups.

Theorem 2. *Consider the linear systematic (n,k) code $\mathcal{C}$ over an abelian group $\mathcal{A}$ of exponent d_m. The parity-check matrix $\mathbf{H}$ is a $m(n-k) \times mn$ echelon matrix over $\mathcal{Z}_{d_m}$. The minimum Hamming distance d of the code is given by the minimum number of linearly dependent columns in matrix $\mathbf{H}$ over the ring $\mathcal{Z}_{d_m}$, where certain sets of columns are accounted for 1.*

Example 2. Let $n = 5$ and $k = 2$. A (5,2,3) linear code over $\mathbf{Z}_2 \times \mathbf{Z}_4$ exists. The code is defined by the parity check matrix over $\mathcal{Z}_4$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 3 & 1 & 2 & 2 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Direct computation shows that he minimum Hamming distance of this code is 3. Its code words have the form

$$(g_1^{x_1} g_2^{x_1},\ g_1^{x_2} g_2^{x_2} \mid g_1^{x_1+x_2} g_2^{3x_1+x_1+2x_2+2x_2},\ g_1^{x_1+x_2} g_2^{x_1+x_1+2x_2},\ g_1^{x_1+3x_1+x_2} g_2^{x_1+x_2+2x_2}).$$

References

[1] G. D. Forney, Jr., "Linear codes over nonabelian groups are not good Hamming distance codes," submitted for publication, September 1991.

[2] J. J. Rotman, *An Introduction to the Theory of Groups*. Dubuque, Iowa: Wm. C. Brown Publishers, 1988.

This research was sponsored by the Italian National Research Council (CNR) under "Progetto Finalizzato Trasporti."

DEBRUIJN SEQUENCES, IRREDUCIBLE CODES AND CYCLOTOMY

E. R. Hauge and T. Helleseth

University of Bergen, Department of Informatics, HiB, N-5020 Bergen, Norway. E-mail: ErikR.Hauge@ii.uib.no, Tor.Helleseth@ii.uib.no
Partly supported by the Norwegian Research Council for Science and the Humanities (NAVF).

Abstract

The cycle join algorithm is applied to construct deBruijn sequences from irreducible cyclic codes. The number of sequences obtained by this construction is shown to be related to the cyclotomic numbers. The Matrix-tree theorem and Gaussian sums give a bound on the number of sequences constructed in this way.

Introduction

A binary deBruijn sequence of length 2^n is a sequence such that all n-tuples occur exactly once in the sequence. The binary deBruijn graph of order n, is a directed graph with 2^n nodes, each labeled with a unique binary n-dimensional vector, and an edge from node $S = (s_0, s_1, ..., s_{n-1})$ to node $T = (t_0, t_1, ..., t_{n-1})$ if and only if $(s_1, s_2, ..., s_{n-1}) = (t_0, t_1, ..., t_{n-2})$. Any nonsingular feedback function $f(s_0, s_1, \cdots, s_{n-1}) = s_0 + g(s_1, s_2, \cdots, s_{n-1})$ decomposes the deBruijn graph into disjoint cycles if we let the successor of the node $S = (s_0, s_1, \cdots, s_{n-1})$ be the node $(s_1, s_2, \cdots, s_{n-1}, f(s_0, s_1, \cdots, s_{n-1}))$. From such a decomposition one can construct deBruijn sequences by the well known cycle join method, which consists of joining all the cycles stepwise into a deBruijn sequence. To join two cycles C_1 and C_2 one finds a node $S = (s_0, s_1, \cdots, s_{n-1}) \in C_1$ and a conjugate node $\widehat{S} = (s_0 + 1, s_1, \cdots, s_{n-1}) \in C_2$. After interchanging the successors of S and $\widehat{S}$ (which corresponds to changing one value of the function g), the two cycles will be joined to one cycle. Repeating this process will eventually lead to a deBruijn sequence.

The number of deBruijn sequences that can be obtained from this construction will depend on the number of conjugate pairs on all pairs of cycles generated by the function f, which is the starting point of our construction.

We apply the cycle join method to construct deBruijn sequences from irreducible cyclic codes, i.e. f is a linear recurrence with an irreducible characteristic polynomial. We give an algebraic expression for the number of deBruijn sequences constructed in this way in terms of the cyclotomic numbers.

Methods and results

Let

$$f(s_0, s_1, \cdots, s_{n-1}) = \sum_{i=0}^{n-1} h_i s_i$$

where the characteristic polynomial $h(x) = x^n + \sum_{i=0}^{n-1} h_i x^i$ is irreducible of degree n over $GF(2)$. Then f will generate E cycles of length e in the deBruijn graph, where $eE = 2^n - 1$ and n is the smallest integer such that $2^n \equiv 1 \pmod{e}$.

Let $h(x)$ be the minimum polynomial of an element $\beta = \alpha^E$ for some primitive $(2^n - 1)$-th root of unity α in $GF(2^n)$. We can express α^j by

$$\alpha^j = \sum_{i=0}^{n-1} a_{j,i}\beta^i, \quad 0 \le j \le 2^n - 1.$$

We define the mapping $\phi : GF(2^n) \to GF(2)^n$ by $\phi(0) = (0, 0, \cdots, 0)$ and $\phi(\alpha^l) = (a_{l,0}, a_{l+E,0}, \cdots, a_{l+(n-1)E,0})$. It follows that ϕ is a vector space isomorphism and that the two elements $\phi(\theta), \phi(\theta + 1) \in GF(2^n)$ are conjugated for all $\theta \in GF(2^n)$, $\phi(1) = (1, 0, ..., 0)$.

For any $eE = 2^n - 1$, the cyclotomic classes C_i, $0 \le i < E$ in $GF(2^n)$ are defined as follows:

$$C_i = \left\{\alpha^{i+jE} | 0 \le j < e\right\}.$$

The cyclotomic number $(i, j)_E$ is defined for $0 \le i, j < E$ by

$$(i, j)_E = \#\left\{(\xi, \xi') | \xi \in C_i, \xi + 1 = \xi' \in C_j\right\}.$$

The cyclotomic numbers have been extensively studied in the literature.

The crucial observation is that under the mapping ϕ above, the cycles in the irreducible cyclic code generated by $h(x)$ correspond to the cyclotomic classes. Further, the number of conjugated pairs between the cycles equals the cyclotomic numbers. The exact number of deBruijn sequences obtainable in this way can now be found from the Matrix-tree Theorem on the graph where the cycles of f are nodes and the edges correspond to conjugated pairs, since each tree in this graph represents a deBruijn sequence. Hence, we can obtain an algebraic expression for the number of deBruijn sequences obtained by this construction in terms of the cyclotomic numbers. Using Gaussian sums to approximate the cyclotomic numbers we are able to show the following result.

Theorem. *The number of deBruijn sequences constructed by the cycle join method starting from the cycles generated by an irreducible polynomial $h(x)$ is at least*

$$\frac{1}{E}\left(\frac{2^n - 2E - (E-1)^2 2^{n/2}}{E}\right)^{E-1}.$$

References

[1] L.D. Baumert and R.J. McEliece, "Weights of irreducible cyclic codes," Information and Control, vol.20, pp. 158-175, March 1972.

[2] H. Fredricksen, "A survey of full length nonlinear shift register cycle algorithms," SIAM Review, vol. 24, pp.195-221, April 1982.

[3] T. Storer, Cyclotomy and Difference Sets, Chicago: Markham: Publishing Company, 1967.

M-SEQUENCES AND DUAL BASES OVER GF(q^m)

John J. Komo and William J. Reid III
Electrical and Computer Engineering Department
Clemson University - 211 Riggs Hall, Box 340915
Clemson, SC 29634-0915 USA

Abstract

An m-sequence over GF(q^m) can be expressed as a vector of m-sequences whose component m-sequences are shifted versions of an m-sequence over GF(q). The amount of shift between components of the m-sequence over GF(q^m) is given as a ratio of elements of the trace dual basis corresponding to the basis expressing GF(q^m) over GF(q). An efficient algorithm, which does not require evaluation of the trace, is developed for obtaining the ratio of trace dual basis elements to the first element. The dual basis can then be completely obtained by evaluating the first dual basis element. Another algorithm is developed which efficiently evaluates this first element. Included in this algorithm is a sequential evaluation of the trace which can be sequentially obtained directly in terms of the coefficients of the primitive polynomial that generates GF(q^m).

Summary

Let $\{\lambda_0,\lambda_1,\cdots,\lambda_{m-1}\}$ be the trace dual basis of the basis $\{1,\gamma,\cdots,\gamma^{m-1}\}$ generated by the primitive polynomial h(x) expressing GF(q^m) over GF(q). Also, let g(x) be the degree mn primitive polynomial with root α that generates the m-sequence c over GF(q) and f(x), which divides g(x), be the degree n primitive polynomial with root α that generates the vector m-sequence d over GF(q^m). Then, for some e_0, d can be expressed as [1]

$$d=\sum_{j=0}^{m-1}\gamma^j T^{e_0+zk_j w} c, \tag{1}$$

where $T^i c$ indicates a left shift of c by i elements, $\gamma=\alpha^{zw}$, $z=(q^{mn}-1)/(q^m-1)$, $\gcd(w,q^m-1)=1$, and $\gamma^{k_j}=\lambda_j/\lambda_0$. The shift of the jth component relative to the 0th component is then given as $zk_jw \bmod (q^{mn}-1)$, where w is used for choosing one of the $\phi(q^m-1)/m$ basis or primitive polynomials for expressing GF(q^m) over GF(q) and $\phi(\cdot)$ is the Euler ϕ function.

The dual basis can be obtained in terms of $\mathbf{B}=\mathbf{A}^{-1}$ where [2]

$$\mathbf{A}=\begin{bmatrix} tr_1^m(1) & tr_1^m(\gamma) & tr_1^m(\gamma^2) & \cdots & tr_1^m(\gamma^{m-1}) \\ tr_1^m(\gamma) & tr_1^m(\gamma^2) & tr_1^m(\gamma^3) & \cdots & tr_1^m(\gamma^m) \\ \vdots & \vdots & \vdots & & \vdots \\ tr_1^m(\gamma^{m-1}) & tr_1^m(\gamma^m) & tr_1^m(\gamma^{m+1}) & \cdots & tr_1^m(\gamma^{2m-2}) \end{bmatrix}. \tag{2}$$

Representing the components of the m-sequence generated by h(x) as $a_n=tr_1^m(\theta\gamma^n)$, a modified **A**, **A'** (exact if $\theta=1$), with initial conditions $a_0=1$ and $a_i=0$, i=1,2,...,m-1 can be evaluated without calculating the trace. Due to the symmetry and triangular nature of **A'**, **B**, the modified **B**, reduces to

$$\mathbf{B}'=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & b_3 & \cdots & b_{m-1} & b_m \\ 0 & b_3 & b_4 & \cdots & b_m & 0 \\ \vdots & \vdots & & & \vdots & \\ 0 & b_{m-1} & b_m & \cdots & 0 & 0 \\ 0 & b_m & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{3}$$

The ratio of dual basis elements can be obtained, using (3), as

$$\gamma^{k_{m-i}}=\lambda_{m-i}/\lambda_0 = b_m\Big[\sum_{j=1}^{i} h_{m-i+j}\gamma^j\Big], \quad i=1,2,...,m-1, \tag{4}$$

where

$$b_m=\begin{cases} \gamma^{(q-2)(q^m-1)/(q-1)}, & q \text{ and } m \text{ odd or } q \text{ even} \\ \gamma^{[(q-3)/2](q^m-1)/(q-1)}, & q \text{ odd and } m \text{ even.} \end{cases} \tag{5}$$

If h(x) is a trinomial of the form $h(x)=x^m+h_1x+h_0$, **B'** is a monomial matrix and the shift factors k_{m-i}, i=1,2,...,m-1, are obtained from $\gamma^{k_{m-i}}=b_m\gamma^i$ as

$$k_{m-i}=\begin{cases} (q-2)\dfrac{q^m-1}{q-1}+i, & q \text{ and } m \text{ odd or } q \text{ even} \\ \left(\dfrac{q-3}{2}\right)\dfrac{q^m-1}{q-1}+i, & q \text{ odd and } m \text{ even} \end{cases} \quad i=1,2,...,m-1. \tag{6}$$

The determination of the dual basis as opposed to the ratio of dual basis elements requires the evaluation of 2m-1 trace functions as shown in **A**. An algorithm for the evaluation of the trace directly in terms of the coefficients of the primitive polynomial h(x) that generates GF(q^m) and traces of smaller powers of α is also developed here as

$$tr_1^m(\alpha^i)=-\sum_{j=1}^{i-1} h_{m-j}tr_1^m(\alpha^{i-j}) - ih_{m-i}, \quad i=1,2,...m. \tag{7}$$

As usual, $tr_1^m(1)=m$ and for $i\geq m$, $tr_1^m(\alpha^i)$ can be expressed as a linear combination of $tr_1^m(\alpha^j)$, j=0,1,..., m-1. Thus, λ_0 can be obtained by solving for the first column of **B** and the remaining dual basis elements obtained by multiplying (3) by λ_0 or the entire dual basis can be obtained by solving for all of **B**.

References

[1] J. J. Komo and M. S. Lam, "Primitive polynomials and m-sequences over GF(q^m)", *IEEE Trans. Inform. Theory*, to appear March 1993.

[2] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*, Kluwer Academic Publishers, 1987.

Linear Recurrences on 2D Convex Lattices and Decoding of Some Codes from Algebraic Curves †

Shojiro Sakata
Toyohashi University of Technology
Department of Knowledge-Based Information Engineering

Abstract: We present a theory of linear recurrences defined on convex lattices in the 2D plane and propose a generalization of the 2D Berlekamp-Massey algorithm which finds a minimal set of linear recurrences capable of generating a 2D array on a 2D covex lattice. Furthermore we show that this algorithm is applicable to decoding efficiently some kinds of algebraic geometry codes, in particular codes introduced by S. Miura and N. Kamiya.

† This manuscript is a revised and extended version of the paper which was presented partly at Symposium on Information Theory and Its Application (SITA'91), at Ibusuki, Japan, on Dec. 11-14, 1991.

BOUNDS FOR LINEAR BLOCK CODES OVER RINGS

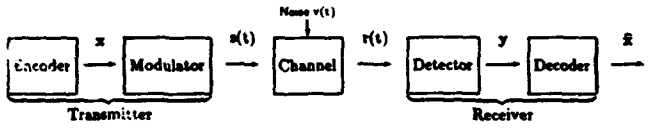
Magnus Nilsson
Department of Electrical Engineering
Linköping University
S-58183 Linköping, Sweden

Abstract

Linear block codes over rings are discussed for q-ary PSK over a Gaussian channel. An upper bound on the minimum Euclidean distance is given and proved. A lower bound on the block error probability is discussed.

Introduction

We discuss bounds for the following PSK-system:



The code is a linear block code over Z_q of block length n and the modulation is q-ary PSK with energy E_N per dimension. The channel adds white Gaussian noise. The detector has q congruent regions and the decoder makes error correction. The modulator, the channel and the detector together form a memoryless additive q-ary channel. Thus, $\mathbf{y} = \mathbf{x} + \mathbf{z} \pmod q$, where the probability of an error pattern $\mathbf{z}$ is a function of the SNR and of the composition of $\mathbf{z}$.

Definition 1. Let q, m and n be positive integers and let H be an m by n matrix with elements in Z_q. Then,

$$C = \{ \mathbf{x};\ \mathbf{x}\mathbf{H}^T = \mathbf{0} \pmod q \}$$

is a linear block code over Z_q. ◆

Definition 2. Let $\mathbf{y}$ be an n-tuple over Z_q and let $c_i(\mathbf{y})$ be the number of symbols in $\mathbf{y}$ which are equal to i or (q - i). Then,

$$\mathbf{c}(\mathbf{y}) = (c_1(\mathbf{y}), c_2(\mathbf{y}), \ldots, c_r(\mathbf{y}))$$

is the composition of $\mathbf{y}$. For odd q, $r = \frac{q-1}{2}$. For even q, $r = \frac{q}{2}$. ◆

Definition 3. Let $\mathbf{y}$ be an n-tuple over Z_q. Then,

$$w_E(\mathbf{y}) = 2\sqrt{2E_N\left(\sum_{i=1}^{r} c_i(\mathbf{y}) \sin^2\left(\frac{i\pi}{q}\right)\right)}$$

is the Euclidean weight of $\mathbf{y}$. ◆

Definition 4. Let $\mathbf{y}$ be an n-tuple over Z_q and let $|y_j|$ be the minimum of y_j and $(q - y_j)$. Then,

$$w_L(\mathbf{y}) = \sum_{j=1}^{n} |y_j| = \sum_{i=1}^{r} i c_i(\mathbf{y})$$

is the Lee weight of $\mathbf{y}$.

Upper Bound on the Minimum Euclidean Distance

Theorem 1. Consider a linear block code C over Z_q, of block length n and including M codewords. Let t be the smallest integer, $t \le n$, such that $M \sum_{i=0}^{t} \binom{n}{i} 2^i > q^n$. The minimum Euclidean distance d_{Emin}, between two codewords in C is then upper bounded by

$$d_{Emin} \le 2\sqrt{2E_N t} \sin\frac{2\pi}{q}$$ ◆

Proof of theorem 1. We first remind that for any linear code, the difference between two codewords is a codeword. The minimum Euclidean distance between two codewords then equals the minimum Euclidean weight of a codeword. Consider the following subset S of the set of n-tuples over Z_q:

$$S = \{ \mathbf{y};\ c_1(\mathbf{y}) \le t \text{ and } c_i(\mathbf{y}) = 0 \text{ for } i > 1 \}.$$

The number of n-tuples in S is $|S| = \sum_{i=0}^{t} \binom{n}{i} 2^i$. By assumption, $M|S| > q^n$. Since the code is a subgroup in Z_q^n, there must be two n-tuples, $\mathbf{y}_1$ and $\mathbf{y}_2$ in S which are in the same coset with respect to C. Then, $\mathbf{x} = \mathbf{y}_1 - \mathbf{y}_2$ is a codeword in C. We need an upper bound on $w_E(\mathbf{x})$. We have $c_i(\mathbf{y}_1) = c_i(\mathbf{y}_2) = 0$ for $i > 1$, which implies that $c_i(\mathbf{x}) = 0$ for $i > 2$. Furthermore, $w_L(\mathbf{y}_1) \le t$ and $w_L(\mathbf{y}_2) \le t$ which implies that $w_L(\mathbf{x}) \le 2t$. Thus, $c_1(\mathbf{x}) + 2c_2(\mathbf{x}) \le 2t$ and $c_i(\mathbf{x}) = 0$ for $i > 2$. Under this condition, $w_E(\mathbf{x})$ is maximum for $c_1(\mathbf{x}) = 0$ and $c_2(\mathbf{x}) = t$. The upper bound is then given by definition 3. ◆

Lower bound on the block error probability

Assume fixed q, n, and SNR. Then, each possible error pattern $\mathbf{z}_i$ has a fixed probability $P(\mathbf{z}_i)$. Assume a linear code with M equally likely codewords. Then, independently of which codeword is transmitted, the code can correct at most one error pattern in each coset. The number of cosets is $\frac{q^n}{M}$. We define V to be the set of the $\frac{q^n}{M}$ most likely error patterns. The best possible linear code then, is a code for which all error patterns in V belong to distinct cosets. Thus,

$$P_e = P(\hat{\mathbf{x}} \ne \mathbf{x}) \ge 1 - \sum_{\mathbf{z}_i \in V} P(\mathbf{z}_i)$$

Consider a list of all possible error patterns, arranged in order of decreasing probability. The list is then actually a list of composition classes. The number of error patterns in each class can easily be calculated as a multinomial coefficient multiplied by a power of two. The bound is then quite easy to calculate because there is at most one composition class which is partly inside and partly outside V.

ON DESIGNS AND FORMALLY SELF-DUAL CODES

George T. Kennedy *
National Security Agency
Fort George G. Meade, Maryland 20755 USA

Vera Pless†
Department of Mathematics
University of Illinois at Chicago
Chicago, IL 60680

Abstract

Binary formally self-dual (f.s.d) even codes are the one type of divisible $[2n, n]$ codes which need not be self-dual. On occasion a f. s. d. even $[2n, n]$ code can have a larger minimum distance than a $[2n, n]$ self-dual code. We give many examples of interesting f. s. d. even codes. We also obtain a strengthening of the Assmus-Mattson theorem. If C is a f. s. d. extremal code of length $n \equiv 2 \pmod 8$[$n \equiv 6 \pmod 8$], then the words of a fixed weight in $C \cup C^{\perp}$ hold a 3-design [1-design]. Finally, we show that the extremal f. s. d. codes of lengths 10 and 18 are unique.

Summary

A code C is formally self-dual code (f.s.d.) if C and $C^{\perp}$ have the same weight distribution. A code is *divisible* if the number of vectors of any weight is divisible by a constant δ, greater than one. The Gleason-Pierce theorem characterizes the fields over which a formally self-dual divisible code can exist and shows that δ must be either 2, 3 or 4. In all cases but one a formally self-dual divisible code is in fact self-dual. We consider this case in this paper, which is the case of binary f.s.d. codes with $\delta = 2$. We call such codes f.s.d. even codes or simply type I codes.

The classification of self-dual codes with vectors of minimal weight 2 is trivial, since a vector of weight 2 in such a code is a direct summand. The existence of f.s.d. even codes with a fixed number of vectors of weight 2 is complicated. We prove the following:

Theorem 1. 1. *If m_i are positive integers such that $\sum_{i=1}^{r} m_i = n (n > 2)$, then there exists a f.s.d. even code C of length $2n$ with $A_2 = \sum_{i=1}^{r} \binom{m_i}{2}$. Furthermore, C is equivalent to its dual.*

2. *If $\sum_{i=1}^{r} m_i = n - r/2$ where r is even, then there exists a f.s.d. even code C of length $2n$ with $A_2 = \sum_{i=1}^{r} \binom{m_i}{2}$. Again C is equivalent to its dual.*

Type I codes which are not self-dual often have a larger minimum distance for a given length than self-dual codes. In fact there are extremal f. s. d. codes where self-dual codes cannot exist. We construct some of these extremal codes and list the open cases where the existence of extremal f. s. d. even codes has not been determined. Also, we give an infinite family of f. s. d. even codes which are not equivalent to their duals, using the Turyn construction.

It is well-known that one can obtain designs from vectors of a fixed weight in an extremal self-dual code by means of the Assmus-Mattson theorem [1]. One can extend the Assmus-Mattson theorem to the words of a fixed weight in $C \cup C^{\perp}$ as follows:

Theorem 2. *Let C be a $[2n, n]$ extremal f. s. d. even code and consider the set S of vectors of a fixed weight in C and in $C^{\perp}$. Then the set S holds a 3-design whenever $2n \equiv 2 \pmod 8$ and a 1-design whenever $2n \equiv 6 \pmod 8$.*

Extremal type I codes exist at length 10 and 18. Delsarte [2] constructed a [10,5,4] f. s. d. code and exhibited the underlying 3-designs by means of inversive planes, while Assmus and Mattson [1] constructed a [18,9,6] f. s. d. code as an extended quadratic residue code. They exhibited the underlying 3-designs as a consequence of a 3-set transitive group action.

We prove the following about these codes:

Theorem 3. *Any two f.s.d. even [10,5,4] codes are equivalent.*

Theorem 4. *Any two f.s.d. even [18,9,6] codes are equivalent.*

References

[1] E. F. Assmus Jr., and H. F. Mattson, Jr., "New 5-designs," *J. Comb. Theory*, vol. 6A, pp 122-151, 1969.

[2] P. Delsarte, "Majority logic decodable codes derived from finite inversive planes," *Inform. Contr.*, vol 18, pp. 319-325, 1971.

*The author thanks the University of Illinois at Chicago for their hospitality while this work was in progress.

†This work was supported in part by NSA Grant MDA 904-91-H-0003.

GREEDY CODES

Richard A. Brualdi *
Department of Mathematics
University of Wisconsin
Madison, WI 53706

Vera Pless†
Department of Mathematics
University of Illinois at Chicago
Chicago, IL 60680

Abstract

Given an ordered basis of F_2^n and an integer d, we define a greedy algorithm for constructing a code of minimum distance at least d. We show that these greedy codes are linear and construct a parity check matrix for them. A special case of this algorithm gives the lexicodes, thereby providing a proof of their linearity which is independent of game theory. For ordered bases which have a triangular form we are able to give a lower bound on the dimension of greedy codes. Some greedy codes are better than lexicodes.

Summary

Let n and d be integers with $0 \leq d \leq n$ and suppose that the set F_2^n of binary n-tuples has been listed in some order. Choosing the first vector on the list and then apply recursively the rule:

Choose the next vector on the list whose (Hamming) distance to each previously chosen vector is at least d.

defines a binary code with minimum distance at least d. Such greedy codes were discussed in [2,3] in the case that the binary n-tuples are listed in lexicographic order.

Let $\mathcal{B}$ denote an ordered basis $y_1, y_2, \ldots, y_n$ of F_2^n. The ordered basis $\mathcal{B}$ induces an order of the vectors of F_2^n defined recursively as follows: Let $V_0 = \{(0,0,\ldots,0)\}$ and let

$$V_i = \langle y_1, \ldots, y_i \rangle \qquad (i = 1, 2, \ldots, n)$$

be the subspace of F_2^n spanned by the vectors $\{y_1, \ldots, y_i\}$. The subspace V_0 contains a unique vector and hence its vectors are ordered. Suppose the vectors in V_{i-1} have been ordered

$$x_1, x_2, \ldots, x_m \qquad (m = 2^{i-1}).$$

We have the partition

$$V_i = V_{i-1} \cup (y_i \oplus V_{i-1})$$

and we order the vectors in V_i by following the vectors $x_1, x_2, \ldots, x_m$ with the vectors $y_i \oplus x_1$, $y_i \oplus x_2, \ldots, y_i \oplus x_m$:

$$x_1, x_2, \ldots, x_m, y_i \oplus x_1, y_i \oplus x_2, \ldots, y_i \oplus x_m.$$

Since $V_n = F_2^n$, this defines an order for the vectors of F_2^n which we call the *order induced by* $\mathcal{B}$ or, for short, the *$\mathcal{B}$-order* of F_2^n.

Let $\mathcal{B}$ be an ordered basis of F_2^n and let d be an integer with $0 \leq d \leq n$. Applying the greedy algorithm (for the chosen d) to the $\mathcal{B}$-order of F_2^n we obtain a code $C = C(\mathcal{B}, \Gamma)$ whose minimum distance is at least d. The code C is the *$\mathcal{B}$-greedy code of length n and designed distance d.* The lexicodes are a special case of $\mathcal{B}$-greedy codes.

Our main result is that $\mathcal{B}$-greedy codes are always linear and we show how to enhance the greedy algorithm in order to determine a parity check matrix of the code. We also show that it suffices to consider only $\mathcal{B}$-greedy codes of even designed distance. The $\mathcal{B}$-greedy codes for which $\mathcal{B}$ is a triangular ordered basis are called *triangular-greedy* codes.

We present computer data which shows that these codes have dimension within one of the best codes known [4].

References

[1] R. A. Brualdi and V. S. Pless, Greedy codes, to appear in JCT (A).

[2] J. H. Conway and N. J.A. Sloane, Lexicographic codes: Error correcting codes from game theory, *IEEE Trans. Inform. Theory*, vol. IT-32, 1986, 337-348.

[3] V. I. Levenstein, A class of systematic codes, *Soviet Math. Dokl.*, 1:1, 1960, 368-371.

[4] T. Verhoeff, An updated table of minimum-distance bounds for binary linear codes, *IEEE Trans. Inform. Theory*, vol. IT-33, 1987, 665-680.

*Research partially supported by NSF Grant DMS-8901445 and NSA Grant MDA904-89-H-2060.

†Research partially supported by NSA Grant MDA 904-91-H-0003.

ON THE UPPER BOUND OF THE SIZE OF THE r-COVER-FREE FAMILIES

MIKLÓS RUSZINKÓ

RESEARCH GROUP for INFORMATICS and ELECTR.
HUNGARIAN ACADEMY of SCIENCES and
MATHEMATICAL INSTITUTE of the HUNG. AC. of SC.
BUDAPEST, P.O.B. 127, 1364 HUNGARY

Abstract

The notion of the r-cover-free families was introduced by Kautz and Singleton in 1964 [17]. They initiated investigating binary codes with the property that the disjunction of any $\leq r$ ($r \geq 2$) codewords are distinct (UD_r codes). This led them to studying the binary codes with the property that none of the codewords is covered by the disjunction of $\leq r$ others (*Superimposed codes, ZFD_r codes;* P. Erdős, P. Frankl and Z. Füredi called the correspondig set system r-cover-free in [7]).

Since that many results have been proved about the maximum size of these codes. Various authors studied these problems basically from three different points of view, and these three lines of investigations were almost independent of each other. This is why many results were found first in information theory ([1], [4], [5], [14], [15], [16], [17]), were later rediscovered in combinatorics ([2], [6], [7], [10]), or in group testing ([12], [13]), and vice versa.

We shall approach this area from the combinatorial side. Our main goal is to estimate the maximal size of the family of subsets of an n-element set with the property that no set is covered by the union of r others.

Summary

Let S be an n-element set. 2^S is the set of all subsets of S. $\binom{S}{k}$ denotes the set of all k-subsets of S ($k \geq 0$). If $|S| = n$, then $|\binom{S}{k}| = \binom{n}{k}$. We denote by $[n]$ the set $\{1, 2, \ldots, n\}$, and $\log x$ is always of base 2. A set system $\mathcal{A} \subseteq 2^S$ is called k-uniform if its members are k- sets. It is usually supposed that the underlying set of the set systems is $[n]$.

We call $\mathcal{F}' \subset 2^S$ r-distinct, if $\bigcup_{i=1}^{k} A_i \neq \bigcup_{j=1}^{\ell} B_i$ for any $\{A_1, A_2, \ldots, A_k\} \neq \{B_1, B_2, \ldots, B_\ell\}$, $1 \leq k, \ell \leq r$; $A_1, A_2, \ldots, A_k, B_1, B_2, \ldots, B_\ell \in \mathcal{F}'$. $\mathcal{F} \subset 2^S$ is r-cover-free, if $A_0 \not\subseteq A_1 \cup A_2 \cup \ldots \cup A_r$ holds for all distinct $A_0, A_1, \ldots, A_r \in \mathcal{F}$. $\mathcal{F}^* \subset 2^S$ is $< r$ part intersecting, if $|A_i \cap A_j| < \frac{1}{r} \min\{|A_i|, |A_j|\}$ for any distinct $A_i, A_j \in \mathcal{F}^*$ holds. We denote by $T'(r,n)$, $T(r,n)$, $T^*(r,n)$ and $T'(r,n,k)$, $T(r,n,k)$, $T^*(r,n,k)$ the maximum cardinality of the corresponding set systems in general and in k-uniform case, resp. We will provide upper bounds on these functions for r fixed and n tending to infinity.

The following upper and lower bounds were proved in [1], [4], [5], [7], [13]: there exist two (absolute) constants c_1, c_2 such that

$$\frac{c_1}{r^2} \leq \frac{\log T(r,n)}{n} \leq \frac{c_2}{r} \qquad (1)$$

for any n. In most papers the lower bound is proved by probabilistic methods. In [13] V.T. Sós and F.K. Hwang used a greedy-type algorithm to generate $< r$ part intersecting families for proving the lower bound. The upper bound was proved using the observation that, by definition, $\sum_{i=1}^{r} \binom{T'}{i} \leq 2^n$. The gap between the upper and lower bounds is rather large. Dyachkov and Rykov obtained a better upper bound [4]:

$$\frac{\log T(r,n)}{n} \leq c_3 \frac{\log r}{r^2} \qquad (2)$$

for some absolute constant c_3 and any n. Their proof is rather involved. Here we shall give a simple and purely combinatorial proof of this result.

References

[1] Nguyen Quang A and T. Zeisel, Bounds on constant weight binary superimposed codes, *Probl. of Control and Information Theory* 17 (1988), 223-230.

[2] N. Alon, Explicit constractions of exponential sized families of k-independent sets, *Discrete Mathematics* 58 (1986), 191-193.

[3] Zs. Baranyai, On the factorization of the complete uniform hypergraph, *Proc. Colloq. Math. Soc. János Bólyai* (10. Infinite and finite sets, Keszthely, Hungary (1973).

[4] A. G. Dyachkov and V.V. Rykov, Bounds on the length of disjunctive codes, *Problemy Peredachi Informatsii,* Vol. 18, No 3 (1982), 7-13.

[5] A. G. Dyachkov and V.V. Rykov, A survey of superimposed codes theory, *Probl. of Control and Information Theory,* Vol. 12, No 4 (1983), 1-13.

[6] P. Erdős, P. Frankl and Z. Füredi, Families of finite sets in which no set is covered by the union of two others, *Journal of Combinatorial Theory, Series A* Vol. 33, No. 2 (1982), 158-166.

[7] P. Erdős, P. Frankl and Z, Füredi, Families of finite sets in which no set is covered by the union of r others, *Israel J. of Math.* Vol. 51. Nos. 1-2 (1985), 79-89.

[8] P. Frankl, On Sperner Families Satisfying an Additional Condition,*Journal of Combinatorial Theory, Series A* Vol. 20, No. 1 (1976), 1-11

[9] P. Frankl and Z. Füredi, Colored packing of sets, *Annals of Discrete Mathematics,* Vol. 34, (1987), 165-178

[10] P. Frankl and V. Rödl, Near perfect coverings in graphs and hypergraphs, *Europ. J. Combinatorics* 6 (1985), 317-326.

[11] R. G. Gallager, Information Theory and Reliable Communication, Wiley (1968), problem 5.8

[12] F. K. Hwang, A method for detecting all defective members in a population by group testing, *J. of the American Statistical Association,* Vol. 67, No 339 (1972), 605-608.

[13] F. K. Hwang and V.T. Sós, Non adaptive hypergeometric group testing, *Studia Sc. Math. Hungarica,* 22 (1987), 257-263.

[14] S. M. Johnson, On the upper bounds for unrestricted binary error-correcting codes, *IEEE Trans. on Inf. Th.,* Vol. it- 17, No. 4 (1971), 466-478.

[15] S. M. Johnson, Improved asymptotic bounds for error-correcting codes, *IEEE Trans. on Inf. Th.,* Vol. it-9, No 4 (1963) 198-205.

[16] S. M. Johnson, A new upper bound for error-correcting codes, *IRE Trans. on Inf. Th.,* Vol. it-8 (1962), 203-207.

[17] W. H. Kautz and R.C. Singleton, Nonrandom binary superimposed codes, *IEEE Trans. on Inf. Th.,* Vol. it-10 (October 1964), 363-377.

Packing Radius
vs
Covering Radius

Patrick Solé,
Philip Stokes,
CNRS, I3S ,
250, rue A. Einstein,
06 560 Valbonne, France

Key words:Binary Codes, Covering Radius, Packing Radius, Asymptotic Bounds

1 Introduction

Let C_i, $i = 1, 2, \ldots$ denote an infinite family of binary codes with length n_i, covering radius r_i, minimum distance d_i. Assume that the limit ρ *(resp.* δ*) of the ratio* $\frac{r_i}{n_i}$ (resp. $\frac{d_i}{n_i}$) for large i exist and call it normalized covering radius (resp. distance). Our aim is to study the set Y_2 (resp. Y_2^{lin}) of points (ρ, δ) of the unit square achieved by binary families of codes (resp. of linear codes). We address the following questions for both domains

1. bounds on the extreme points
2. convexity
3. continuity at the border.

Both sets split naturally into four subdomains according to the position of ρ and δ w.r.t. $\frac{1}{2}$.

2 Convexity

Question 2 is still unsolved. A weaker result is the following

Theorem 1 *Let* $(x, y) \in Y_2^{lin}$ *(resp.* $\in Y_2$*). Then every point of the line between* (x, y) *and* $(1, 0)$ *lies in* Y_2^{lin} *(resp.* Y_2*).*

3 Upper Left Corner

$(0.5 \leq \delta \leq 1,\ 0 \leq \rho < 0.5)$) The Plotkin bound shows that this corner is empty.

4 Lower Left Corner

$(0 \leq \delta \leq 0.5,\ 0 \leq \rho < 0.5)$) Let $B(\delta)$ denote any bound on the rate as a function of the distance. Eliminating the rate by the sphere covering bound yields

$$\rho \geq H^{-1}(1 - B(\delta)),$$

where H is the entropy function. Taking B to be the Elias Bound yields the following weak but elegant result

$$\delta \leq 2\rho(1 - \rho).$$

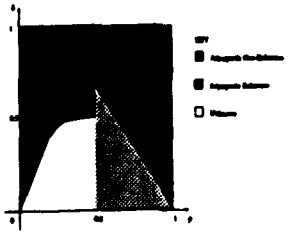


Figure 1: Linear Codes

The MR2W (bound of four) bound yields a more precise but less explicit result. Nonconstructive results [1] of Cohen and Frankl entail that the intersection of the first axes' bissector with this corner lies entirely in Y_2^{lin}. Theorem 1 shows then that the whole triangle $\{(0,0),(0.5,0.5),(1,0)\}$ is in Y_2^{lin}. Manin [3] showed the existence of a continuous function $\alpha_2(\delta)$ which is the "true upper bound" on the rate. Analogously define $\beta_2(\delta)$ as the "true lower bound" on ρ. We do not know if β_2 is continuous in this corner. All that is known is $\delta \geq \beta_2(\delta) \geq H^{-1}(1 - \alpha_2(\delta))$.

5 Right Half-Square

This means $\rho \geq 0.5$.

5.1 Linear Codes

A simple construction shows that the line segment $\{\rho = 0.5, 0 \leq \delta \leq \frac{2}{3}\}$ lies entirely in Y_2^{lin}. By Theorem 1 the triangle spanned by this segment and the point $(1, 0)$ lies entirely in Y_2^{lin}. Now recall the Janwa's bound [2]

$$r_i \leq n_i - \sum_{i=1}^{k} \lceil \frac{d_i}{n_i} \rceil.$$

Families of linear codes with $k_i = 1$ (for i large enough) lie on $\rho = 1 - \frac{\delta}{2}$. Families of linear codes with $k_i \geq 2$ (for i large enough) lie under $\rho = 1 - \frac{3\delta}{4}$, which is a side of the preceding triangle. This settle the three questions for this corner in the linear case.

5.2 Unrestricted Codes

It is easy to see that $\rho \leq 1 - \frac{\delta}{2}$ is valid for all (families of) codes with at least two words. The Plotkin bound yields $\delta \leq \frac{2}{3}$ for families of codes with at least three words. A careful study of 3-word codes based on the Sloane-Mattson [5, 4] linear programming methodology shows that there are such codes on the line $\rho = \delta$. So Y_2 consists of Y_2^{lin}, plus the triangle with sides the bissector and the first two Janwa bounds. The status of the triangle $(1/2, 2/3)(4/7, 4/3)(2/3, 2/3)$ is still unresolved.

6 Acknowledgement

We thank G.D. Cohen, S. Litsyn, H-F. Mattson, jr for helpful discussions.

References

[1] G.D. Cohen, P. Frankl,"Good Coverings of Hamming Spaces with Spheres", Discr. Math 56 (1985) 125-131.

[2] H. Janwa, "Some New Upper Bounds on the Covering Radius of Binary Linear Codes" IEEE Trans. on Information Th., IT-35 (1989) 110-122.

[3] Y. Manin, "What is the maximum number of points of a curve over F_2?" J. of the Fac. of Sc. Univ. Tokyo 26 (1981) 715-720.

[4] H.F. Mattson jr. "Simplifications to ..." J. Comb. Th. ser. A 57,2 (1991) 311-315.

[5] N.J.A. Sloane "A New Approach to the Covering Radius of Codes", J. Comb. Th. ser. A 46 (1986) 61-86.

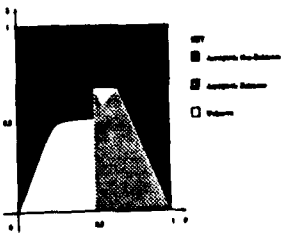


Figure 2: Unrestricted Codes

Constructive Non-Existence Proofs for Linear Covering Codes

René Struik
Eindhoven University of Technology
Dept. of Mathematics and Computing Science
P.O. Box 513, 5600 MB Eindhoven, the Netherlands

1 Introduction

A code $C \subseteq \mathbb{F}_2^n$ *has covering radius* (at most) r, if $d(x, C) \leq r$ for all $x \in \mathbb{F}_2^n$. For *linear* codes, the covering radius is the highest weight of any coset leader of the code. A basic question concerning the covering radius of codes is to determine $K(n, r)$, the minimum cardinality of any block-code of length n and covering radius r. For linear codes this question amounts to determining $l(m, r)$, the minimum length of a linear code with codimension m and covering radius r. We show how techniques from coding theory can be successfully applied to improve bounds on $l(m, r)$ found in the literature [3-7].

2 Preliminaries

Let C be a code of length n with covering radius r.
A trivial lowerbound for the size of a covering code is given by the *Sphere-Covering Bound*

$$|C| \sum_{i=0}^{r} \binom{n}{i} \geq 2^n \tag{1}$$

The Van Wee bound [1] improves on this bound, whenever $(r+1) \mid (n+1)$:

$$|C| \left\{ \sum_{i=0}^{r} \binom{n}{i} - \frac{\binom{n}{r}}{\lceil \frac{n-r}{r+1} \rceil} \left(\left\lceil \frac{n+1}{r+1} \right\rceil - \frac{n+1}{r+1} \right) \right\} \geq 2^n \tag{2}$$

As a direct consequence we obtain

$$\text{If } n \text{ is } \textit{even}, \text{ then } K(n, 1) \geq 2^n/n \tag{3}$$

Blokhuis and Lam [2] showed that arbitrary coverings and sphere-coverings can be linked.

Definition 1 Let $S \subset \mathbb{F}_2^k$ and let A be a $k \times n$ matrix. S is said to r-cover $\mathbb{F}_2^k$ using matrix A, if $\{ s + w\, A^T \mid s \in S \text{ and } wt(w) \leq r \} = \mathbb{F}_2^k$. □

Lemma 2 If S r-covers $\mathbb{F}_2^k$ using matrix A, then the set $C := \{ w \in \mathbb{F}_2^n \mid w\, A^T \in S \} \subseteq \mathbb{F}_2^n$ has covering radius r. In particular, $K(n, r) \leq |S|\, 2^{n-k}$. □

Linear codes can be slightly modified without changing the covering radius, as is demonstrated by the following trivial 'inversion property.' For codes with even covering radius this property was already mentioned in [4].

Lemma 3 The linear codes with parity check matrices $H = \left(\begin{array}{c|c} 1 & u \\ \hline 0 & X \end{array} \right)$, resp. $H' = \left(\begin{array}{c|c} 1 & u \oplus 1 \\ \hline 0 & X \end{array} \right)$, have the same covering radius. □

Lemma 4 If C is an $[n, k, d]$-one weight code without zero-positions, then $d(2^k - 1) = n 2^{k-1}$. In particular $2^{k-1} \mid d$. □

3 New Bounds for Linear Covering Codes

A linear covering code imposes restrictions on the form of its dual code. This observation enables us to transform the problem of designing a 'good' linear covering code into the problem of designing a (dual) linear code with a lot of structure imposed onto it. Techniques from coding theory might show that such a dual code can not exist. We demonstrate the main idea for covering radius two.

Lemma 5 Let C be an $[n, n-m]$-code with covering radius two. Then the weights $w \neq 0$ in $C^\perp$ satisfy the following properties:

1. $w\,(n+1-w) \geq 2^{m-1}$
2. $w\, 2^{(n-w)-(m-1)} \geq K(n-w, 1)$
3. if weight w can not occur, then weight $(n+1)-w$ can not occur either

PROOF Suppose code $C^\perp$ has a codeword of weight $w \neq 0$. We can put the parity-check matrix H for code C into the following form:

$$H = \left(\begin{array}{c|c} A_0 & A_1 \\ \hline 0 \;\cdots\; 0 & 1 \;\cdots\; 1 \end{array} \right) \updownarrow m$$

$$\underbrace{\qquad\qquad}_{n-w} \quad \underbrace{\qquad\qquad}_{w}$$

1. All syndromes of the form $(*, 1)$ should be the sum of one or two columns of matrix H, hence $w + w\,(n-w) = w\,(n+1-w) \geq 2^{m-1}$
2. The columns of matrix A_1 1-cover $\mathbb{F}_2^{m-1}$ using matrix A_0. Now application of Lemma 2 proves the statement.
3. Via elementary row operations on matrix H we obtain a zero-column in matrix A_1. Now we can apply the 'inversion property' (Lemma 3). □

Remark 6 Notice that Property 1 is weaker than Property 2, since Property 2 together with the sphere-covering bound implies Property 1. Often significantly better bounds for $K(n, 1)$ are known. □

As an application of Lemma 5 we prove the bound $l(2m-1, 2) \geq 2^m + 1$ for $m \geq 3$, improving by one the minimum value of $l(2m-1, 2)$ implied by the Van Wee bound. This bound was conjectured by Brualdi, Pless and Wilson [3], but up to now only the case $m = 6$ has been settled [4,5]. The proof is surprisingly simple.

Example $l(2m-1, 2) \geq 2^m + 1$ for all $m \geq 3$

PROOF Suppose C is an $[n = 2^m, 2^m - (2m-1)]$-code with covering radius two. We infer from Property 1 of Lemma 5 that, for $m \geq 3$, code $C^\perp$ does not contain the all-one vector. If a codeword of weight $w \neq 0$ occurs in $C^\perp$, then we have $K(v, 1) \leq w\, 2^{v-(2m-2)}$ with $v + w = n$, according to Property two of Lemma 5.
For *even* v we have the lowerbound $K(v, 1) \geq 2^v/v$, cf. equation (3). Thus we obtain the inequality $v\, w \geq \frac{1}{4} n^2$ for even v. Since $v + w = n$ we have in fact equality and $w = n/2 = 2^{m-1}$. We infer that the even weight subcode of $C^\perp$ of dimension $k \geq 2m-2$ is in fact a one-weight code with $d = 2^{m-1}$, hence it satisfies the divisibility constraint $2^{k-1} | d$ (cf. Lemma 4). However, for $m \geq 3$ this divisibility constraint is not satisfied.
Hence $l(9, 2) \geq 33$, $l(11, 2) \geq 65$, $l(13, 2) \geq 129$, etc. □

Remark 7 In a similar way we can prove the next bounds: $l(16, 2) \geq 363$, $l(18, 2) \geq 725$, $l(20, 2) \geq 1449$, $l(22, 2) > 2897$. □

Our approach can be extended into several directions, enabling us to prove the bounds $l(6, 2) = 13$, $l(7, 2) = 19$, $l(8, 2) \geq 25$, $l(9, 2) \geq 34$, $l(8, 3) = 14$, $l(9, 3) \geq 17$, $l(10, 3) \geq 21$, $l(12, 3) \geq 31$, $l(13, 3) \geq 38$.

References

[1] G.J.M. van Wee, "Improved Sphere Bounds on the Covering Radius of Codes," *IEEE Trans. Inform. Theory*, Vol. IT-34, pp. 237-245, March 1988.

[2] A. Blokhuis, C.W.H. Lam, "More Coverings by Rook Domains," *Journal of Combinatorial Theory*, A 36, pp. 240-244, 1984.

[3] R.A. Brualdi, V.S. Pless, R.M. Wilson, "Short Codes with a Given Covering Radius," *IEEE Trans. Inform. Theory*, Vol. IT-35, pp. 99-109, January 1989.

[4] R.A. Brualdi, V.S. Pless, "On the Length of Codes with a Given Covering Radius," in *Coding Theory and Design Theory, Part 1*, D.R. Chaudhuri, Ed., New York: Springer-Verlag, 1990, pp. 9-15.

[5] O. Ytrehus, "Binary [18,11]2 Codes Do Not Exist-Nor do [64,53]2 Codes," *IEEE Trans. Inform. Theory*, Vol. IT-37, pp. 349-351, March 1991.

[6] A.R. Calderbank, N.J.A. Sloane, "Inequalities for Covering Codes," *IEEE Trans. Inform. Theory*, Vol. IT-34, pp. 1276-1280, September 1988.

[7] J. Simonis, "The Minimal Covering Radius t[15,6] of a Six-Dimensional Binary Linear Code of Length 15 is Equal to 4," *IEEE Trans. Inform. Theory*, Vol. IT-34, pp. 1344-1345, September 1988.

Perfect Tilings of Binary Spaces

Gerard Cohen
ENST, Département Informatique
46 rue Barrault, Paris, France

Simon Litsyn
Tel-Aviv University, EE Dept.
Tel-Aviv 69978, Israel

Alexander Vardy
IBM Almaden Research Center
650 Harry Road, San Jose, CA

Gilles Zémor
ENST, Département Réseaux
46 rue Barrault, Paris, France

Abstract. We study partitions of the space $\mathbb{F}_2^n$ of all the binary n-tuples into disjoint sets, such that each set is an additive coset of a given set V. Such a partition is called a perfect tiling of $\mathbb{F}_2^n$ and denoted (V, A), where A is the set of coset representatives. A sufficient condition for a set V to be a tile is given in terms of the cardinality of $V+V$. A perfect tiling (V, A) is said to be proper if V generates $\mathbb{F}_2^n$. We show that the classification of perfect tilings can be reduced to the study of proper perfect tilings. We then prove that each proper perfect tiling is uniquely associated with a perfect binary code. A construction of proper perfect tilings from perfect binary codes is presented. Furthermore, we introduce a class of perfect tilings obtained by iterating a simple recursive construction. Finally, we generalize the well-known Lloyd theorem, originally stated for tilings by spheres, for the case of arbitrary perfect tilings.

Given a body in the n-dimensional Euclidean space, is it possible to tile the space with exact copies of this body? This problem has been extensively studied in the classical literature, see [6] and references therein. We study here the binary version of this problem. Let $\mathbb{F}_2^n$ denote the n-dimensional Hamming space, i.e. the set of all binary n-tuples with addition term by term modulo 2. A given set (body) V tiles $\mathbb{F}_2^n$ if it is possible to perfectly cover $\mathbb{F}_2^n$ with disjoint additive cosets of V. Note that the set of coset representatives A is also a tile of $\mathbb{F}_2^n$. Without loss of generality we assume that both V and A contain the **0** element. Evidently, each element x of $\mathbb{F}_2^n$ must have a unique representation of the form $x = v + a$, where $v \in V$ and $a \in A$. Thus we have the following definition of a perfect tiling: (V, A) is a perfect tiling of $\mathbb{F}_2^n$ if $V + A = \mathbb{F}_2^n$ and $(V+V) \cap (A+A) = \{\mathbf{0}\}$.

If both V and A are groups, (V, A) is a perfect tiling of $\mathbb{F}_2^n$ iff $A = \mathbb{F}_2^n / V$. Hence in the sequel we consider only nonlinear tilings where at least one of the sets V, A is not a group. A well-known example of a nonlinear tile is a sphere, in which case the set of coset representatives is a perfect binary code. Tilings by generalized spheres have been studied in [2] and [3]. The following proposition shows that many more perfect tilings exist.

Proposition 1. *If $|V+V| < 2|V|$ there exists a group A, such that (V, A) is a perfect tiling.*

In particular, since $|V| \le |V+V| \le \binom{|V|}{2} + 1$, any set of cardinality 4 is a tile by Proposition 1. If $|V+V|$ is large it is sometimes possible to show that no tiling is possible. Certain bounds on the cardinality of $V+V$ for a given set V may be found in [7].

We shall say that (V, A) is a *proper* perfect tiling of $\mathbb{F}_2^n$ if (V, A) is a perfect tiling of $\mathbb{F}_2^n$ and V generates $\mathbb{F}_2^n$, i.e. $\langle V \rangle = \mathbb{F}_2^n$, where $\langle V \rangle$ denotes the span of V. We now prove that the classification of perfect tilings can be reduced to the study of proper perfect tilings.

Proposition 2. *A set V is a tile of $\mathbb{F}_2^n$ iff it is a tile of $\langle V \rangle$. Furthermore all the sets A, such that (V, A) perfectly tiles $\mathbb{F}_2^n$, can be constructed as follows. Denote $m = 2^{n-r} - 1$, where r is the rank of V.*

1. *Let $A_0, A_1, \ldots A_m$ be some $m+1$, not necessarily distinct, subsets of $\langle V \rangle$ such that for all i, $0 \le i \le m$, (V, A_i) is a proper perfect tiling of $\langle V \rangle$.*
2. *Let $c_0 = \mathbf{0}, c_1, \ldots c_m$ be a set of representatives of $\mathbb{F}_2^n / \langle V \rangle$.*
3. *For $1 \le i \le m$, let v_i be any element of $\langle V \rangle$.*

Then $A = A_0 \cup (v_1 + c_1 + A_1) \cup \ldots \cup (v_m + c_m + A_m)$.

The foregoing proposition shows that all the perfect tilings may be constructed from proper perfect tilings. Therefore, we shall henceforth assume that $n = \mathrm{rank}(V)$, and identify $\mathbb{F}_2^n$ with $\langle V \rangle$. With an appropriate choice of basis for $\mathbb{F}_2^n$, it may be further assumed that $V \supset B_n(\mathbf{0}, 1)$, where $B_n(\mathbf{0}, 1)$ is a Hamming sphere of radius 1 in $\mathbb{F}_2^n$. Some of the facts which we were able to demonstrate for proper perfect tilings are listed below.

1. If $|V+V| = 2|V|$, V is a tile iff $(V+V)$ is not a group.
2. Obviously, if $n = |V| - 1$ then $V = B_n(\mathbf{0}, 1)$ is a tile, and A is a perfect binary code. If $n = |V| - 2$ then V is also a tile, and A is a shortened Hamming code.
3. If $|V| \le 8$ and (V, A) is a proper perfect tiling, then A is a group.

Let (V, A) be a perfect tiling, and let $H(V)$ be an $n \times (|V|-1)$ matrix having the elements of $V \backslash \{\mathbf{0}\}$ as its columns. For $x \in \mathbb{F}_2^{|V|-1}$, set $s(x) = H(V)x^t$ and define the codes C and C_0 as follows:

$$C = \{c \in \mathbb{F}_2^{|V|-1} : s(c) \in A\} \qquad C_0 = \{c \in \mathbb{F}_2^{|V|-1} : s(c) = \mathbf{0}\}$$

Proposition 3. *The code C is a perfect binary code with minimum distance 3, and C_0 is a linear subcode of C. Furthermore, if (V, A) is a proper perfect tiling then $|C| = |C_0| \cdot |A|$, $\mathrm{rank}(C) = |V| + \mathrm{rank}(A) - \mathrm{rank}(V) - 1$, and C is linear iff A is a group.*

Proposition 3 shows that each perfect tiling is uniquely associated with a perfect binary code, and provides a means for constructing perfect codes from perfect tilings. The converse construction is also possible.

Proposition 4 (Converse of Proposition 3). *Let C be a perfect binary code of length n with minimum distance 3. Let Γ be a linear code of dimension γ, such that $C + \Gamma = C$. Let $H(\Gamma)$ be a parity check matrix of Γ. If $V \backslash \{\mathbf{0}\}$ is the set of rows of $H(\Gamma)^t$ and $A = \{H(\Gamma)c^t : c \in C\}$, then (V, A) is a proper perfect tiling of $\mathbb{F}_2^{n-\gamma}$.*

Note that γ is possibly 0, in which case V is a sphere and $A = C$. If $\gamma \neq 0$, many non-equivalent perfect tilings may be constructed from the same perfect binary code. In this case the construction of Proposition 4 is not explicit, as there is no obvious way to find the code Γ. Several explicit constructions of proper perfect tilings from perfect binary codes will be presented elsewhere.

The correspondence between sets V, A such that $V + A = \mathbb{F}_2^n$ and coverings by spheres of radius 1 has been initially noticed in [1]. The relevance of their rank, however, seems to have been overlooked. We will hereafter elaborate on this issue. First we show that many proper perfect tilings have a recursive structure analogous to the structure of perfect tilings exhibited in Proposition 2. Suppose that (V, A) is a proper perfect tiling of $\mathbb{F}_2^n$, with $\mathrm{rank}(A) < \mathrm{rank}(V) = n$. Then (A, V) is a perfect tiling of $\mathbb{F}_2^n$. Applying Proposition 2 to (A, V) yields

$$V = V_0 \cup (a_1 + c_1 + V_1) \cup \ldots \cup (a_m + c_m + V_m),$$

where for $i = 0, 1, \ldots m$, (A, V_i) is a proper perfect tiling of $\langle A \rangle$, $c_0, c_1, \ldots c_m$ are the representatives of $\mathbb{F}_2^n / \langle A \rangle$, and $a_i \in \langle A \rangle$. The same argument can now be applied to each of the tilings (A, V_i), provided that $\mathrm{rank}(V_i) < \mathrm{rank}(A)$ for all $i = 0, 1, \ldots m$. This defines a class of tilings obtained by recursively iterating the construction of Proposition 2. The recursion terminates only if a proper perfect tiling with $\mathrm{rank}(A) = \mathrm{rank}(V)$ is encountered. Such a tiling is said to be of *full rank*. Full-rank perfect tilings have been constructed by Etzion and Vardy in [4]. In view of the foregoing discussion they may be considered as the "building blocks" of all the perfect tilings.

For demonstrating the non-existence of certain tilings the following generalization of the Lloyd theorem may be useful. Let $\chi_u(V) = \sum_{v \in V} (-1)^{\langle u, v \rangle}$ be a character of the group algebra QF_n (cf. [5], chap. 5). For a perfect tiling (V, A), define the sets $U, N(U), A'$ and $N(A')$ as follows:

$$U = \{u : \chi_u(V) = 0\} \qquad N(U) = \{j : \exists u \in U \text{ with } \mathrm{wt}(u) = j\},$$

$$A' = \{a' : \chi_{a'}(A) \neq 0\} \qquad N(A') = \{j : \exists a' \text{ with } \mathrm{wt}(a') = j\}.$$

Note that the set A' may be regarded as the code formally dual to A.

Proposition 5. *In the above notation $N(A') \subseteq N(U)$ and $|U| \ge |V|$.*

References

[1] A. Blokhuis and C.W.H. Lam, "More coverings by rook domains," *J. Combin. Theory A*, vol.36, pp. 240–244, 1984.

[2] G. Cohen and P. Frankl, "On tilings of the binary vector space," *Discrete Math.*, vol.31, pp. 271–277, 1980.

[3] M. Deza, "The effectiveness of noise correction or detection," *Problems of Inf. Trans.*, vol.1 (3), pp. 29–39, 1965.

[4] T. Etzion and A. Vardy, "Perfect binary codes: constructions, properties, and enumeration," *IEEE Trans. Inform. Theory*, submitted.

[5] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, New York: North-Holland, 1977.

[6] C.A. Rogers, *Packing and Covering*, Cambridge University Press, 1964.

[7] G. Zémor, "Subset sums in binary spaces," *Europ. J. Combin.*, vol.13, pp. 221–230, 1992.

THE UNIQUENESS OF THE (9, 18, 4) CONSTANT-WEIGHT-4 CODE

by H. F. Mattson, jr.
CIS, 4-116 CST
Syracuse University
Syracuse, NY 13244-4100

We prove that the code of the title is unique by showing it can be extended to a constant-weight-4 code of type (10,30,4). The uniqueness of the latter code was proved by Witt in 1938 (in the language of Steiner systems). Acknowledgement. The author is grateful for support from ENST, Paris, and INRIA, Rocquencourt, where this work was done.

OPTIMIZATION OF TRANSMITTER PULSES FOR TWO-USER DATA COMMUNICATIONS

Michael L. Honig and Upamanyu Madhow
Bellcore
445 South St.
Morristown, New Jersey 07960

SUMMARY

Two-user data communications is considered in which the users each transmit Pulse-Amplitude Modulated data signals through linear, time-invariant channels with transfer functions $H_1(f)$ and $H_2(f)$, respectively. The received signal is the sum of the outputs of these channels plus white Gaussian noise. Assuming that the symbol rate is the same for each user, and that the users are not allowed to coordinate their transmissions on a per symbol basis, we study the problem of optimizing their transmitted pulse shapes.

Two types of receivers are considered: the matched filter detector, which attempts to demodulate each user independently, while treating interference from the other user as wide-sense stationary noise, and the Minimum Mean Squared Error (MMSE) linear detector, which jointly demodulates both users simultaneously. For each case necessary conditions are derived for the transmitted pulse shapes that minimize the Mean Squared Error (MSE), subject to an average power constraint, and conditions are given for which the corresponding solution is unique. Our results generalize those in [1], in which the MMSE linear transmitter and receiver filters for a single-user channel are derived.

User i, $i \in \{1, 2\}$, generates a sequence of pulses

$$s_i(t) = \sum_k b_k^{(i)} \delta(t - kT) \tag{1}$$

where $\{b_k^{(i)}\}$ is the sequence of transmitted data symbols from user i, and $1/T$ is the symbol rate, which is assumed to be the same for both users. The received signal is then

$$y(t) = h_1 * p_1 * s_1(t) + h_2 * p_1 * s_2(t) + n(t) \tag{2}$$

where $p_1(t)$ and $p_2(t)$ are the pulse shapes for each user, $h_1(t)$ and $h_2(t)$ are the impulse response functions associated with $H_1(f)$ and $H_2(f)$, respectively, "$*$" denotes convolution, and $n(t)$ is white Gaussian noise with spectral density σ_n^2.

For the matched filter detector the output of $H_i(f)$ is the input to the filter with transfer function $P_i^*(f)H_i^*(f)$, where $P_i(f)$ is the Fourier Transform of $p_i(t)$. The output of this filter is sampled at rate $1/T$, which produces the estimated sequence of symbols $\{\hat{b}_k^{(i)}\}$. The MSE for the matched filter receiver is

$$\sum_{i=1}^{2} E[(b_k^{(i)} - \hat{b}_k^{(i)})^2] = T\sigma_b^2 \sum_{i=1}^{2} \int_{-1/(2T)}^{1/(2T)} \left\{ (\|Q_i\|^2 - 1)^2 + |Q_1^* Q_2|^2 + \xi \|Q_i\|^2 \right\} df \tag{3}$$

where $\sigma_b^2 = E[(b_k^{(i)})^2]$, $\xi = \sigma_n^2/\sigma_b^2$, and

$$[Q_i(f)]_k = P_i[f - (k-1-K)/T]\, H_i[f - (k-1-K)/T], \tag{4}$$

$i = 1, 2$, $k = 1, \ldots, 2K+1$. The three terms in the integrand can be classified as MSE due to ISI, multiple-access interference, and noise.

The MMSE linear detector in general offers a significant performance improvement relative to the matched filter detector. The MMSE linear detector for this multi-user channel consists of matched filters followed by symbol-rate samplers, and a 2-input/2-output digital filter. The MSE in this case is given by [2]

$$\text{MMSE} = T\sigma_n^2 \int_{-1/(2T)}^{1/(2T)} \frac{2\xi + \|Q_1\|^2 + \|Q_2\|^2}{(\xi + \|Q_1\|^2)(\xi + \|Q_2\|^2) - |Q_1^* Q_2|^2}\, df. \tag{5}$$

We now wish to find transmitter pulses, specified by $P_1(f)$ and $P_2(f)$, to minimize the expressions for MSE given by (3) and (5) subject to the average power constraints

$$T \int_{-1/(2T)}^{1/(2T)} \left(\sum_{k=-K}^{K} |P_i(f - k/T)|^2 \right) df \le \Pi_i, \quad i = 1, 2. \tag{6}$$

To specify the solution to these optimization problems we need the following notation. For every $f \in [-1/(2T), 1/(2T)]$ define $\bar{k}_i(f)$ to be any integer for which $|H_i(f - \bar{k}_i/T)| \ge |H_i(f - k/T)|$ for all integers k. Provided that $|H_i(f)|$, $i = 1, 2$, satisfy some relatively weak conditions (i.e., $|H_1(f)|$ cannot be a constant times $|H_2(f)|$ on a set of positive measure), then the MMSE transmitter filter for the MMSE linear detector is given by

$$|P_i(f - \bar{k}_i/T)|^2 = \frac{1}{|H_i(f - \bar{k}_i/T)|^2} \left(\frac{|H_i(f - \bar{k}_i/T)|}{\sqrt{\lambda_i}} - \xi \right), \tag{7}$$

$f \in G_i(f)$, where $|f| < 1/(2T)$, $G(f) = G_{i1}(f) \cap G_{i2}(f)$,

$$G_{i1}(f) = \{ f : |H_i(f - \bar{k}_i/T)| > \xi\sqrt{\lambda_i} \} \tag{8a}$$

$$G_{i2}(f) = \left\{ f : \frac{|H_i(f - \bar{k}_i/T)|^2}{|H_j(f - \bar{k}_j/T)|^2} > \frac{\lambda_i}{\lambda_j} \right\} \tag{8b}$$

where $i \ne j$. For $f \notin G_{i1} \cap G_{i2}$ and for $k \ne \bar{k}_i$, $|P_i(f - k/T)| = 0$. The constants λ_1 and λ_2 are selected to satisfy the constraint (6). We show that the solution to (7)-(8) is unique, subject to appropriate restrictions on $H_i(f)$.

Note that where $|P_i(f)| \ne 0$, it has the same form as the MMSE transmitter filter for the single user channel with transfer function $H_i(f)$. The MMSE transmitter filters for the matched filter receiver also has this property. Since meas $(G_{12} \cap G_{22}) = 0$, the preceding results imply that for the MMSE receiver and the type of multiple-access channel considered, Frequency Division Multiple Access (FDMA) is optimal. This is also true for the matched filter detector. Specific examples of $H_1(f)$ and $H_2(f)$ along with optimized pulse shapes and an associated comparison of MSE are planned for presentation at the conference.

REFERENCES

1. T. Berger and D. W. Tufts, "Optimum Pulse Amplitude Modulation, Part I: Transmitter-Receiver Design and Bounds from Information Theory," *IEEE Trans. on Inform. Theory*, Vol. IT-13, No. 2, pp. 196-208, April 1967.

2. M. Honig, P. Crespo, and K. Steiglitz, "Suppression of Near- and Far-End Crosstalk by Linear Pre- and Post-Filtering", *IEEE Journal on Selected Areas in Comm.*, Vol. 10, No. 3, pp. 614-629, April 1992.

Optimum Sequence Multisets for Symbol-Synchronous Code-Division Multiple-Access Channels

Marcel Rupf and James L. Massey

Institute for Signal and Information Processing
ETH Zentrum, CH-8092 Zürich, Switzerland

A discrete-time symbol-synchronous code-division multiple-access (S-CDMA) system is considered where K independent users spread their real-valued encoded symbols $B_k, k = 1, \ldots, K$, by individual real signature sequences $\vec{s}_k$ of length L 'chips' and, then, transmit the L-dimensional symbols $B_k\vec{s}_k$ over a Gaussian multiple-access (GMAC, [1]) channel with noise correlation matrix $E[\vec{N}\vec{N}^T] = \sigma^2 I_L$ (T denotes transposition and I_L is the $L \times L$ identity matrix). Moreover, the same symbol-energy constraint is assumed for all users, i.e., $E[B_k^2] \leq \mathcal{E}_c$ and $\vec{s}_k^T\vec{s}_k = L$. For convenience, the L-dimensional observation vector at the output of the GMAC channel is written as

$$\vec{Y} = S\vec{B} + \vec{N} \qquad (1)$$

where the sequence matrix $S = [\vec{s}_1, \ldots, \vec{s}_K]$ and the symbol vector $\vec{B} = [B_1, \ldots, B_K]^T$. Finally, an optimum multiuser receiver is supposed which makes a joint maximum-likelihood decision of all information data.

The goal of this presentation is to find those sequence multisets, consisting of K not necessarily different sequences, which enable the K users to communicate reliably and fairly with maximum sum rate. As a consequence, equation (1) is viewed as a S-CDMA channel having a capacity region $\mathcal{C}(S)$ which is a function of the sequence matrix. The criterion of goodness for a sequence multiset is chosen to be the largeness of the symmetric capacity $C_{sym}(S)$ per chip where $C_{sym}(S)$ is defined by the maximum achievable equal-rate point in the capacity region $\mathcal{C}(S)$. It is achieved with zero-mean Gaussian distributed encoded symbols of maximum allowed variance $\mathcal{E}_c$ [2].

Theorem 1: Let the real sequence matrix $S = [\vec{s}_1, \ldots, \vec{s}_K]$ consists of K L-dimensional sequences $\vec{s}_k$ of equal energy L. Then,

$$C_{sym}(S) \leq \frac{1}{2}\log\left(1 + K\frac{\mathcal{E}_c}{\sigma^2}\right) \quad \text{[bits/chip]}$$

with equality if and only if the L rows of S are orthogonal and have equal norm K, i.e. $SS^T = KI_L$.

This upper bound is equal to the sum capacity of a GMAC with noise variance σ^2 and K chip-inputs of equal energy $\mathcal{E}_c$ [1, p.378]. Therefore, it can be concluded that (chip-)dimensions can be used most efficiently in a fair communication as long as $SS^T = KI_L$ in S-CDMA. Note that a necessary condition for $SS^T = KI_L$ is $K \geq L$. Moreover, $SS^T = KI_L$ is the necessary and sufficient condition for a sequence multiset to meet Welch's lower bound on the sum of the squares of the inner products between all pairs of K equal-energy sequences [3].

In the presentation, Theorem 1 will also be generalized for the case of two-dimensional modulation. Additionally, further properties of Welch-bound-equality sequence multisets will be mentioned.

References

[1] Cover, T.M., Thomas, J.A.: "Elements of Information Theory", John Wiley and Sons, Inc., 1991.

[2] Verdú, S.: "Capacity Region of Gaussian CDMA Channels: The Symbol-Synchronous Case", Proc. 24th Allerton Conf., pp. 1025-1034, October 1986.

[3] Massey, J.L., Mittelholzer, Th.: "Welch's Bound and Sequence Sets for Code-Division Multiple-Access Systems", Sequences 91, Positano, Italy, 17-22 June 1991 (Springer-Verlag, Ed. R. Capocelli).

Optimally Orthogonal Time-Limited Signals Under RMS Bandwidth Constraints

DARA PARSAVAND and MAHESH K. VARANASI

Department of Electrical and Computer Engineering
University of Colorado, Boulder, CO 80309
e-mail: parsavan@prony.colorado.edu

Abstract

The determinant of the correlation matrix between n time-limited unit-energy signals can be seen as a measure of orthogonality of the signal set. The problem of designing a signal set that maximizes this determinant is considered under the average as well as the maximum root mean square (RMS) bandwidth constraints.

Summary

In multiuser communication, an important signal design problem is to choose a set of unit-energy signature signals that are optimally orthogonal. This corresponds to an autocorrelation matrix that is as close as possible to the identity matrix. If there is no constraint on the bandwidth, it is possible to choose orthogonal signals. However a nontrivial constraint on the bandwidth necessitates that the signals be non-orthogonal. The optimality measure on the correlation matrix is defined in this paper to be its determinant. This measure is chosen due in part to its significance in the PAM synchronous Gaussian CDMA channel where the capacity region was characterized in the high signal-to-noise ratio regions via the total asymptotic efficiency which in turn was shown to be upper bounded (achievably) by the determinant of the correlation matrix [1]. The bandwidth of each time-limited signal is defined to be its root mean square (RMS) bandwidth (cf. [2]).

In this paper we consider the problem of finding the optimally orthogonal unit-energy signals whose maximum RMS bandwidth is bounded by B_0. A solution to this problem is found by solving the problem with a weaker constraint, that of an average RMS bandwidth bounded by B_0, and establishing (constructively) the existence of a solution to the latter problem that consists of signals with equal RMS bandwidths. These signals will therefore also be a solution to the maximum RMS bandwidth problem, for if there were another signal set meeting the latter constraint and having a higher determinant, this set would supplant the original solution to the average RMS bandwidth problem.

The main result in [2] provides a closed-form solution for the set of signals that achieve the minimum average RMS bandwidth among all signal sets that have a given correlation matrix R. The signals have the following form:

$$s_k(t) = (2/T)^{1/2} \sum_{j=1}^{n} c_{kj} \sin(j\pi t/T), \quad 0 \le t \le T, \quad 1 \le k \le n, \quad (1)$$

where T is the duration of the signal and n is the number of signals. The solution for the coefficient matrix $C = \{c_{kj}\}$ is given by $C = \Lambda^{1/2}V$, where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$, $\lambda_i \ge \lambda_{i+1}$ are the ordered eigenvalues of R and V is the matrix of eigenvectors of R in its spectral decomposition, $R = V\Lambda V^T$. Furthermore, these signals have individual and average RMS bandwidths given by

$$B_i = (2T)^{-2}(V\Lambda\Pi V^T)_{ii} \quad \text{and} \quad B = (2T)^{-2}n^{-1}\mathrm{trace}(\Lambda\Pi), \quad (2)$$

where Π is defined to be a diagonal matrix with $\Pi_{ii} = i^2$.

In light of the result of [2] stated above, the average RMS bandwidth problem can be equivalently posed as that of finding a nonnegative definite unit-diagonal correlation matrix R with maximum determinant under the constraint that the corresponding minimum average bandwidth given by (2) is bounded by B_0. The problem can be formulated entirely in terms of the n eigenvalues of the correlation matrix since $\det(R) = \det(\Lambda)$, the bandwidth constraint is expressed only in terms of Λ, and the constraints on R can also be represented by positivity and trace constraints on Λ. The average RMS problem is now expressed as:

$$\begin{aligned} \max \quad & \det(\Lambda) && (3)\\ \text{subject to} \quad & \lambda_i \ge \lambda_{i+1} \ge 0, && (4)\\ & \mathrm{trace}(\Lambda) = n, && (5)\\ & \mathrm{trace}(\Lambda\Pi) \le b \quad \text{with} \quad b = B_0(2T)^2 n. && (6) \end{aligned}$$

Non-trivial solutions for Λ exist when b is restricted to the range $(n, n(n+1)(2n+1)/6)$.

The first result in this paper simplifies the nonlinear optimization problem (3) - (6) by showing that the ordering constraint in (4) can be relaxed to only a positivity constraint and the inequality constraint in (6) can be changed to the corresponding equality constraint. The new constraints are thus $\Lambda \ge 0$, $\mathrm{trace}(\Lambda) = n$ and $\mathrm{trace}(\Lambda\Pi) = b$. A proof of this result involves showing the suboptimality of any set of eigenvalues that either violates the ordered property or fails to use all the bandwidth given in the constraint. The problem is then solved using the Lagrange multiplier technique which involves finding the Lagrange multipliers numerically as a solution to a set of nonlinear equations. These equations are then solved by standard numerical techniques.

The next result of the paper gives a constructive procedure for finding a positive unit-diagonal definite matrix R of size n with the specified optimal eigenvalues. The procedure involves finding the optimal correlation matrix by starting with $R^{(1)} = \Lambda$ and performing at most $n-1$ rotations given by, $R^{(k+1)} = U^{(k)}R^{(k)}U^{(k)T}$, such that with $V = U^{(n-1)}U^{(n-2)}...U^{(1)}$, the correlation matrix $R = V\Lambda V^T$ has unit-diagonal elements and in addition, the matrix $V\Lambda\Pi V^T$ has equal diagonal elements. This ensures equal bandwidths of the optimal signal set, thereby solving the maximum RMS problem simultaneously. The problem of finding the optimal signal set is now identical to the problem solved in [2].

Finally, if n is such that a Hadamard matrix of dimension n exists, a single unitary (Hadamard) transformation is also shown to yield a signal set that solves the maximum RMS bandwidth problem. Hadamard matrices exist for all dimensions which are a power of 2 but also many others, (cf. [3]).

References

[1] Verdú, S., "Capacity region of Gaussian CDMA channels: the symbol-synchronous case," *Proc. of the Twenty-fourth Allerton Conference on Communication, Control and Computing*, Allerton, IL, pp. 1025-1034, October, 1986.

[2] Nuttall, A.H., "Minimum rms bandwidth of M time-limited signals with specified code or correlation matrix," *IEEE Transactions on Information Theory*, vol. IT-14, pp. 699-707, September 1968.

[3] Hall Jr., Marshall, *Combinatorial Theory*, Blaisdell Publishing Company, Waltham, MA, 1967.

ON ACHIEVABLE INTER-USER ORTHOGONALITY FOR MULTI-USER COMMUNICATION SYSTEMS IN MULTIPATH FADING ENVIRONMENTS

JÜRG RUPRECHT

Swiss PTT General Directorate, R&D, Mobile Communications VD 2, 3000 Berne 29, Switzerland

Abstract — **This paper proposes and compares three broadband modulation/demodulation schemes for use in a multipath fading environment. They are all based on CDMA, are of a broadband nature in order to combat frequency selective fading and achieve a certain degree of orthogonality in order to enhance spectral efficiency.**

I. Introduction

In multi-user communication systems where the users access the same channel, the optimum receiver is in the general case a joint detection receiver, i.e., the data symbols of all users have to be detected jointly. For a large number of users, this receiver is very complex and in most cases practically not implementable. To avoid joint detection, orthogonal modulation/demodulation schemes are desirable where a single user detector has the same performance independently of the number of active users.

In an AWGN environment, FDMA, TDMA and, for proper spreading code choices, CDMA are examples of such orthogonal modulation/demodulation schemes. When properly implemented, FDMA and TDMA keep inter-user orthogonality even in multipath environments, whereas conventional CDMA causes inter-user interference due to the lack of a sufficient number of orthogonal spreading codes. In order to overcome the spectral nulls of multipath fading channels, mainly wideband communication systems based on TDMA and CDMA are currently considered for future mobile communications systems. CDMA offers many advantages, but suffers from the above-mentioned inter-user interference in multipath environment when a single user detector is applied. This presentation suggests and compares several approaches to orthogonalize the users in a synchronized CDMA system in a multipath fading environment.

II. Model

A discrete-time multi-user communication system as shown in Figure 1 is considered. Each user k ($k = 0, 1, \ldots K-1$) modulates its information sequence $b_k[.]$ with symbol rate R_S by converting it into the corresponding transmission sequence $x_k[.]$ with chip rate R_C. The total transmission sequence $x[.] = \sum_{k=0}^{K-1} x_k[.]$ is then transmitted through a common noiseless channel with impulse response $h[.]$ and is further distorted by a noise sequence $z[.]$. The corresponding received sequence $y[.] = h[.] * x[.] + z[.]$ then serves each demodulator $\#k$ as input for its estimate $\hat{b}_k[.]$ of $b_k[.]$. The goal is to design *orthogonal modulation/demodulation schemes*, i.e., the demodulator performance is the same independently of the number $k \leq K$ of active users.

III. Conventional Access Techniques

In an AWGN environment, FDMA (where the users are assigned non-overlapping spectra), TDMA (where the users are assigned non-overlapping time instants) and CDMA (where the users are assigned orthogonal spreading codes) provide orthogonal modulation/demodulation schemes. In a multipath fading environment, these access techniques perform as follows:

- FDMA still provides an orthogonal modulation/demodulation scheme. The spectral nulls, however, degrade the FDMA performance significantly when not combined with frequency hopping and/or interleaving.
- TDMA still provides an orthogonal modulation/demodulation scheme, if the users are separated in time such that no inter-user interference occurs. The system bandwidth is limited because of equalizer complexity.
- Conventional CDMA does no longer provide an orthogonal modulation/demodulation scheme, and the Qualcomm approach [2] provides only orthogonality on every single channel path, such that the system becomes interference limited for single user detectors.

Thus, FDMA and TDMA provide orthogonal modulation/demodulation schemes even in multipath fading environments, but yield the stated disadvantages. On the other hand, CDMA offers the above-mentioned advantages.

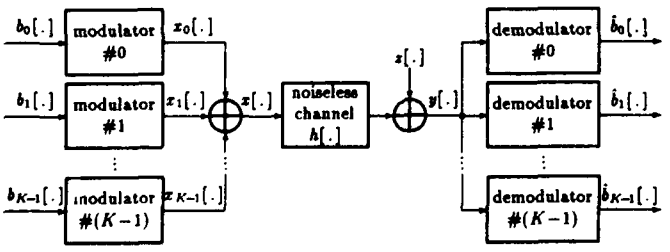


Figure 1: Multi-user communication system

IV. Proposed Access Techniques

We therefore suggest and compare the following approaches to orthogonalize the users in a sCDMA system in a multipath fading environment:

- *Code frequency division multiple access (CFDMA)* is a new accessing scheme. It spreads an narrowband FDMA system with a spreading sequence in a CDMA fashion. CFDMA is orthogonal on each channel path, but different channel paths interfere.
- *Code time division multiple access (CTDMA)* [1] spreads a conventional TDMA system in a CDMA fashion, where all users spread with the same sequence. An inverse filter receiver garantees an orthogonal modulation/demodulation scheme even in a multipath fading environment.
- *Code code division multiple access (CCDMA)*, as proposed by Qualcomm as CDMA [2], scrambles a CDMA system (spread by orthogonal Walsh codes) by an additional PN sequence. As CFDMA, CCDMA is orthogonal on each channel path, but different channel paths interfere.

V. Conclusion

CTDMA is the only scheme that keeps full modulation/demodulation orthogonality. On the other hand, the number of possible users is limited by the maximum excess delay of the channel impulse response. In CFDMA and CCDMA, only partial modulation/demodulation orthogonality can be provided, but user capacity of the system does no longer depend as severely on the maximum excess delay of the channel. These accessing schemes are proposed and compared in performance for different situations.

Literature

[1] Ruprecht, J., Neeser, F.D., Hufschmid, M., "Code time division multiple access: An indoor cellular system", *Proceedings of the 42nd IEEE Vehicular Technology Conference*, pp. 736-739, Denver, 1992.

[2] Salmasi, A., Gilhausen, K.S., "On the system design of code division multiple access (CDMA) applied to digital and personal communication networks", *Proceedings of the 41th IEEE Vehicular Technology Conference*, pp. 57-62, St. Louis, 1991.

CHANNEL CODING FOR ASYNCHRONOUS FIBEROPTIC CDMA COMMUNICATIONS

M. Dale
TASC Corp.
Reston, VA

R. Gagliardi
Univ of Southern Calif
Los Angeles, CA

Code Division Multiple Access (CDMA) has been proposed as a possible format for fiberoptic networks. The baseline CDMA uses on-off keying (OOK) of binary data with a unique coded pulse sequence transmitted for each on-bit. Multiple accessing is achieved by having multiple sources, each with its own code sequence, superimpose their transmissions over a common fiber. The fibers can then be interconnected via STAR or other fiber systems to form the distribution network Data bits are separated out at a receiving terminal by recognizing (correlating) the proper sequence of the desired source. Pulse code sequences can be passively generated from an initial OOK laser pulse by serial or parallel delay lines , and sequence correlation can be achieved optically by corresponding matched delay lines. After correlating the desired sequence to a peak value, photodetection followed by threshold comparison can be used to detect the presence or absence of each bit. Minimal interference multiple accessing is achieved by using only sets of pulse code sequences that have low pairwise crosscorrelations. Optical CDMA has the advantage of permitting completely asynchronous transmitters, relatively simple, off-the-shelf laser sources, standard photodetectors, and improved power levels due to the laser pulsing. In addition, pulsed CDMA combines the higher speeds of optical signals with the more developed electronic processing to provide maximum performance efficiencies in converting digital data to optical transmission.

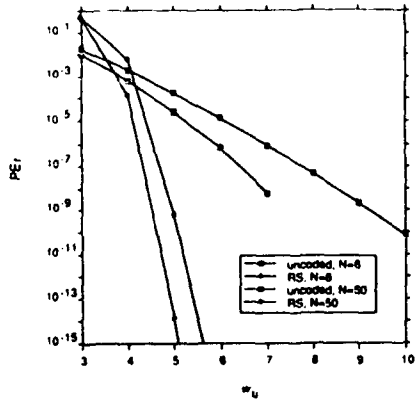
A prime disadvantage with CDMA is the sacrifice in per-channel data rate (relative to the speeds available in the laser itself) that occurs in the insertion of code addressing. Another important CDMA concern is the development of digital crosstalk between channels when multiplexing many simultaneous sources. Channel crosstalk is the ultimate limit in link performance, and produces an asymptotic floor to the error probability (PE) that can only be reduced by slowing the data rate with longer and higher weight CDMA sequences . This raises the question of whether external channel coding can be more effective in reducing the PE floor.

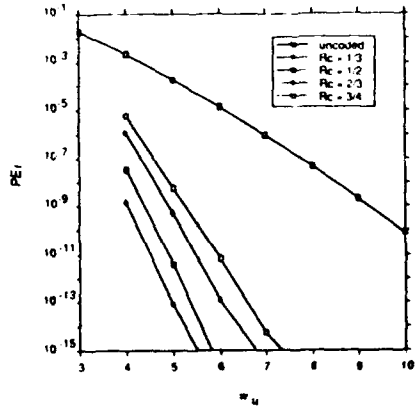
In this paper we consider the use of external channel coding in the form of either forward error correction or modified waveform encoding, to aid in mitigating this crosstalk buildup and produce more efficient individual channel performance. While the advantages of channel coding are well known for the classical Gaussian noise channel, the application to the optical CDMA crosstalk channel is somewhat diverse, and care must be used inserting commonly accepted coding "gains".

The channel coding is applied directly to the CDMA fiber links. The channel coder converts the data bits to binary synbols, which are then sent over the fiber link as OOK symbols encoded with the transmitter code sequence. Forward error correction in the form of Reed-Solomon (RS) block codes and convolutional coding (CC) were considered. It was assumed that all systems use the same laser pulse width, number of transmitters, and channel data rate. The CDMA code sequence were adjusted to accommodate each type of coding. Figure 1 and 2 show example results for RS and CC, indicating the reduction in the PE crosstalk floor from the uncoded case, as a function of the CDMA sequence weight.

in conclusion,we have shown that channel coding can indeed be more effective in reducing the PE floor. Both Reed-Solomon block codes and convolutional codes were considered, with both showing improved PE performance over the uncoded OOK-CDMA system at the same information bit rate. The convolutional codes tended to produce lower PE floors for the lower code weight values and hence would tend have the highest network capacities. This result is important since the channel coding is applied with electronic hardware external to the optics. Hence the channel coding should have limited impact on the overall system cost.

The use of PPM as a channel coding technique was also considered, since is also reduces the PE floor at the expense of data rate. It was shown that the PPM system was not as effective as the uncoded and error correction systems at the same data rate. Furthermore PPM requires a modification of the optical encoding and decoding processing from the standard OOK-CDMA format.





A SIGNALING TECHNIQUE FOR MULTIPLE ACCESS LASER COMMUNICATIONS

John E. Hershey, Nabeel A. Riza, and A. A. Hassan
GE Corporate Research and Development Center, P.O.Box 8,
Schenectady NY 12301 USA.

Summary

A multi-dimensional signaling technique is described for use in asynchronous multiple access laser communications. This technique is based on interferometric signaling and can be thought of as temporal and spatial coding of light. Transmitter implementation requires a coherent source (a laser), signal modulation electronics, and some special optics. Spatial modulation is accomplished via an aperture populated with liquid crystal (LC) modulators with different users having different LC distributions. These distributions are chosen to yield low cross-correlation between the (Fraunhofer) diffraction patterns. The large spatial bandwidth (e.g., 10^6 pixels) of each laser transmitter aperture is utilized for user coding, while temporal coding is used for information signals. Signal recovery is based on incoherent optical detection, spatial sampling, and electronic or optical matched filtering of the received optical beam Fresnel/Fraunhofer diffraction pattern. With electronic filtering, low to medium (e.g., 3 Mbps) data rates can be achieved. With a lenslet array-based incoherent optical correlator, up to 100 Mbps data rates can be tolerated.

Assume the liquid crystal modulators are evenly placed on a circular plane of radius R. For a laser with angular frequency ω it can be shown that the light intensity $I(x,y)$ at position (x,y) of the receiver plane at a distance L from the transmitter is approximately given by

$$I(x,y) = \sum_{p=1}^{N}\sum_{q=1}^{N} \alpha_p\, \alpha_q\, \cos\left\{\frac{R\omega}{Lc}\right. \left.[(\cos\theta_p - \cos\theta_q)\, x + (\sin\theta_p - \sin\theta_q)\, y]\right\},$$

where c is the speed of light, $\theta_i = 2\pi i/N$ is the angular spacing of the liquid crystal modulators, and α_i is a phase modulation equal to +1 or -1. In general, α_i is a complex valued code symbol

$$\alpha_i = r_i \exp\{j\ \phi_i\},$$

r_i being the amplitude modulation and ϕ_i is the phase modulation of symbol i; this yields a more elaborate expression for the received intensity. A codeword (or signaling mask) for user k is given by

$$A_k = (\alpha_1(k),\ \ldots\ , \alpha_N(k)),$$

and the multiple access codebook is given by

$$\mathrm{A} = \{A_1,\ \ldots\ , A_U\},$$

where N can be thought of as the block length of the code, and U is the number of users. As in sequences used in spread spectrum applications, A is designed to have codewords with low cross-correlations and auto-correlations.

At the end of each signaling interval, a 2-D correlator in the receiver crosscorrelates the sampled intensity with stored user specific intensity functions to decide what transmitters are on. The capability of this system to discriminate among multiple users is demonstrated, and preliminary results on the design of A is shown.

VARIABLE WEIGHT OPTICAL ORTHOGONAL CODES FOR CDMA NETWORKS WITH MULTIPLE PERFORMANCE REQUIREMENTS

Guu-chang Yang
Department of Electrical Engineering
National Chung-Hsing University
Taichung, Taiwan, R.O.C.

Abstract

An optical orthogonal code (OOC) is a collection of binary sequences with good auto- and cross-correlation properties; they were defined by Salehi and others as a means of obtaining code division multiple access (CDMA) on fiber optic networks. Up to now all work on OOC's have assumed that the weight of each codeword is the same. In this paper we develop bounds on the size of OOC's when this assumption is removed. In addition, we demonstrate construction techniques for building such "variable weight" OOC's. The results demonstrate that it is possible to assign codewords with different weights among the users. Changing the weight of a user's signature sequence affects that user's performance; therefore this approach is useful for CDMA fiber optic networks with multiple performance requirements among the users.

Summary

1. Background and Motivation

As the demand for personal communication services continues to rise, multiple access techniques become ever more important. Code-division multiple access (CDMA) is a kind of spread spectrum technology that enables many users to share the same channel without interference by employing a unique signature sequence to distinguish different users' transmission.

Optical orthogonal codes (OOC's) were introduced by Salehi *et. al* [1-2] as a means of obtaining code division multiple access among asynchronous users on fiber optic networks. An OOC is a family of (0,1) sequences with good auto- and cross-correlation properties, and a variable weight OOC is an OOC in which the weight of each codeword is not constant over the code.

Throughout this summary, we use W, L, and Q, to denote the sets $\{w_0, w_1, \ldots, w_p\}$, $\{\lambda_0, \lambda_1, \ldots, \lambda_p\}$, and $\{q_0, q_1, \ldots, q_p\}$, respectively.

Definition: An (n, W, L, λ_c, Q) variable weight optical orthogonal code $\mathcal{C}$ is a collection of binary n-tuples such that the following three properties hold:

- (Weight Distributution) Every n-tuple in $\mathcal{C}$ has a Hamming weight contained in the set W; furthermore, there are exactly $q_i \cdot |\mathcal{C}|$ codeword of weight w_i - i.e., q_i indicates the fraction of codewords of weight w_i.
- (Auto-correlation Property) For any $\mathbf{x} = [x_0, x_1, \ldots, x_{n-1}] \in \mathcal{C}$ with Hamming weight $w_i \in A$ and any integer τ, $0 < \tau < n$,
$$\sum_{t=0}^{n-1} x_t x_{t\oplus\tau} \leq \lambda_i.$$
- (Cross-correlation Property) For any $\mathbf{x} = [x_0, x_1, \ldots, x_{n-1}] \in \mathcal{C}$ and any $\mathbf{y} = [y_0, y_1, \ldots, y_{n-1}] \in \mathcal{C}$ such that $\mathbf{x} \neq \mathbf{y}$ and any integer τ,
$$\sum_{t=0}^{n-1} x_t y_{t\oplus\tau} \leq \lambda_c.$$

  Note: OOC's were defined in terms of periodic correlation; thus the addition in the subscripts above - denoted "$\oplus$" - is all modulo-n.

The definition of a variable weight OOC is a generalization of the defintion for OOC given in [1-2]. The use of OOC's for multiple access is described in [1-3].

- The auto-correlation constraint guarantees that each signature sequence is unlike cyclic shifts of itself. This property is used to enable the receiver to obtain *synchronization*.
- The cross-correlation constraint guarantees that each signature sequence is unlike cyclic shifts of the *other* signature sequences. This property is used to enable the receiver to estimate its message in the presence of interference from other users.

A reasonable "figure of merit" for a code is the number of interfering users necessary to cause the code to fail. For instance, assume synchronization has been achieved; then the only errors the i^{th} user can make are $0 \rightarrow 1$ errors, and they can only occur when enough *other* users interfere to make the correlation at the i^{th} receiver exceed $w_{\pi(i)}$. (Here, $w_{\pi(i)}$ is the Hamming weight of i^{th} user's codeword.). Since each of those other users can contribute at most λ_c to the correlation, the performance "figure of merit" is $w_{\pi(i)}/\lambda_c$. In a similar vein, the synchronization "figure of merit" is $(w_{\pi(i)} - \lambda_{\pi(i)})/\lambda_c$ for multiple-access synchronization and $w_{\pi(i)} - \lambda_{a_{\pi(i)}}$ for single-user synchronization. (Here, $\lambda_{\pi(i)}$ is the auto-correlation associated with codeword's of weight $w_{\pi(i)}$.

From above, we can see that the weight of a user's signature sequence will strongly affect that user's performance. Therefore, by assigning codewords with different weights we are able to accomodate multiple performance requirements among the network's users.

2. New Results

This paper will detail new techniques for analyzing variable weight OOC's and related sequences. New methods of constructing such codes have been found and new bounds on the size of such codes have been derived.

2.1. A New Bound

Define $\phi(n, W, L, \lambda_c, Q)$ to be the cardinality of an optimal variable weight optical orthogonal code with the given parameters – i.e.,

$$\phi(n, W, L, \lambda_c, Q) \triangleq \max\{|\mathcal{C}|: \mathcal{C} \text{ is an } (n, W, L, \lambda_c, Q) \text{ variable weight OOC}\}.$$

We have derived a new upper bound on $\phi(n, W, L, \lambda_c, Q)$ for $\lambda_{a_i} \geq \lambda_c$ $(\lambda_{a_i} \in L)$.

Theorem: Let $\lambda_{a_i} \geq \lambda_c$ $(\lambda_{a_i} \in L)$. Then

$$\phi(n, W, L, \lambda_c, Q) \leq \frac{(n-1)(n-2)\ldots(n-\lambda_c)}{\sum_{i=0}^{p} q_i w_i (w_i - 1)(w_i - 2)\ldots(w_i - \lambda_c)/\lambda_{a_i}}.$$

We note at this point that the technique used to prove this theorem is immediately applicable to binary codes employed for CDMA when the auto-correlation and/or the cross-correlation constraints are specified in terms of *aperiodic* correlation as well.

2.2. New Constructions

Several new approaches for constructing variable weight OOC's have been found. Among them:

- We can use the balanced incomplete block design technique [3] to construct $(n, \{w+1, w\}, \{2,2\}, 1, Q)$, $(n, \{2w, w\}, \{2,2\}, 1, Q)$, $(n, \{2w+1, w\}, \{2,2\}, 1, Q)$, $(n, \{2w+1, w+1\}, \{2,2\}, 1, Q)$, and $(n, \{2w+1, w+1, w\}, \{2,2,2\}, 1, Q)$ variable weight OOCs for even w. $(n, \{w+1, w\}, \{1,1\}, 1, Q)$, $(n, \{2w, w\}, \{2,1\}, 1, Q)$, and $(n, \{2w+1, w\}, \{2,1\}, 1, Q)$ variable weight OOCs for odd w. Among these constructions, the cardinality of the $(n, \{w+1, w\}, \{1,1\}, 1, Q)$ variable weight OOC reaches the upper bound of the last section; hence it is optimal.
- We have generalized the recursive construction method of [4] to construct variable weight OOC's. The recursive construction technique uses the codes which are constructed by previous techniques to provide infinite families of codes.

References

[1] J. A. Salehi, "Code Division Multiple Access Techniques in Optical Fiber Networks-Part I: Fundamental Principles," *IEEE Transactions on Communications*, pp. 824-833, Aug., 1989.

[2] J. A. Salehi and C. A. Brackett, "Code Division Multiple Ac cess Techniques in Optical Fiber Networks-Part II: Systems Performance Analysis ," *IEEE Transactions on Communications*, pp. 834-850, Aug., 1989.

[3] G. C. Yang and Thomas Fuja, "Optical Orthogonal Codes with Unequal Auto- and Cross-correlation Constraints," *Proceedings of The Twenty-Sixth Annual Conference on Information Sciences and Systems*, Princeton, New Jersey, March, 1992.

[4] M. J. Colbourn and C. J. Colbourn, "Recursive Constructions for Cyclic Block Designs," *J. Statistical Planning and Inference*, vol 10, pp. 97-103, 1984.

OPTICAL SPECTRAL AMPLITUDE CODE DIVISION MULTIPLE ACCESS SYSTEM

Maïté Brandt-Pearce and Behnaam Aazhang
Department of Electrical and Computer Engineering
Rice University, Houston, Texas 77251-1892

A system is proposed and analyzed which fully utilizes the bandwidth available in the optical medium. This optical code-division multiple-access system illustrated in Figure 1 has its signature sequences on-off encoded on the frequency bands of the optical beam. Two suboptimal detector options for multi-user operation are considered: an optimized single-user detector and a multistage detector. The analysis is performed using a large deviation theory approximation and further verified by simulation. The advantages of this system over the conventional time-encoded system include the larger number of low crosscorrelation sequences available and the implementation of efficient decoders for low error probability detection.

1. Summary

Two of the possible optical sources with bandwidths broad enough are: a coherent mode-locked laser source and an incoherent source, such as a superfluorescent fiber source. The mode-locked laser is more difficult to implement than the incoherent source yet less noisy due to its Poisson photoelectron count statistics, compared to the doubly stochastic Poisson statistics of the incoherent source. This paper concentrates on the laser based system, which can be considered a best case yielding a lower error probability, mentioning the differences with the incoherent source system when applicable.

The novelty of this system is in the CDMA encoding, which is achieved by spatially spreading the optical spectrum and on-off modulating the resulting frequency bands. The spreading is accomplished by the use of a diffraction grating and the encoding is done via an amplitude mask. The system is composed of K users, labelled $k = 1, \ldots, K$, each transmitting continuous binary information $b_k^{(\ell)} \in \{0,1\}$, $\ell = \ldots, -1, 0, 1, \ldots$. The symbol $b_k^{(\ell)}$ on-off modulates the optical beam, which is then encoded by a mask representing the sequence $A_{k,j} \in \{0,1\}, j = 1, \ldots, J$. The mask allows frequency bands corresponding to $A_{k,j} = 1$ to pass and blocks the other frequencies. The total integrated intensity of the modulated signal in one frequency band over the bit period $(0, T]$ is

$$r_j = b_1^{(1)} A_{1,j} \int_0^T |E_{1,j}(t)|^2 dt + \sum_{k=2}^{K} A_{k,j} \left[b_k^{(0)} \int_{T-\tau_k}^{T} |E_{k,j}(t)|^2 dt + b_k^{(1)} \int_0^{T-\tau_k} |E_{k,j}(t)|^2 dt \right] + \Lambda_d$$

where $E_{k,j}(t)$ is the optical field of user k in frequency band j and τ_k is its delay with respect to the desired user, user one. The number of low crosscorrelation encoding sequences available to the spectral amplitude encoded system is a factor of J more than for the time-encoded system, since the auto-correlation constraint can be completely relaxed. The sequences of interest for the frequency encoded system are codes with fixed weight w and minimum distance $2(w-1)$, for a maximum overlap of one frequency band.

The detection is based on spreading the signal spatially exactly as in the encoding stage, and then detecting the individual frequency bands using a photodetector array integrating over the entire bit interval. One of two proposed detection algorithms then follows: an optimized single-user detector or a multistage detector [1]. The degradations to the system considered in this model are the multiple access interference, the noise due to photoelectron statistics, and the dark current noise. The photoelectron counts of the J frequency bands, labelled $N_j, j = 1, \ldots, J$, are Poisson distributed with parameter r_j for the laser based system. The statistics for a truly incoherent source, i.e., thermal light, depend on the shape of the spectrum but in this case can be approximated by a negative binomial distribution. For the laser system, letting $\Lambda_{k,j}(t_1, t_2) = \int_{t_1}^{t_2} |E_{k,j}(t)|^2 dt$ and letting the interference in frequency band j be $I_j = \sum_{k=2}^{K} A_{k,j} \left[b_k^{(0)} \Lambda_{k,j}(T - \tau_k, T) + b_k^{(1)} \Lambda_{k,j}(0, T - \tau_k) \right] + \Lambda_d$, the optimized single-user detector has the form

$$\sum_{j=1}^{J} \ln \left\{ \frac{\sum_\lambda (\Lambda_{1,j} A_{1,j} + \lambda)^{N_j} e^{-\lambda T} p_{I_j}(\lambda)}{\sum_\lambda (\lambda)^{N_j} e^{-\lambda T} p_{I_j}(\lambda)} \right\} \begin{matrix} > \\ = \\ < \end{matrix} \gamma, \qquad (1)$$

with $p_{I_j}(\lambda)$ a convolution of all possible interference distributions, and the multi-stage detector has the form

$$\sum_{j=1}^{J} N_j A_{1,j} \ln \left(1 + \frac{\Lambda_{1,j}}{\hat{I}_j} \right) \begin{matrix} > \\ = \\ < \end{matrix} \gamma, \qquad (2)$$

where $\hat{I}_j$ is an estimate of this interference based on the previous stage of detection and γ is the threshold.

The primary performance analysis tool is large deviation theory [1], through which the probability of error is derived by computing the expected value over uniformly distributed delays τ_k of the synchronous OCDMA system in [1]. Using this scheme, an approximation to the performance is obtained, which is verified by simulation to be very accurate for all sequences considered. The coherent laser optical source case is analyzed, considering a correlation detector, an optimized single-user detector, and a multistage detector. The advantage gained by the increase in code size is quantified and compared to time-encoded systems, as illustrated in Figure 2. More than twice as many users can be supported using spectral encoding than time encoding if a multistage detector is employed. Yet as explained before, at such bandwidths, the time-encoded system can only employ a correlation detector, whose performance is shown to be unacceptable for a large number of users.

REFERENCES

[1] M. Brandt-Pearce and B. Aazhang, "Unequal received power effects on single-user and multi-user detectors for optical cdma," *Proceedings of the 22nd Annual Conference on Information Sciences and Systems, Princeton University, Princeton, NJ*, March, 1992.

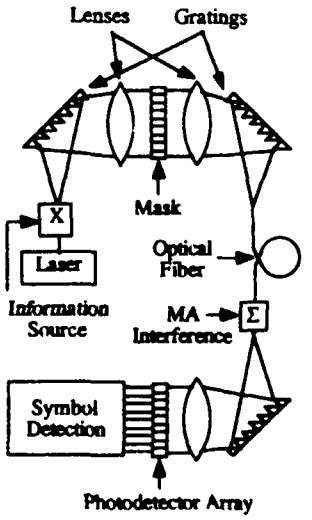


Figure 1: Optical spectral amplitude encoded communication system.

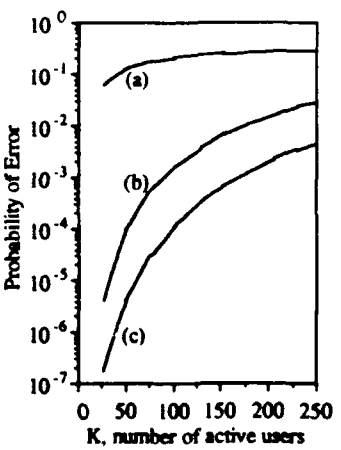


Figure 2: (a) Time encoded system using correlation detector, (b) ... using multistage detector, (c) Spectral encoded system using multistage detector

This work is supported in part by NASA/Johnson Space Center under grant NGT-50447 and by the Advanced Technology Program of the Texas Higher Education Coordinating Board under Grant 003604-018.

IMPROVED CONCATENATED CODING/DECODING FOR DEEP SPACE PROBES

Dale C. Linne von Berg
Stephen G. Wilson

Department of Electrical Engineering
University of Virginia
Charlottesville, VA 22901

Deep space communication systems are traditionally designed to use a concatenated coding scheme employing an inner and an outer code in order to obtain reliable communication at low SNR. The Consultative Committee on Space Data Systems, CCSDS, uses a concatenated coding system with a rate 1/2, memory order 6, convolutional inner code and a Reed Solomon (RS) block outer code over GF(256). Eight RS words are organized into eight columns in a 255x64 bit data frame. Before the data frame is transmitted, the data is trellis encoded by rows. On the receiving end, decoding is performed in two steps: the inner decoder uses the maximum-likelihood (soft) Viterbi algorithm and the outer decoder uses RS algebraic decoding methods. Paaske has shown that the performance of the CCSDS scheme can be improved by providing multiple-pass feedback from the outer to inner decoders [1]. The decoded output from the outer decoder is used by the inner decoder to pin states in subsequent decoding passes. As Collins has shown [2], the error correcting performance of a trellis code is improved when reliable side information is used to pin states during decoding, roughly gaining in energy efficiency by $10\log_{10} f$, where f is the pinning fraction.

In this paper, we study two issues related to further improvement of the performance of this concatenated coding system, without increasing the transmission power, the bandwidth, or the decoder complexity. First, we investigate the options of byte versus bit interleaving between the outer and inner systems. This has a subtle effect--byte interleaving is a better choice as far as Reed-Solomon decoding goes, since inner decoder errors are not diffused throughout the interleaving array. However, we find that for a given pinning fraction, it is more advantageous to scatter the side-information uniformly in time, i.e., use bit interleaving. At the bottom line, however, we find that byte interleaving is slightly superior.

Second, we investigate the design of RS outer coding with a variable rate in various columns, keeping the overall frame redundancy constant. Thus some codeword columns have less parity than nominal, while others have much greater parity. This variable-rate feature is no real complication for decoding due to the highly flexible decoding algorithms for RS codes. Driven by a comment of Paaske that the multi-pass scheme succeeds with high probability if one column is successful, we design one column with high redundancy (but not so high as to exhaust the available parity symbols), then incrementally design the redundancy of other columns to provide high probability of success given that the previous columns have been correctly decoded. Optimizing this profile has been done experimentally by simulating the entire system. At $E_b/N_0 = 1.5$ dB we suggest that a parity profile across the eight columns of (26,28,32,36,100,8,4,22) is near-optimal in terms of maximizing probability of correct decoding of a frame. At 2 dB a more uniform profile (26,28,30,32,90,16,14,20) seems preferable. We note the asymmetry of the profile accrues from the fact that pinning has an asymmetric effect on cleaning out errors in the Viterbi decoding process; known-state information prunes away more errors "ahead of" the decoder than behind it in time. The improvement in link efficiency with this "optimized" scheme is about 0.4 dB, clearly not a huge amount, but fractions of a decibel are precious in this regime.

Simulations also show that use of larger-than-normal decoder delay is helpful at very low SNR conditions. The usual rule of thumb (five constraint lengths) suggests 30 bit decoder delay, but we observe that 60 bit delay is a simple way to gain another 0.15 dB. Finally, relaxing the restriction on bandwidth can significantly improve performance at low SNR. By replacing the rate 1/2 inner code with a rate 1/4 code optimized at low SNR [3], the bandwidth is expanded by a factor of two, but by keeping the memory order of the inner code constant, the decoder complexity remains constant and the required transmission power can be reduced. Operation at $E_b/N_0 = 1$ dB is feasible in this case.

References

[1] E. Paaske, "Cost-Efficient Methods for Improving Coding Gains in Concatenated Coding Systems for Deep Space Missions", Int'l Symposium on Information Theory, p. 297, Budapest, 1991.
See also "Improved Decoding for a Concatenated Coding System Recommended by CCSDS", IEEE Trans. on Communications, Vol. 38, pp. 1138-1144, August 1990.

[2] O. Collins, "Pruning the Trellis", Int'l Symposium on Information Theory, p. 50, Budapest, 1991.

[3] P. J. Lee, "New Short Constraint Length, Rate 1/N Convolutional Codes Which Minimize the Required SNR for Given Desired Bit Error Rates", IEEE Trans. on Communications, Vol. 33, pp. 171-177, Feb. 1985.

[4] D. C. Linne von Berg, "Improved Concatenated Coding/Decoding for Deep Space Probes," M.S. thesis, University of Virginia, 1992.

CHANGING THE CODING SYSTEM ON A SPACECRAFT IN FLIGHT

Kar-Ming Cheung, Dariush Divsalar, Sam Dolinar, Ivan Onyszchuk, Fabrizio Pollara, and Laif Swanson
Jet Propulsion Laboratory, California Institute of Technology

For many years, the data stream sent by American deep-space missions has been protected by error-correcting codes. For *Galileo*, a probe and orbiter currently en route to explore Jupiter and its moons, NASA's standard constraint length 7, rate 1/2 code is available, but a constraint length 15, rate 1/4 convolutional code was developed for the mission and a prototype Viterbi decoder for long constraint length codes has been built. For parts of the mission, the convolutional code was to be the inner code in a concatenated system with an outer (255,223) Reed-Solomon code.

Data compression has been used in deep space much less than channel coding for three reasons. First, source models are neither as developed nor as simple as the deep-space channel model (AWGN). Second, most data compression algorithms require substantial complexity for encoding, and calculations on a spacecraft are much more difficult than on the ground. But third and most important, scientists are very slow to accept the distortions that come with much compression, when they are unlikely to be able to gather this data again in their careers. *Galileo's imaging system includes the option for lossless* data compression.

Last year, during its trip to Jupiter, *Galileo*'s collapsible high-gain antenna was scheduled to unfurl, but efforts to open it have not been successful yet. If the high-gain antenna were to remain closed, all communication would be via a low-gain antenna. Without any additional changes, this would cause the achievable data rate to drop by four orders of magnitude. A small part of this loss is due to the fact that the (15, 1/4) encoder is not accessible for data going through a low-gain antenna.

If the high-gain antenna remains unusable, we will change *Galileo's* error-correcting codes and data compression algorithm. Of course, the spacecraft hardware is completely inaccessible, but at the data rates we are now considering (about 100 bits per second) a lot can be done in software, even on *Galileo*'s somewhat old computers.

Software convolutional encoding would be easy, except that *Galileo*'s design requires that most communication through the low-gain antenna must pass through the (7, 1/2) encoder after leaving the spacecraft's computers. This means that any code must be realizable as a concatenated code, with a (7, 1/2) code on the inside. An (11, 1/2) code concatenated with a (7,1/2) code yields a (14, 1/4) code, and many of these (14, 1/4) codes perform nearly as well as the best known (14, 1/4) codes. But no such code has taps on both ends of all connection vectors, a property of "good codes" and a requirement to be decoded in a straightforward way by our hardware Viterbi decoder. Again, we are saved by data rate: we can decode the resulting (14, 1/4) convolutional code in software. In addition, we may change the outer Reed-Solomon code from (255, 223) to a system with words of different parity in each interleaved frame; this would mean that some words in each frame are almost certain to decode, giving more information about the state of the convolutional encoder; this information can then be used by a second Viterbi decoder to "redecode" with substantially smaller error rate.

Until now, data compression algorithms for deep-space have been limited to lossless compression, and thus to about 3.6 bits/pixel, while slightly lossy algorithms like the proposed Joint Photographic Experts Group (JPEG) standard show almost no visual degradation at less than 1 bit/pixel, compared to our original 8 bits/pixel. Because the communications rate with the low-gain antenna is so low, many of the planned images would not be sent at all, and so the small distortions introduced by data compression are now much more attractive to the scientists.

A JPEG standard 8x8 Discrete Cosine Transform (DCT) is not possible within our constrained memory and computation resources, but we intend to implement a similar multiplication-free integer transform, the Integer Cosine Transform (ICT), which can compress a typical planetary image 10:1 with an RMS error of 1 (out of 256) gray level, (peak SNR 48 dB), or 20:1 with an RMS error of 2, with memory requirements of 4K bytes for code and 7K bytes for buffer, using 32 adds and 12 shifts for each 8-point DCT. This algorithm will be used on *Galileo's* images if a low-gain mission is required.

A Modification of Generalized Concatenated Codes and its Applications

Yan Gao, Uwe Dettmar and Ulrich K. Sorger
Institut für Netzwerk- und Signaltheorie, Technical University of Darmstadt
Merckstraße 25, 6100 Darmstadt, Germany

Abstract– We propose a modification of generalized concatenated codes, which allows the construction of some best known binary linear codes in a very simple way. As another application we show that by using this method we can generate a big class of optimal linear unequal error protection codes (LUEP codes) very easily and that most of the constructions given by van Gils [1] are special cases of this new method. A big advantage of our method is, that all constructed codes can be decoded very easily by the well known Blokh–Zyablov–Zinov'ev algorithm with a slightly modified metric.

Summary

Up to now all constructions of generalized concatenated codes (GC codes, [2][3]) are restricted to have outer codes A_i of constant length n_a and only one inner code B together with its partitions. In this paper we construct binary GC codes which have outer codes A_i of different lengthes n_{ai} and hence different binary inner codes $B^{(j)}$ in the different columns of the code matrix.

Denote by $A_i : (q_i; n_{ai}, k_{ai}, d_{ai})$ the outer code A_i over $GF(q_i)$ and of length n_{ai}, dimension k_{ai} and minimum distance d_{ai} and by $B_i^{(j)} : (n_b^{(j)}, k_{bi}^{(j)}, d_{bi}^{(j)})$ the ith subcode of the jth binary inner code with $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n_{a,max} = \max\{n_{ai}\}$. By concatenating outer and inner codes the symbols of the outer code A_i is used to label the subcode $B_{i+1}^{(j)}$ and its cosets obtained by partitioning the ith subcode $B_i^{(j)}$. The new GC code has the parameters:

$$n = \sum_{j=1}^{n_{a,max}} n_b^{(j)}, \qquad k = \sum_{i=1}^{m} k_{ai} \log_2(q_i),$$

$$d_{min} \geq \min_{\forall i} \left(\min_{\forall \mathcal{J}_i} \left(\sum_{j \in \mathcal{J}_i} d_{bi}^{(j)} \right) \right)$$

where $\mathcal{J}_i \subset \{1, ..n_{ai}\}$ with $|\mathcal{J}_i| = d_{ai}$.

Table 1 shows some best known codes constructed in this way. The inner code $B^{(j)}$ and its subcodes for $j = 1, 2, \cdots, n_{am}$ are given in the first column, and the inner codes $B^{(n_{am}+1)}$ and $B^{(n_{am}+2)}$ are given in the 2^{nd} and 3^{rd} columns.

The same idea can be used to construct optimal LUEP codes. In the following we give some examples.

Construction I: First we consider a two-level GC code ($m = 2$), where we take $A_1 : (2^{k_{b1}}; n_a, k_{a1}, \mathbb{s}_1)$ and $A_2 : (2^{k_{b2}}; n_a, k_{a2}, \mathbb{s}_2)$ as outer codes and $B_1 : (n_b, k_{b1} + k_{b2}, d_{b1})$ and its subcode $B_2 : (n_b, k_{b2}, d_{b2})$ as inner codes. A_1 and A_2 are LUEP codes with nonincreasing separation vectors $\mathbb{s}_1 = (s_{11}, s_{12}, \cdots, s_{1k_{a1}})$ and $\mathbb{s}_2 = (s_{21}, s_{22}, \cdots, s_{2k_{a2}})$. If $d_{b1} s_{1k_{a1}} \geq d_{b2} s_{21}$, then the GC code is a binary $(n_a n_b, k_{a1} k_{b1} + k_{a2} k_{b2}, \mathbb{s})$ LUEP code, where $\mathbb{s} = (d_{b1} s_{11} 1_{k_{b1}}, d_{b1} s_{12} 1_{k_{b1}}, \cdots, d_{b1} s_{1k_{a1}} 1_{k_{b1}}, d_{b2} s_{21} 1_{k_{b2}}, \cdots, d_{b2} s_{2k_{a2}} 1_{k_{b2}})$ ($s 1_{k_{bi}}$ denotes the k_{bi}-vector with all components equal to s). It can be seen that the *Construction 1, 3A* and 5 in [1] are special cases of the above construction, where some special codes are used as outer and inner codes to obtain optimal LUEP codes.

Construction II: The GC code with $A_1 : (2^{(n_a - k_b)}; n_a + n', k_{a1}, \mathbb{s}_1)$ and $A_2 : (2^{k_b}; n_a, k_{a2}, \mathbb{s}_2)$ as outer codes and $B_1 : (n_b, n_b, 1)$ and $B_2 : (n_b, k_b, d_{b2})$ as inner codes is a modified GC code, where the length of the first outer code A_1 is larger than the length of A_2. In this case the last n' symbols of a codeword in A_1 are not concatenated with other codes and just appended to the first $n_a n_b$ concatenated bits. The *LUEP code* has the parameters $n = n_a n_b + (n_b - k_b) n'$, $k = (n_b - k_b) k_{a1} + k_b k_{a2}$ and $\mathbb{s} = (s_{11} 1_{(n_b - k_b)}, \cdots, s_{1k_{a1}} 1_{(n_b - k_b)}, d_{b2} s_{21} 1_{k_b}, \cdots, d_{b2} s_{2k_{a2}} 1_{k_b})$, where $s_{1k_{a1}} \geq d_{b2} s_{21}$. Optimal LUEP codes can be obtained if some special codes are used as outer and inner codes. The codes from *Constructions A,C,E,F,I,J* and *K* in [1] can also be obtained in this way.

Inner codes			Outer codes	GC code
(8,4,4)	(7,3,4)	(4,1,4)	$(2^3; 10, 7bits, 8)$	(75,11,32)
(8,1,8)			(2;8,4,4)	
(8,4,4)	(7,3,4)	(4,3,2)	$(2^3; 10, 3, 8)$	(75,13,30)
(8,1,8)			(2;8,4,4)	
(8,4,4)	(7,3,4)	(3,1,3)	$(2^3; 10, 7bits, 8)$	(74,11,31)
(8,1,8)			(2;8,4,4)	
(8,4,4)	(4,3,2)	(4,3,2)	$(2^3; 10, 3, 8)$	(72,13,28)
(8,1,8)			(2;8,4,4)	
(6,6,1)	(5,5,1)		(2;16,1,16)	(95,52,14)
(6,5,2)	(5,4,2)		$(2^4; 16, 10, 7)$	
(6,1,6)			(2;15,11,3)	

Table 1: Some modified GC codes

Construction III: The GC code with $A_1 : (2^{k_b}; n_a, k_{a1}, \mathbb{s}_1)$ and $A_2 : (2; n_a, k_{a2}, \mathbb{s}_2)$ as outer codes and $B_1 : (n_b, k_b + 1, d_b)$ and $B_2 : (n_b, 1, n_b)$ as inner codes is a special case of *Construction I*. If we add to the outer code A_2 a parity bit, which is not concatenated with the inner code B_2, for $s_{1k_{a1}} d_b > s_{21} n_b$ we obtain a new $(n_a n_b + 1, k_{a1} k_b + k_{a2}, (s_{11} d_b 1_{k_b}, \cdots, s_{1k_{a1}} d_b 1_{k_b}, (n_b - 1)\mathbb{s}_2 + 2\lceil \mathbb{s}_2/2 \rceil))$ LUEP code , where $\lceil \mathbb{s}_2/2 \rceil$ denotes $\lceil s_{2i}/2 \rceil$ for all $i = 1, 2, \cdots, k_{a2}$. Here if we take some special codes as outer and inner codes, the same codes can be obtained as from the *Construction 3B*, B and H in [1]

A big advantage of the new construction method is that the codes can be decoded very easily up to half of their minimum distance. The decoding algorithm is quite similar to the well-known Blokh-Zyablov-Zinov'ev algorithm but with a slightly modified metric.

References

[1] W. Gils, *Design of Error-Control Coding Schemes for Three Problems of Noisy Information Transmmission, Storage and Processing.* PhD thesis, Eindhoven Univ. of Technology, Eindhoven,The Netherlands, 1988.

[2] E. Blokh and V. Zyablov, "Coding of generalized concatenated codes," *Probl. Inform. Transm.*, vol. 10, no. 3, pp. 218–222, 1974.

[3] V. Zinov'ev, "Generalized concatenated codes for channels with error bursts and independent errors," *Probl. Inform. Transm.*, vol. 17, no. 4, pp. 53–56, 1981.

PERFORMANCE OF CONCATENATED CODING SYSTEMS FOR CHANNELS WITH MEMORY

G. Ferland
Department of Electrical and Computer Engineering
Royal Military College of Canada

ABSTRACT

This paper presents the analysis of concatenated coding systems with inner convolutional codes and outer burst-error-correcting codes. Contrary to most studies, interleavers are not required to perform the analysis. A modelling technique for the outer channel resulting from a convolutional decoder operating over a wide variety of inner channels is presented. The proposed outer model is simple and is determined for inner channels with and without memory. The resulting finite state outer channel model is entirely characterized by the transition probabilities between the states and a technique to compute these transition probabilities is proposed. For memoryless channels, hard and soft decision decoding are both considered. The well known Gilbert and Fritchman models are used to represent channels with memory. Once the outer channel model is determined, it is used to compute the bit error performance of the entire concatenated coding system considering both convolutional and block outer codes, the Massey λ-diffuse and the Reed-Solomon codes respectively. Several examples are presented to illustrate the applicability of the method. Results are obtained both by analytical methods and computer simulation.

SUMMARY

Concatenated coding schemes are used in digital communication systems. They have the ability to correct long error sequences and provide very high reliability. This paper presents a methodology to evaluate the performance of concatenated coding systems based on inner convolutional codes and outer burst-error-correcting codes. A typical configuration is a short constraint length inner convolutional code followed by a Reed-Solomon (RS) outer code. An essential aspect of our work is that we can also consider the use of an outer convolutional code.

In this work particular attention is paid to the modelling of the outer channel formed by the inner encoder, the inner channel and the inner decoder. When a maximum likelihood decoder like the Viterbi decoder is used, the outer channel exhibits a tendency to produce bursts and should be modelled by a channel with memory. Often an interleaver is used to remove the memory from the outer channel, its presence is not required in this study but can be accounted for if needed. Since a maximum likelihood decoder generates error events of various lengths, the decoding process is very difficult to analyze. To circumvent this difficulty, we have modelled the errors at the output of a sub-optimum decoder called the sliding window decoder (SWD). This decoder operates on a finite window size L and approaches the performance of the Viterbi decoder as L increases to infinity. The decoding process of the SWD is modelled by a 2-state Markov chain, a correct and an incorrect state, characterized by the transition probabilities p_{ci} and p_{ic}. The decoding model (e.g. the outer model) has been determined for both memoryless channels and channels with memory. The memoryless channels are the BSC and the AWGN channel with BPSK modulated signals and soft decision decoding. Quantization with a finite number of intervals is considered and arbitrary metric assignments can be used. Results for a (2,1,2) code and integer metric assignments are illustrated in Figure 1. Gilbert's [1] and Fritchman's [2] models are used to represent channels with memory.

Then the bit error performance of the concatenated coding system considering both convolutional and block outer codes is determined from the outer model mentioned above. The outer convolutional code considered are the λ-diffuse Massey [3] codes which are entirely defined by their burst-error-correcting capability B and the required guard space G. The outer block code considered are the well known RS codes which can correct both multiple bursts and random errors. These codes are specified by their block length and their symbol-error-correcting capability. All the theoretical results have been verified against simulations results and show good agreement. Figure 2 shows the simulated versus analytical performance results for both random error and burst error channels. On this Figure the detrimental effect of memory on the performance of the concatenated system can be observed. Results also indicates that the Reed-Solomon codes outperform the Massey Diffuse codes for the same error-correcting capability. Finally we find that soft decision decoding on the inner code with a finite number of levels (Q = 4 and 8) can provide most of the gain in performance for the concatenated system that would be obtained using infinitely fine quantization.

References:

[1] E.N. Gilbert, "Capacity of a Burst-Noise Channel", BSTJ, Sept. 1960.

[2] B.D. Fritchman, "A Binary Channel Characterization Using Partitioned Markov Chains", IEEE Trans. on Inf. Theory, Vol. IT-13, April 1967.

[3] J.L. Massey, "Implementation of Burst-Correcting Convolutional Codes", IEEE Trans. on Inform. Theory, Vol. IT-11, July 1965.

Acknowledgement:

P.J. McLane, Department of Electrical Engineering, is thanked for his Ph.D. supervision of this research. The research was performed at Queen's University.

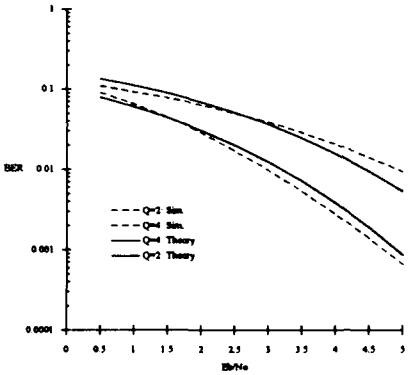


Figure 1 Performance of a (2,1,2) code: BPSK modulation and soft decision decoding.

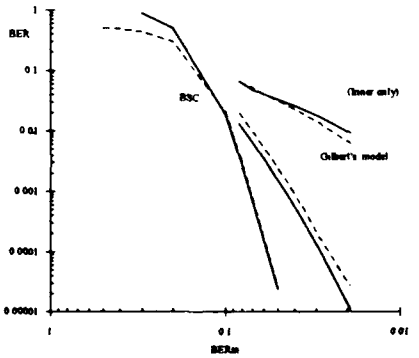


Figure 2 Performance of a Concatenated Coding system: BSC compared to a Gilbert model.

A Performance Analysis for Adaptive Rate, Trellis Coded Hybrid-ARQ Protocols

Lars K. Rasmussen and Stephen B. Wicker
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract

Hybrid-ARQ protocols are known to improve the reliability of communication systems at the expense of the total throughput. Systems based on trellis coded modulation (TCM) can be modified for use in type-I hybrid-ARQ schemes. In order to regain some of the lost throughput due to retransmission requests, various adaptive code rate algorithms can be applied. In this paper the concept of averaged diversity combining is considered. Multiple received copies of the same data packet are combined on a symbol-by-symbol basis to improve the performance. A simple example of an adaptive rate, TCM hybrid-ARQ protocol is developed and performance bounds on both bit error rate and throughput are derived for AWGN channels. To verify the bounds simulation results have been obtained and are presented.

1 Summary

In this paper we examine the performance of a trellis coded hybrid-ARQ protocol (TCM-HARQ) with packet combining. The Yamamoto and Itoh algorithm [1] is used to generate retransmission requests as investigated by Wicker and Rasmussen [2]. The derivation of the throughput bounds for the combining protocol follows the approach of Kallel and Haccoun [3]. The expected number of transmissions is bounded as follows:

$$1+\sum_{i=1}^{\infty}\left\langle [P_l(R_1)]^{i-1}\cdot\prod_{j=1}^{i}P_l(R_j)\right\rangle \le Tr \le 1+\sum_{i=1}^{\infty}P_u(R_i) \quad (1)$$

Here, $P(R_L)$ is the probability of a retransmission after L packets have been combined. The lower index on P indicates whether a lower or an upper bound is to be used. The averaged diversity combining of L packets is equivalent to a decrease of the effective noise variance by a factor $1/L$. Introducing the noise improvement factor $1/L$, the bounds on both throughput and bit error rate (BER) developed by Wicker and Rasmussen [2] can then be applied. For the BER the upper bound is as follows:

$$\begin{aligned} P_B \;\le\; & P_u(B_1)-\sum_{i=1}^{\infty}\left\langle\left[(P_l(R_1))^i\left\{\prod_{j=1}^{i}P_l(R_j)\right\}-P_u(R_{i+1})\right]\right. \\ & \left.\cdot\,[P_l(B_1)-P_u(B_{i+1})]\right\rangle \end{aligned} \quad (2)$$

Here, $P(B_L)$ is the probability of bit error after L packets have been combined. The lower bound is derived the same way.

$$P_B \ge P_l(B_1)\cdot\left(1-\sum_{i=1}^{\infty}\left\langle\left[P_u(R_i)-(P_l(R_1))^i\left\{\prod_{j=1}^{i+1}P_l(R_j)\right\}\right]\right\rangle\right) \quad (3)$$

For low signal-to-noise ratios (SNR) the lower bound tends to break down. An approximation is thus more useful.

$$\begin{aligned} P_B \;\simeq\; & P_l(B_1)-\sum_{i=0}^{\infty}\left\langle P_u(R_1)^{i+2}\cdot[P_u(R_{i+1})-P_u(R_{i+2})]\right. \\ & \left.\cdot\,[P_l(B_1)-P_u(B_{i+2})]\right\rangle \end{aligned} \quad (4)$$

A simple 2-state, 4-PSK TCM-HARQ protocol has been investigated in detail and simulation data obtained. The results are shown in Figures 1 and 2. For the BER in Figure 1 the approximation is noted to be very good. The awkward behavior of the BER should be noted here as more and more packets are combined. Under conditions in which all packets are combined with at least one other packet, a decrease in the BER is observed. This phenomenon is easily explained, for the combination of packets is equivalent to improving the SNR. At some point a majority of accepted packets will consist of combined packets and thus the effective SNR will be improved. In Figure 2 a substantial improvement in throughput is observed. At SNR's where no throughput is expected for the non-combining case it is now possible to get close to 50 % of full throughput through packet combining. The simulated throughput follows the lower bound very closely. This bound is thus a good approximation to the real throughput.

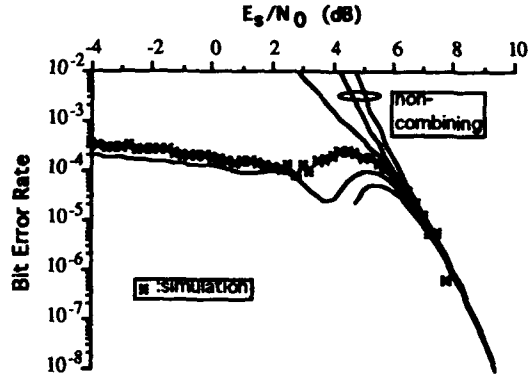


Figure 1: BER performance of a TCM-HARQ system based on a (2,1,1) convolutional encoder and QPSK modulation.

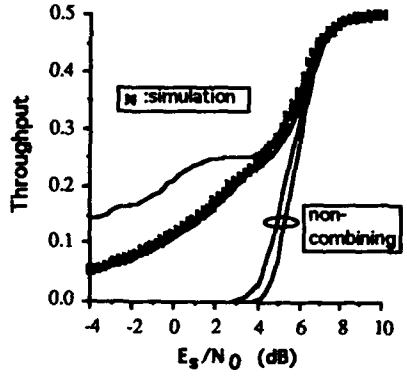


Figure 2: Throughput performance of a TCM-HARQ system based on a (2,1,1) convolutional encoder and QPSK modulation.

References

[1] Yamamoto H. and Itoh K. Viterbi decoding algorithm for convolutional codes with repeat request. *IEEE Trans. Info. Th.*, IT-26(5):540–547, Sept. 1980.

[2] Wicker, S.B. and Rasmussen, L.K. A performance analysis for trellis coded hybrid-ARQ protocols. In *Proc. IEEE Supercomm, ICC '92*, pages 323.7.1–5, June 14-18 1992. Chicago, IL.

[3] Kallel, S and Haccoun, D. Sequential decoding with ARQ and code combining: A robust hybrid FEC/ARQ system. *IEEE Trans. Com.*, COM-36(7):773–780, July 1988.

Further Results on the Convolutionally Coded ARQ with GVA decoding

Takeshi Hashimoto

Dept. Elect. Eng., Univ. Electro-Communications
Chofugaoka 1-5-1, Chofu, Tokyo 182, JAPAN
e-mail: hasimoto@liszt.ee.uec.ac.jp

Coded ARQ, or hybrid ARQ, is an effective scheme in order to attain high reliability and high throughput for channels with a moderate noise intensity. In 1980, Yamamoto and Itoh proposed a coded ARQ based on a convolutional code and the Viterbi decoding and show that a good bit error probability is attained. As the channel noise increases, however, its performance rapidly deteriorates because of increased retransmission requests and decoding errors. In 1990 ISIT, the author proposed a new coded ARQ scheme which exploits, for error detection, the error propagation caused by reduced-complexity decoding of convolutional codes with an extremely large constraint length and showed, by a random coding argument, the attainability of both high reliability and high throughput. This asymptotic result, however, does not assure that a good performance is practically attained; more realistic discussions have been expected.

In this paper, we presents a simple scheme for constructing the desired code, say the ARQ code hereafter, from a convolutional code with a short constraint length and from a BCH block code. In this construction, the convolutional code takes the role of error correction and the block code takes the role of propagating a decoding error to the next ARQ block. This ARQ code possesses a particular unit-memory structure between ARQ blocks besides its original trellis structure.

Recently, we found that Kudryashov also discussed a coded ARQ scheme using such a unit-memory structure between ARQ blocks and considered its performance by a random coding argument. Although his coding scheme is basically block coding and no particular code construction scheme is suggested, both works show an intimate relationship. Stimulated by his work, next, we show that the performance bound presented before can be considerably strengthened.

We show, using the above mentioned ARQ code and the generalized Viterbi decoding algorithm, that a good performance is obtained for a considerably large channel noise. Especially interesting is that high reliability is attained near R_{comp}, the computational cut-off rate of the channel. This is not expected for coded ARQ schemes based on sequential decoding algorithms. In the simulation, at least several thousands ARQ blocks of block length about 500 bits are transmitted and "high reliability" means that no incorrect acceptance of erroneously decoded ARQ block is observed. For a larger channel noise, however, the throughput becomes sensitive to the rule which decides whether a particular ARQ block is decoded incorrectly. Our original scheme, as well as Kudryashov's scheme, uses the threshold decision $\log \frac{P(\mathbf{y}|\mathbf{x})}{q(\mathbf{y})} > T$ for error detection. We also consider the use of error-detecting codes for error decision and compare the performance through a theoretical analysis and simulation. It is shown that the use of error-detecting code increases the robustness of the scheme and allow us to attain high reliability at rates above R_{comp}.

AN ADAPTIVE TRANSMISSION SCHEME FOR METEOR-BURST COMMUNICATION[1]

Guy Bégin
Dept. of Mathematics and Computer Science, Université du Québec à Montréal
Montréal (Québec) Canada H3C 3P8

Abstract — We consider the use of rate-compatible variable-rate punctured codes for implementing an adaptive transmission scheme for MB communication. The performance of the scheme is investigated both theoretically and through computer simulation. The results indicate that a rate-adaptive strategy leads to a more efficient use of available received power than fixed coding rate strategies.

I Introduction

Meteor-burst (MB) communication is an attractive means of beyond line of sight radio communication for several applications [1]. In MB communication, propagation is achieved through reflection of transmitted signals from trails of ionized particles created by meteors entering the atmosphere. MB channels are characterized by random availability and received signal levels that decay with time. For the most prevalent type of trails, the signal-to-noise ratio (SNR) is modeled as exponentially decaying, i.e., $\mathrm{SNR}(t) = \mathrm{SNR}_0 e^{-t/\tau}$. The initial signal-to-noise ratio SNR_0 and the decay parameter τ are random variables which differ from trail to trail [2].

Error control coding provides powerful means for dealing with uneven received power [3]. With a fixed coding rate scheme however, power is wasted at the beginning of trails, while at the end of trails noise conditions are severe. We consider a solution that relies on the use of variable-rate punctured convolutional codes for implementing an adaptive transmission scheme for MB communication.

II Adaptive transmission scheme

Error performance may be considered satisfactory as long as the residual Bit Error Rate (BER) is kept below an acceptable level. The efficiency of the transmission scheme is then measured by the throughput achieved. Our scheme relies on the use of rate-compatible variable-rate punctured codes [4] for adapting the error correcting power to the variations of received power. Punctured codes with coding rates close to one are used for the first transmitted bits of a trail which require almost no error protection and, in parallel with the decay of received signal power, more and more redundancy is added to the transmission by progressively decreasing the coding rate.

III Performance of the scheme

The performance of the scheme has been investigated both theoretically and through computer simulation. The theoretical performance is obtained by modeling the time-varying MB channel as a rapid succession of stationary channels with decreasing values of average SNR. Error performance is obtained using classical union bound arguments, assuming additive white Gaussian noise. Several combinations of modulation types and quantization have been considered. Computer simulations were conducted using a similar model.

Both theoretical and simulation results indicate that the rate-adaptive strategy leads to a more efficient use of available received power than fixed coding rate strategies. As Fig. 1 shows, a plateau is observed on BER curves. This plateau corresponds to a range where received power in excess of what is required for the target BER is exchanged for additionnal throughput. Indeed, the average bit rate is seen to increase with increasing initial SNR in this range (Fig. 2).

The rate-adaptive strategy clearly outperforms fixed coding rate strategies with or without interleaving, achieving better overall throughput (Fig. 2). Some trails that do not provide sufficient SNR for sustaining transmission with the fixed rate strategies may be exploited with the adaptive scheme, which has the further advantage of providing continuous improvements in throughput.

[1]This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Fonds de l'UQAM.

References

[1] L.B. Milstein *et al.*, "Performance of Meteor-Burst Communication Channels," *IEEE J. Select. Areas Commun.*, Vol. SAC-5, pp. 146–154, Feb. 1987.

[2] G.R. Sugar, "Radio Propagation by reflection from Meteor Trails," *Proc. IEEE*, Vol. 52, pp. 117–136, Feb. 1964.

[3] S.Y. Mui, "Coding for Meteor Burst Communications," *IEEE Trans. Commun.*, Vol. COM-39, pp. 647–652, May 1991.

[4] J. Hagenauer, "Rate Compatible Punctured Convolutional Codes and their Applications," *IEEE Trans. Commun.*, Vol. COM-36, pp. 389–400, April 1988.

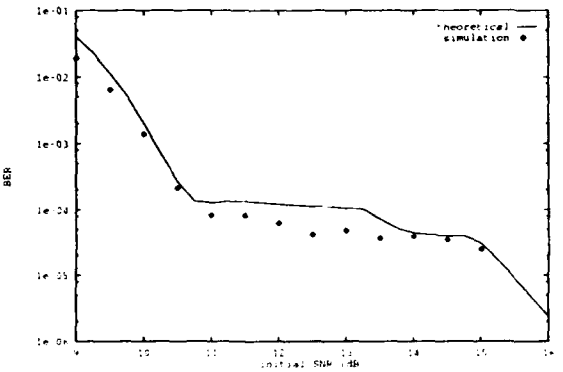


Figure 1: BER versus initial symbol energy-to-noise ratio. $M = 3$, $\tau = 2$, non-coherent FSK, hard quantization.

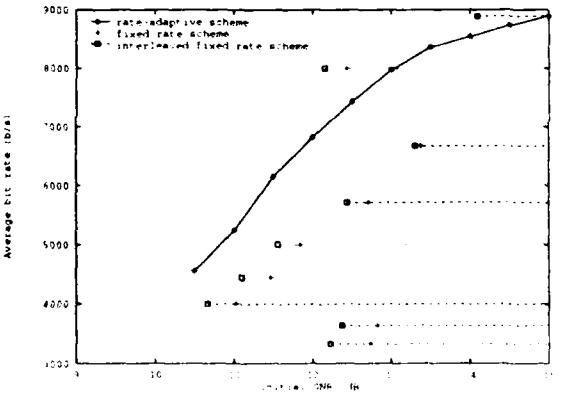


Figure 2: Throughput versus initial symbol energy-to-noise ratio. $M = 3$, $\tau = 2$, non-coherent FSK, hard quantization.

REAL CONVOLUTIONAL CODES EMBEDDED IN MULTICHANNEL DEMULTIPLEXERS FOR FAULT TOLERANCE*

Professor Robert Redinbo
Department of Electrical and Computer Engineering
University of California
Davis, CA 95616-5294

SUMMARY

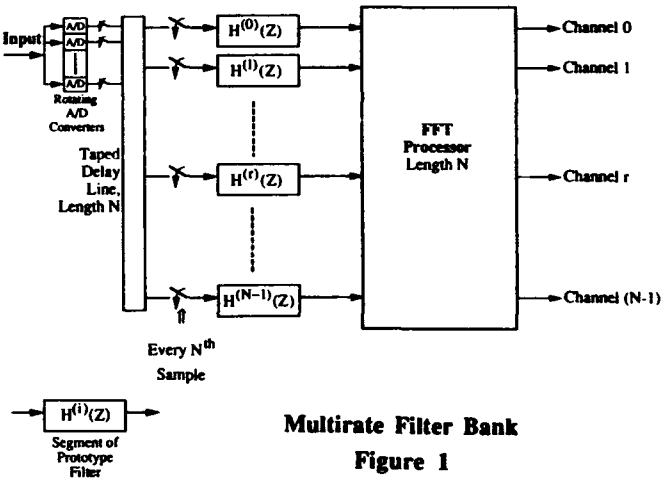
Many future communication systems employ a large number of channels tightly packed in separate frequency bands which need to be demultiplexed at a central site to achieve network connectivity. Mobile cellular and very small aperture satellite (VSAT) communication systems are two important examples where numerous channels are carried in a spectral segment through a central processing location. The practical feasibility of such systems rests on efficiently sharing common processing resources for simultaneously demultiplexing channels. This increased efficiency introduces a heightened susceptibility to both permanent and temporary failures in the underlying demultiplexing digital hardware. This paper demonstrates how real convolutional codes can be used for defining embedded parity streams which detect processing failures.

Multirate digital signal processing techniques capitalize on the narrowband nature of individual output channels permitting lower sample rates for most internal processing operations. The wide input spectrum requires a high sample rate whereas individual information channels need much lower sample rates. An efficient realization of the multichannel demultiplexer separates channels by spectrally shifting a prototype filtering operation. However, it may be segmented into many short filter sections, each operating at a lower sampling rate, whose outputs are passed through an appropriately-sized fast Fourier transform (FFT). The FFT realization is a major reason for the increased efficiency. Furthermore, the input A/D converters may be used in a rotating fashion to achieve the very high speed sampling necessary at the input. Contiguous samples are relegated to their respective positions in the lower speed processing sections for the remaining necessary operations. There are a wide range of efficient configurations of banks of subfilters for most practical applications. Figure 1 dramatizes the sharing nature of a multirate filter bank demultiplexer.

Any failure, whether permanent or temporary, virtually anywhere in an efficient multirate, multichannel realization impacts many channels simultaneously. One effective protection approach employs an algorithm based fault tolerance methodology, wherein parallel parity generating channels simultaneously produce a few parity samples for comparison with related parity values generated from the system outputs. The parallel parity producing resources may be used as standby units for replacing failed resources. Of course, then the protection levels are reduced accordingly. Real convolutional codes, derived from burst-correcting binary convolutional codes, are ideally suited for determining the parity values employed in the algorithm based fault tolerance approach. These codes are used in a detection mode only, and system diagnostic and reconfiguration phases may follow the detection of failures.

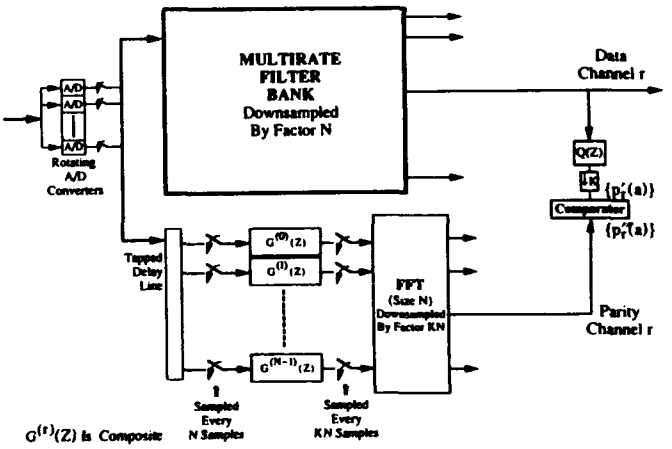
Rate K/(K+1) systematic binary burst-correcting convolutional codes produce one parity sample for every K information positions while still detecting the onset of a burst within a constraint length. When these codes are viewed over the real numbers and are judged by a real Hamming distance metric, excluding roundoff errors, it is known that their real error-detecting levels are at least as good as the binary precursor codes. The parity positions in these real number convolutional codes dictate a parity generating, finite impulse response filter with Z transfer function Q(Z) containing only 0 and 1 weights. The parity filter operates at a rate reduced by factor K and has a memory span determined by the constraint length of the original binary code.

The parity filter is used to produce low rate parity values from each demultiplexer channel. On the other hand, comparable parity values are generated in parallel with the demultiplexer from the original high-speed input samples, by combining the demultiplexing prototype filter H(Z) with the parity transfer function Q(Z). The required parity subsystem is very similar to the main demultiplexer, however, in the parallel parity process, the computational rates in the segmented filter sections and the FFT are reduced further by the factor K, a design parameter of the real code. The error-detecting fault tolerance capability for a generic channel r is depicted in Figure 2, where the down sampling feature in parity filter Q(Z) is indicated by ↓K. The comparator contains a threshold tolerance to accommodate roundoff noise introduced by the different computational paths for related parity values. The use of rotating A/D converters, necessary for high-speed performance, leads to easily detectable error energy in well-defined spare channels when an A/D converter fails.



Multirate Filter Bank

Figure 1



Protection of Demultiplexer Channels

Figure 2

* This was supported by NASA Lewis Research Center through grant NAG-3-1166, the National Science Foundation through grant MIP-9002664 and SDIO through the Office of Naval Research grant N00014-92-J-1759.

Performance Evaluation of Trellis-Coded Vector Quantization

René J. van der Vleuten Jos H. Weber

Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands

Abstract

We propose a new construction of trellis-coded vector quantizers (TCVQs), based upon our construction of trellis-coded quantizers (TCQs). The new construction yields TCVQs with a higher performance and is simpler than the previous construction given in [1].

The performances of the new TCVQs have been determined for the memoryless Gaussian, Laplacian, and uniform sources. The experiments that have been performed for various computational complexities and vector dimensions show that, at a constant rate and complexity, the performances of the TCVQs decrease as the dimension increases.

Thus, for memoryless sources, at the same rate and computational complexity, our TCQs are superior to our TCVQs.

Summary

A first constructive design method for trellis-coded quantizers (TCQs) and its extension to trellis-coded vector quantizers (TCVQs) have been given in [1, 2]. Recently, we proposed a new construction of TCQs for the rate of 1 b/sample [3, 4] as well as its extension to the higher rates [5]; our TCQs outperform those of [2]. Here, we present the extension of our construction to TCVQs.

The new construction yields TCVQs with a higher performance and is simpler than that of [1], since it does not make use of convolutional codes and uses trellises corresponding to a shift register. In [1], only two experiments were presented; these showed that the TCVQs outperformed the TCQs of [2]. We, however, performed various experiments which show that our TCQs are superior to our TCVQs, at the same computational complexity.

The TCVQs we consider have 2^ν states, with two branches entering and leaving each state. Each branch is assigned a set of representation symbols, according to the structure defined in [5]. Of course, in this case the representation symbols are not scalars, but vectors. Specifically, for quantizing at R b/sample using N-dimensional representation vectors ($N = 1, 2, \ldots$), each set contains 2^{NR-1} vectors.

To compare the performances of the TCVQs with those of our TCQs, experiments have been performed for memoryless Gaussian, Laplacian, and uniform sources. For the experiments, as in [1, 2, 3, 4, 5], a training set of $N \cdot 100\,000$ independent random samples was used. To optimize the codebook, 100 iterations were performed using an algorithm based on that described in [6], but extended for TCVQ and adapted to maintain the structure defined in [5]. Representation symbols onto which no input symbols are mapped are updated to zero (the average input value). The initial codebook was constructed from independent random samples from the distribution to be coded (maintaining of course the structure defined in [5]).

Complex.	TCQ:MF	TCVQ:FMW	TCVQ:VW	TCQ:VW
32	4.92	5.05	5.15	5.16
64	5.13	5.22	5.34	5.39

Table 1: SNRs (in dB), at the same complexity, for the TCQs of [2] (TCQ:MF), the TCVQs of [1] (TCVQ:FMW), the new TCVQs (TCVQ:VW), and the TCQs of [5] (TCQ:VW) for the Laplacian source, at $R = 1$.

As a measure of computational complexity we use the number of evaluations of the (single-sample) distortion function necessary to quantize one sample. Thus the complexity equals the product of the number of states, the number of branches (sets) per state, and the number of vectors per set: $2^\nu \cdot 2 \cdot 2^{NR-1} = 2^{\nu+NR}$. Table 1 shows a comparison, at the same complexities, of the performances of the TCQs of [2], the TCVQs of [1], the new TCVQs, and our TCQs [5] for quantizing the Laplacian source at 1 b/sample. Even though our TCVQs outperform those of [1], our TCQs are still superior.

In [1], only the two experiments shown in Table 1 were presented. We performed various experiments for dimensions N equal to 2, 4, and 8, and complexities up to 4096. They show that, at a constant complexity, the performances of the TCVQs decrease as the dimension increases.

Thus, for memoryless sources, at the same rate and computational complexity, our TCQs are superior to our TCVQs.

References

[1] T. R. Fischer, M. W. Marcellin, and M. Wang, "Trellis-coded vector quantization," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1551–1566, Nov. 1991.

[2] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. COM-38, pp. 82–93, Jan. 1990.

[3] R. J. van der Vleuten and J. H. Weber, "A new constructive design method for trellis waveform coders," in *Thirteenth Symp. Inform. Theory in the Benelux*, (Enschede, The Netherlands), pp. 15–22, June 1–2, 1992.

[4] R. J. van der Vleuten and J. H. Weber, "A new construction of trellis waveform coders," in *Signal Processing VI: Theories and Applications (EUSIPCO-92)*, (Brussels, Belgium), pp. 1477–1480, Elsevier Sci. Pub., Aug. 24–27, 1992.

[5] R. J. van der Vleuten and J. H. Weber, "A new construction of trellis-coded quantizers," in *Joint DIMACS/IEEE Workshop on Coding and Quantization*, (Piscataway, U.S.A.), Oct. 19–21, 1992. To be published.

[6] L. C. Stewart, R. M. Gray, and Y. Linde, "The design of trellis waveform coders," *IEEE Trans. Commun.*, vol. COM-30, pp. 702–710, Apr. 1982.

An Efficient Algorithm for Optimal Tree Pruning with Application to VQ

Xiaolin Wu * Yonggang Fang †

A vector space can be recursively partitioned into k convex regions with $k-1$ cutting halfplanes. Such a k-partition can be embedded into a binary tree of k leaves. This tree data structure was independently developed by researchers in VQ [6], pattern classification [1], databases [3] and computer graphics [4]. It was given different names in different research communities: quantizer tree, classification tree, k-d tree, and binary spatial partitioning (BSP) tree. In this abstract, we will use the term BSP tree. Each leaf of the BSP tree corresponds to a resulting convex region of the k-partition, and each internal node of the tree and its two sons correspond to a bipartition by a cutting halfplane. The BSP tree has a wide range of applications: source coding, pattern recognition, artificial intelligence, computer graphics, etc. However, for clarity and relevancy to the information theory symposium, we will study the problem in the framework of VQ. The results apply to other applications straightforwardly.

A tree-structured vector quantizer (TSVQ) has two attractive advantages: low design complexity and low decoding complexity compared with its unstructured counterparts. However, these computational advantages are gained at the expense of codebook optimality. Namely, the k-partition embedded into the BSP tree is not, in general, a Voronoi diagram on the k centroids. Two avenues were opened to improve the performance of TSVQ: 1) adaptive partitioning strategy and 2) optimal tree pruning. When growing the spatial partitioning tree we can optimize the cutting halfplanes one at a time by principal analysis [9], or optimize several cutting halfplanes together by dynamic programming as proposed by Wu [10], and/or elaborate on the order of tree growth using a look-ahead scheme as suggested by Riskin and Gray [7], and Wu and Zhang [9], rather than blind recursion. To further improve the performance of TSVQ, one can generate an initial BSP tree of larger size and then optimally prun it back to the required size. The first work on pruned TSVQ was due to Chou *et al.* [2], which was developed from an earlier work by Breiman *et al.* [1] on classification trees. The pruning of a BSP tree is an optimization problem of minimizing some objective function (say, quantization distortion) under certain constraints (e.g., the size or the expected height of the tree, the entropy of the leaves, etc.). The algorithm of [2] can find points on the convex hull of the objective function. However, given an arbitrary constraint value, this algorithm can only obtain the minimum through time sharing. Recently, Lin *et al.* [8] showed that optimally pruning an n-node initial tree to a k-node tree to minimize the total quantization distortion can be effected in $O(nk)$ time without time sharing.

In this research we found that the expected execution time of Lin *et al.*'s algorithm can be reduced from $O(n^2)$ to $O(n \log n)$ if $O(n) = O(k)$ (a common case in practice). We observed that the mechanic bottom-up testing in the search for the optimal subtree as conducted by the algorithm of [8] was often unnecessary and wasteful. Decisions can be made at a much earlier stage as to which subtrees can never be part of the final optimal k-node tree and which top portion of the initial BSP tree must stay in the final optimal k-node tree. Consequently, the search domain, and hence the algorithm execution time is drastically reduced.

Let $E(v)$ be the quantization distortion of the quantizer cell corresponding to a BSP tree node v. Then for each internal node v, we define its partitioning profit to be $\Delta(v) = E(v) - [E(v.leftson) + E(v.rightson)]$. $\Delta(v)$ quantifies how much reduction the corresponding bipartition of v brings to the total distortion. We sort all the internal nodes of an initial BSP tree T using $\Delta(v)$ as the key in descending order, and denote by $\lambda(v)$ the rank of v in this sorted list. Based on the ranking of the partitioning profits, we introduce a special kind of pruned trees of T, called *stable roofs*. A *stable roof* is defined to be a subtree R of T rooted at the root of T such that all the $\lfloor |R|/2 \rfloor$ internal nodes of R have the $\lfloor |R|/2 \rfloor$ largest partitioning profits among the internal nodes of T, namely, $\lambda(v) \leq \lfloor |R|/2 \rfloor$ if v is an internal node of R, where $|R|$ is the total number of nodes in R. By the above definition, the initial BSP tree T is itself a stable roof. In general, there may exist many stable roofs $R_i \subseteq T$. The subscript i in R_i denotes the size of the stable roof R_i, i.e., $i = |R_i|$. For simplicity we assume that all $\Delta(v)$ are distinct so that no two stable roofs have the same size.

Now investigate our problem of constructing the optimal pruned k-node tree $T_{opt}(k)$ from the initial BSP tree T, $k < n = |T|$. If there exists a stable roof R_k, then $R_k = T_{opt}(k)$, because any other internal nodes of T that are not the internal nodes of R_k have smaller profits, hence they cannot reduce the total quantization distortion further by replacing any of the internal nodes of R_k. Even if R_k does not exist, we may find two stable roofs R_i and R_j such that $i < k < j$. It is easy to see, in this case, that $R_i \subset T_{opt}(k)$ (in fact, R_i is also a stable roof of $T_{opt}(k)$), and that $v \notin R_j$ implies $v \notin T_{opt}(k)$. Therefore, we only need to prun the nodes between R_i and R_j, i.e., those v such that $v \in R_j$ but $v \notin R_i$, in search for $T_{opt}(k)$. The worst case is when T is the only stable roof of T. We have no bounds on the shape of $T_{opt}(k)$. Fortunately, this worst-case does not occur often in practice.

The stable roofs R_i and R_j such that $i < k < j$ serve as the upper and lower bounds on the shape of $T_{opt}(k)$. If the gap between R_i and R_j has m nodes, then the optimal pruning can be done in $O(m^2)$ time using the dynamic programming paradigm. In our experiments, m was very small independent of k, so the cost of $O(m^2)$ can be considered negligible. The family of stable roofs with respect to T can be found by the standard graph connected component algorithm. Starting with the initial graph consisting of the root of T and its two edges, we insert into the graph the internal nodes v with their edges in the descending order of $\Delta(v)$. In the insertion process, we dynamically detect the connected components containing the root of T and examine if they are stable roofs. Computing R_i and R_j requires $O(n \log n)$ time, which is spent on sorting $\Delta(v)$ and detecting stable roofs. Thus our claim on the time complexity of computing $T_{opt}(k)$.

References

[1] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Belmont, CA: Wadsworth.

[2] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Trans. Inf. Theory.*, vol. IT-35, no. 2, pp. 299-315, 1989.

[3] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. Math. Software*, Vol. 3, No. 3, pp. 209-226, Sept. 1977.

[4] H. Fuchs, G. Z. M. Kedem, and B. F. Naylor, "On visible surface generation by a priori tree structure," *Computer Graphics*, vol. 14, no. 3, p. 124-133, July 1980.

[5] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization for speech coding," *Proc. of IEEE*, vol. 73, pp. 1551-1588.

[6] E. A. Riskin and R. M. Gray, "A greedy tree growing algorithm for the design of variable rate vector quantizers," *IEEE Trans. Signal Proc.*, Nov. 1991.

[7] J. Lin, J. Storer, and M. Cohn, "On the complexity of optimal tree pruning for source coding," *Proc. of Data Compression Conference*, IEEE Computer Society Press, pp. 63-72, 1991.

[8] X. Wu and K. Zhang, "A better tree-structured vector quantizer," *Proc. of IEEE Data Compression Conference*, IEEE Computer Society Press, pp. 392-401, 1991.

[9] X. Wu, "Vector Quantizer design by constrained global optimization," *Proc. of IEEE Data Compression Conference '92*, IEEE Computer Society Press, p. 132-141, March, 1992.

*Dept. of Computer Science, Univ. of Western Ontario, London, Ont., Canada N6A 5B7

†On leave to the first author's department from Dept. of Electronic Engineering, Tsinghua University, Beijing, P. R. China

Asymptotic Entropy Constrained Performance of Tesselating and Universal Randomized Lattice Quantization

Tamás Linder
Department of Telecommunications
Technical University of Budapest
1521 Stoczek u. 2, Budapest, Hungary

and

Kenneth Zeger
Coordinated Science Laboratory
University of Illinois
1308 W. Main St., Urbana, IL 61801

ABSTRACT

Two results are given. First, using a result of Csiszár, the asymptotic (i.e., high resolution/low distortion) performance for entropy constrained tessellating vector quantization, heuristically derived by Gersho, is proven for all sources with finite differential entropy. This implies, using Gersho's Conjecture and Zador's formula, that tessellating vector quantizers are asymptotically optimal for this broad class of sources, and generalizes a rigorous result of Gish and Pierce from the scalar to vector case. Second, the asymptotic performance is established for Zamir and Feder's randomized lattice quantization. With the only assumption that the source has finite differential entropy, it is proven that the low distortion performance of the Zamir-Feder universal vector quantizer is asymptotically the same as that of the deterministic lattice quantizer.

SUMMARY

Let Q_N^k denote an N-level k-dimensional vector quantizer, and let X^k be the k-dimensional random vector with density f and differential entropy $h(f)$ to be quantized. Let the rth power quantization distortion be defined in the usual way,

$$D_r(Q_N^k(X^k)) = \frac{1}{k} E\|X^k - Q_N^k(X^k)\|^r,$$

where $\|\cdot\|$ denotes the Euclidian norm, and $r > 0$. Denote the Shannon entropy of Q_N^k by $H(Q_N^k)$, and for $H > 0$ let

$$D_r(H,k,r) = \inf_N \inf_{H(Q_N^k(X^k)) \le H} D_r(Q_N^k(X^k)), \tag{1}$$

the distortion of an optimal k-dimensional vector quantizer with entropy H. Gersho [2] heuristicly derived the asymptotic performance of quantizers given by the tessellation of $\mathcal{R}^k$ by a convex polytope P. He found that if $Q_{d,P}^k$ denotes the tesselating quantizer with rth power distortion d, then

$$\lim_{d\to 0} d\, 2^{\frac{2}{k}H(Q_{d,P}^k)} = l(P)\, 2^{\frac{2}{k}h(f)}, \tag{2}$$

where $l(P)$ is the normalized rth moment of P. We prove (2) in great generality. Our Theorem 1 establishes the asymptotic entropy constrained performance of lattice quantizers without any smoothness or compact support condition on the density. Thus the often quoted formula

$$\lim_{\Delta\to 0}[H(Q_\Delta) + \frac{1}{2}\log 12 D_2(Q_\Delta)] = h(f), \tag{3}$$

on the asymptotics of uniform quantizers is proved for all densities such that Q_Δ has finite Shannon entropy for some step size Δ, and $h(f) < \infty$, strengthening Gish and Pierce's result.

Ziv [4] introduced a randomized ("dithered") quantization scheme for scalar random variables which was extended by Zamir and Feder [3] for lattice vector quantizers.

We prove in Theorem 2 that for a large class of densities the asymptotic performance of the randomized lattice quantizer and the asymptotic performance of the ordinary lattice quantizer are the same.

Theorem 1 *If $|h(f)| < \infty$ and $H(Q_{d,P}^k(X^k)) < \infty$ for some $d > 0$, then*

$$\lim_{d\to 0} d\, 2^{\frac{2}{k}H(Q_{d,P}^k)} = l(P)\, 2^{\frac{2}{k}h(f)}. \tag{4}$$

Furthermore, if Zador's formula holds for f, $l(P) = C(k,r)$, and Gersho's conjecture holds, then

$$\lim_{d\to 0} \frac{D_e(H(Q_{d,P}^k),k,r)}{d} = 1, \tag{5}$$

i.e., the quantizer $Q_{d,P}^k$ is asymptotically optimal.

A standard technique using the vector Shannon lower bound on the kth order rate-distortion function $R_k(d)$ then gives for mean squared distortion

$$\limsup_{d\to 0}[\frac{1}{2}H(Q_{d,P}^k) - R_k(d)] \le \frac{1}{2}\log 2\pi e l(P). \tag{6}$$

The condition for (6) to hold is that $E\|X^k\|^2 < \infty$, $|h(f)| < \infty$, and $H(Q_{d,P}^k(X^k)) < \infty$ for some $d > 0$.

Theorem 2 *Suppose the conditions of Theorem 1 hold. Then the rate $\bar{H}(Q_{d,V}^k)$ of the randomized lattice quantizer with basic cell V and rth power distortion d satisfies*

$$\lim_{d\to 0} d\, 2^{\frac{2}{k}\bar{H}(Q_{d,V}^k)} = l(V)\, 2^{\frac{2}{k}h(f)}. \tag{7}$$

i.e., the asymptotic performance of the randomized lattice quantizer is the same as the asymptotic performance of the ordinary (non-randomized) lattice quantizer given by (4).

Corollary 1 *For $r = 2$, $|h(f)| < \infty$, and $E\|X^k\|^2 < \infty$,*

$$\limsup_{d\to 0} \frac{1}{k}\bar{H}(Q_{d,V}^k) - R_k(d) \le \frac{1}{2}\log 2\pi e l(V). \tag{8}$$

ACKNOWLEDGEMENTS

The research was supported in part by Hewlett-Packard Co., and the National Science Foundation under Grants No. NCR-90-09766 and NCR-91-57770.

References

[1] I. Csiszár, "Generalized entropy and quantization problems," in *Trans. of the Sixth Prague Conf. on Inf. Theory, Stat. Decision Functions, Random Processes*, pp. 29-35, 1973.

[2] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373-380, July 1979. performance

[3] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers", *IEEE Trans. Inform. Theory*, IT-38, No. 2, March 1992.

[4] J. Ziv, "On universal quantization," *IEEE Trans. Inform. Theory* vol. IT-31, pp. 344-347, May 1985.

Joint Source and Channel Coding Applied to the Pyramid Vector Quantizer

Michael J. Ruf
German Aerospace Research Establishment (DLR)
Institute for Communications Technology
D - 8031 Oberpfaffenhofen, Germany
Phone: + 49 - 8153 - 28864
Fax: + 49 - 8153 - 281442

Pavel Filip
German Aerospace Research Establishment (DLR)
Applied Data Systems Division
D - 8031 Oberpfaffenhofen, Germany
Phone: + 49 - 8153 - 281367
Fax: + 49 - 8153 - 281448

We consider the problem of transmitting images over noisy channels. Source coding is done by a the fixed rate scheme (DCT and Product Pyramid VQ with (L-K)-thresholding) [1], that allows us to calculate very precisely the expected mean square error caused by data compression. This together with the a-priori knowledge of the bit-sensitivity of the compressed image data enables us to erform a highly efficient equal or unequal error protection for image transmission over noisy channels.

Given the overall rate $R = R_S + R_C$, where R_S and R_C denotes the source rate and the channel rate respectively in bits per pixel, and the channel SNR E_s/N_0, one can calculate the optimal ratio of source to channel rate. First of all this requires the knowledge of mean=$\mathbb{E}(\rho(\vec{X}))$ and variance=$var(\rho(\vec{X}))$ of source vectors $\vec{X}$ for the rate R_S to calculate the mean square error caused by source coding mse_S [1]. Furthermore, we need the bit-sensitivity $\bar{A}_i$ of the compressed data. Therefore, it is sufficient to look at one of the coding units (CU) the whole image is divided into. These bit-sensitivities can be derived before images transmission, again with the knowledge of mean and variance of the source vectors $\vec{X}$.

For the given channel SNR, we can compute the bit error probability p_b of our channel code, a 64-state RCPC-code [2] and with the average contribution $\bar{A}_i$ of an error of bit i on any compressed CU, we obtain the mean square error caused by channel errors

$$mse_C = \sum_{i=0}^{s-1} p_{b_i} \cdot \bar{A}_i \ , \qquad A_i \in (\bar{A}_{j,dc}, \bar{A}_{pyr}, \bar{A}_{m,rad}),$$

where s denotes the number of bits of the compressed CU. Finally, we have to optimize the overall $mse = mse_S + mse_C$ for the given rate. Hereby, we either use an equal error protection (EEP) channel code for the whole data or even better, several levels of protection for the different sensitive bits (UEP).

All these a-priori calculations allow us to adapt the source and channel coding to the SNR so as to obtain the lowest *mse* for a fixed rate or to calculate the required rate for a given maximum *mse* before transmission. Simulation results (see figure (1) - unprotected, EEP, UEP) show the performance of our scheme and gains of up to 5 (8) dB (unprotected) in PSNR and up to 4 (8) dB in E_S/N_0 compared to JPEG (Joint Photographic Experts Group) especially for very noisy channels and low overall rates R.

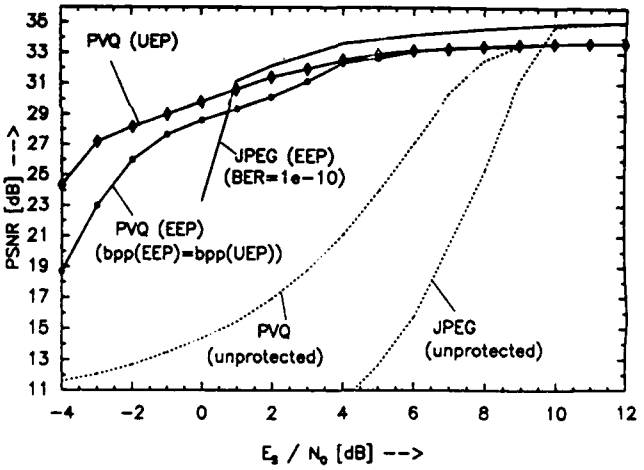


Figure 1: PSNR vs. channel E_S/N_0 of the PVQ and JPEG for LENA at 0.5 bits per pixel

References

[1] P. Filip and M. J. Ruf, "A fixed-rate product pyramid vector quantzation using a Bayesian model," in *Proc. of IEEE Globecom '92*, (Orlando, Fl.), pp. 08B.02.1-5, 1992.

[2] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. COM-36, pp. 389–400, Apr. 1988.

Finite-State Vector Quantization over Noisy Channels †

Yunus Hussain and Nariman Farvardin
Electrical Engineering Department
and Institute for Systems Research
University of Maryland
College Park, Maryland 20742

Summary

A finite-state vector quantizer (FSVQ) is a finite-state machine used for data compression. An FSVQ can be viewed as a collection of memoryless vector quantizers (VQ's) where each input vector is encoded using a VQ associated with the current encoder state; the current state and the codevector selected (or the corresponding index) determine the next encoder state [1].

Unlike an ordinary memoryless VQ, an FSVQ (with K states) can take advantage of the memory between successive source vectors by incorporating a feedback structure which enables it to choose the appropriate VQ (from the set of K different VQ's), given the past behavior [1]. While in the noiseless case this feedback structure leads to a performance gain over ordinary memoryless VQ, in the presence of channel noise, the same feedback structure renders FSVQ very sensitive to the channel error propagation and leads to severe degradation in the performance of the FSVQ as compared to the memoryless VQ. Indeed, FSVQ designed as in [1] falls apart in the presence of channel noise.

In this paper, we propose two modified FSVQ systems (NC-FSVQ1 and NC-FSVQ2) that are robust to channel noise. In order to design NC-FSVQ1, we first redesign the FSVQ system taking into account the channel noise, under the assumption that the decoder has perfect knowledge of the encoder state sequence. This leads to the design of an FSVQ system in which each state VQ is a channel-optimized VQ [2]. The transmission of "protected" encoder state indices is needed to allow the decoder to track the encoder state sequence. In order to keep the overhead information (consisting of encoder state indices) low, the encoder state index is transmitted periodically, say once for every n input vectors, and then given the state indices at times k and $k+n$ and the received codewords at times $i = k, k+1, \ldots, k+n-1$, the Viterbi algorithm, with the maximum likelihood criterion, is used to estimate the encoder state indices at times $i = k+1, k+2, \ldots, k+n-1$ at the decoder. Since some of the encoder state indices at times $i = k+1, k+2, \ldots, k+n-1$ may be incorrectly estimated by the decoder, we have developed an algorithm to do a judicious indexing of the codevectors among the states, so that if an encoder state is incorrectly decoded while the codeword is received correctly, the error introduced is not substantial (it should be noted that this indexing provides protection against state index error, while protection against error in the received codeword within a state is implicitly provided by the channel optimized state VQs). The resulting FSVQ system called NC-FSVQ1 performs significantly better than the ordinary FSVQ [1] in the presence of channel noise (memoryless binary symmetric channel assumed).

We have used NC-FSVQ1 to encode the Gauss-Markov (G-M) source with a correlation coefficient $\rho = 0.9$. When the channel is noisy, NC-FSVQ1 outperforms ordinary FSVQ significantly at all levels of channel noise. We observed that for a block size of 4, under noisy channel conditions, NC-FSVQ1 performed close to or better than the channel-optimized VQ (CO-VQ) [2]; at $\epsilon = 0.005 - 0.05$, the performances are close, while at $\epsilon = 0.1$, NC-FSVQ1 outperforms CO-VQ by 0.4-0.9 dB with a decoding delay of 6 vectors (ϵ is the bit error rate). The performance gain of NC-FSVQ1 over CO-VQ increases as ρ increases and for $\rho = 0.95$, the gain is more than 2 dB.

Although the NC-FSVQ1 offers robustness against channel noise, there are two problems associated with such a scheme. First, it suffers from a delay at the receiver side. The second problem relates to the overhead involved in explicitly transmitting the protected encoder state information; the next encoder state information is implicitly contained, at least partially, in the current transmitted codeword and it seems that protecting the codeword (or part of it) instead of the state (as is done in NC-FSVQ1) might lead to a similar performance (in terms of distortion) at a lower overhead rate. Let us now focus on the second issue. In a given FSVQ, the state information is embedded in the codeword in an unstraightforward way. In other words, we do not know which bits in the codeword should be protected in order to effectively protect the state information. In an effort to resolve the above issue, we modify the FSVQ design algorithm [3] such that all the state information is forced to be contained in $l = \log_2 K$ most significant bits of the codeword. In addition, we modify the next-state function such that the state at time $n+1$ depends only on the codeword at time n (independent of the state at time n, see [3]). Under these conditions, using the development in [4], we have formulated the joint source-channel coding problem for the modified FSVQ system, developed necessary conditions of optimality and based on these conditions described a design algorithm leading to the so-called NC-FSVQ2 system [3].

We also used NC-FSVQ2 to encode the G-M source.Under noisy channel conditions, NC-FSVQ2 performs significantly better than ordinary FSVQ and as compared to CO-VQ, it achieves a gain of 0.4-0.8 dB at $\epsilon = 0.005$ and 0.7-1.0 dB at $\epsilon = 0.1$. Again, this gain increases as ρ increases. In contrast to NC-FSVQ1, NC-FSVQ2 does not have any delay at the decoder and there is no need for a separate channel code. We also used NC-FSVQ1 and NC-FSVQ2 to encode the speech LSP parameters [5] and achieved noticeable gains over ordinary FSVQ and CO-VQ under noisy channel conditions [3].

References

[1] J. Foster, R.M. Gray and M.O. Dunham, "Finite-State Vector Quantization for Waveform Coding," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 348-359, May 1985.

[2] N. Farvardin and V. Vaishampayan,"On the Performance and Complexity of Channel-Optimized Vector Quantizers," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 155-160, January 1991.

[3] Y. Hussain, *Design and Performance Evaluation of a Class of Finite-State Vector Quantizers*, Ph.D. Dissertation, Electrical Engineering Department, Univ. of Maryland, College Park, MD, 1992.

[4] J.G. Dunham and R. M. Gray, "Joint Source and Noisy Channel Trellis Encoding," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 516-519, Jul. 1981.

[5] N. Sugamura and F. Itakura, "Speech Data Compression by LSP Speech Analysis-Synthesis Technique," *IECE Trans.*, vol. J64-A, No. 8, pp. 599-605, Aug. 1981 (in Japanese).

†This work was supported in part by National Science Foundation grants NSFD MIP-86-57311 and NSFD CDR-85-00108, and in part by grants from NTT Corporation and General Electric.

Average Number of Facets per Cell in Tree-Structured Vector Quantizer Partitions

Kenneth Zeger
Coordinated Science Laboratory
University of Illinois
1308 W. Main St., Urbana, IL 61801

and

Miriam R. Kantorovitz
Department of Mathematics
University of Illinois
Urbana, IL 61801

ABSTRACT

Upper and lower bounds are derived for the average number of facets per cell in the encoder partition of binary Tree-Structured Vector Quantizers. The achievability of the bounds is described as well. It is shown in particular that the average number of facets per cell for unbalanced trees must lie asymptotically between 3 and 4 in $\mathcal{R}^2$, and each of these bounds can be achieved, whereas for higher dimensions it is shown that an arbitrarily large percentage of the cells can each have a linear number (in codebook size) of facets. Analogous results are also indicated for balanced trees.

SUMMARY

A binary *Tree-Structured Vector Quantizer (TSVQ)* Q can formally be defined recursively by *cutting* (or *splitting*) one cell of an existing TSVQ by a hyperplane. As in general VQ, TSVQ's also partition $\mathcal{R}^d$ into a finite set of convex polytopal cells. This follows from the fact that every encoding region is a finite intersection of half-spaces. It will be assumed throughout that the intersection of any cell-splitting hyperplane with a face of the split cell is of lower dimension then that of the face itself or equivalently that a general position restriction holds.

A *facet* of a convex polytope in $\mathcal{R}^d$ is any $(d-1)$-dimensional face of the polytope. Two cells in a quantizer partition are *neighbors* if each has a distinct facet, one of which is a subset of the other. Equivalently, two cells are neighbors if the intersection of their closures has dimension $d-1$. For a VQ encoder partition in general position, there is a one-to-one correspondence between the facets of a cell and the cell's neighbors. However, for TSVQ, it is possible that one cell could be adjacent to several other cells via the same facet; in general, the number of facets per cell is less than or equal to the number of neighbors of the cell. Often, however, these two quantities are very similar or equal. For a given convex polytopal partition Ω of $\mathcal{R}^d$ into n cells, define

1) $F_d(n)$ = average number of facets per cell in Ω.
2) $G_d(n) = nF_d(n)$
3) $M_d(n)$ = maximum number of facets of a cell in Ω.

Note that since every cell of any vector quantizer with n codevectors cannot share more than one facet with any other cell we obtain the trivial upper bound $F_d(n) \leq n-1$. In two dimensions, a straightforward application of Euler's theorem for planar graphs shows that $F_2(n) \leq 6$ (i.e. not restricted to TSVQ).

In this paper we derive several bounds on $F_d(n)$ for TSVQ and point out the achievability of these. Specifically, it is shown that for 2-dimensional unbalanced TSVQ, the average number of facets per cell is asymptotically bounded above by 4 and below by 3, and that the bounds are achievable. For higher dimensional spaces an upper bound of $n/2$ and a lower bound of 3 are given. It is also shown that $n/4$ and 3 respectively are achievable in this case. At present, it is an open question as to whether the $n/2$ bound is achievable. In $\mathcal{R}$, it is trivially always the case that $F_1(n) = 2 - 2/n$.

In $\mathcal{R}^2$, it is shown that if an asymptotically large fraction of the TSVQ cells are bounded, then $F_2(n) \approx 4$. This would lend some support to the assumption made in [1] that $F_d(n) = 2d$ for the case $d = 2$. However, for $d > 2$, this might not be the case. It is shown analogously that for balanced TSVQ with $d > 2$ the upper bound on the average number of facets per cell is reduced to $\log_2 n$. It should be emphasized, though, that the achievability of the bounds presented are best and worst cases, over the class of all TSVQ's, and it is a question for future study as to how likely they are to occur for various practical TSVQ systems.

Proposition 1 *For unbalanced TSVQ, the average number of facets per cell satisfies*

$$\begin{aligned} 3 - 4/n &\leq F_d(n) \leq n/2 - 1/2 \quad \text{for} \quad d > 2, n \geq 1 \\ 3 - 4/n &\leq F_d(n) \leq 4 - 7/n \quad \text{for} \quad d = 2, n \geq 3. \end{aligned} \tag{1}$$

The next several results exhibit the bounds' achievability.

Proposition 2 *For every $d > 2$ and $n > 1$ there exists an unbalanced TSVQ such that $F_d(n) \geq n/4$.*

Proposition 3 *For $d = 2$ and every $n > 2$ there exists an unbalanced TSVQ such that $F_d(n) = 4 - 7/n$.*

Proposition 4 *For every $d \geq 2$ and $n > 1$ there exists an unbalanced TSVQ such that $F_d(n) = 3 - 4/n$.*

The following corollary shows that there exist d-dimensional TSVQ's such that an arbitrarily large fraction of the cells each have a linear number (in codebook size) of facets.

Corollary 1 *For every $d > 2$, $n \geq 1$, and $\alpha \in (0,1)$, there exists a TSVQ with n cells such that at least αn of the cells each have at least $(1-\alpha)n$ facets.*

For balanced trees similar results are obtained, though with a reduction from linear to logarithmic bounds. The results are stated in terms of the number of cells n, in the TSVQ, though it should be remembered that balanced trees only exist when n is some integer power of 2. In the following proposition, the achievability of the lower bound for $d \geq 2$ and the upper bound for $d = 2$ are analogous to the unbalanced case. However, it is unknown at present whether the upper bound $\log_2 n$ is achievable; in fact it is unknown whether, for a fixed $d > 2$, it is possible to exhibit balanced TSVQ's such that $F_d(n)$ is unbounded.

Corollary 2 *For balanced TSVQ,*

$$\begin{aligned} 3 - 4/n &\leq F_d(n) \leq \log_2 n \quad \text{for} \quad d > 2, n > 0 \\ 3 - 4/n &\leq F_d(n) \leq 4 - 8/n \quad \text{for} \quad d = 2, n > 0. \end{aligned} \tag{2}$$

ACKNOWLEDGEMENTS The research was supported in part by Hewlett-Packard Co., and the National Science Foundation under Grants No. NCR-90-09766 and NCR-91-57770.

References

[1] D. Neuhoff and D. Lee, "On the Performance of Tree-Structured Vector Quantization", *IEEE Int. Symp. Info. Theory (ISIT)*, Budapest, Hungary, June 1991.

A METHOD FOR EXAMINING VECTOR QUANTIZER STRUCTURES

Erik Agrell
Department of Information Theory
Chalmers University of Technology
S–412 96 Göteborg, Sweden

***Abstract* — This paper presents how to study the geometry of Voronoi regions in an arbitrary vector quantizer. Methods to find the location, the extent, and the neighbors of any region are summarized. Application to fast encoding is emphasized.**

I. INTRODUCTION

It is well known that a vector quantizer (VQ) performs better, in terms of signal-to-noise ratio, than a scalar quantizer [4]. The improvement increases with the dimension, but the price paid is complexity. In particular, the encoding process is slower. In the case of *nearest neighbor* quantization, which this paper considers, the straightforward encoding method is calculating the squared Euclidean distance

$$d(\mathbf{w},\mathbf{r}_i)=\|\mathbf{w}-\mathbf{r}_i\|^2 \tag{1}$$

between an input vector $\mathbf{w}$ and every reconstruction vector $\mathbf{r}_i$; $i=1,\ldots,n$, and selecting the codeword that gives the minimum distance. The set of vectors that are encoded as a certain codeword k according to this rule is called the *Voronoi region* (VR)

$$V_k=\{\mathbf{w}: d(\mathbf{w},\mathbf{r}_k)\le d(\mathbf{w},\mathbf{r}_i);\ i=1,\ldots,n\} \tag{2}$$

Sometimes suboptimal VQs are accepted in order to decrease the encoding time. Several structures have been developed for which fast search algorithms exist, e.g. lattice or multistage coders. However, there are also methods to improve the encoding speed for arbitrary VQs, without paying with signal-to-noise ratio. The methods often require precomputing some geometrical properties of the VRs. A new method to obtain such information is presented here, as well as an encoding algorithm based on the precomputed data.

II. EXAMINING THE GEOMETRY OF VORONOI REGIONS

Some relevant types of problems concerning the structure of given VRs are:

1. What values of a certain component may vectors in this VR take on?
2. On which side of a certain hyperplane lies this VR, or is the VR intersected?
3. Have these two VRs a common face?

The three questions are closely related. All of them have applications in different algorithms for the design of fast encoders, see below. Probabilistic methods have been proposed to obtain approximate, or likely, answers to them [2], [3]. In this section, deterministic methods, based on different applications of *linear programming*, are presented for solving these and related problems reliably.

Consider the following standard formulation of a linear programming problem:

$$\min \mathbf{c}^T\mathbf{x} \quad \text{when} \begin{cases} A\mathbf{x}=\mathbf{b} \\ \mathbf{x}\ge 0 \end{cases} \tag{3}$$

Much research and much literature have been devoted to methods for solving it. Two of the main approaches are the *simplex method* and *Karmarkar's algorithm*, both having numerous variations [1]. From optimization theory it is known that there exists a *dual* problem to (3),

$$\max \mathbf{b}^T\mathbf{w} \quad \text{when } A^T\mathbf{w}\le\mathbf{c} \tag{4}$$

the solution of which is strongly connected to the solution of (3). Actually, both mentioned methods generate the solution of (4) as a by-product when solving (3).

The inequality constraints in (4) form a convex polytope. They describe the VR V_k (2) of a certain codeword k if

$$A=\left[\mathbf{a}_1\ldots\mathbf{a}_{k-1}\,\mathbf{a}_{k+1}\ldots\mathbf{a}_n\right] \qquad \mathbf{c}=\left[c_1\ldots c_{k-1}\,c_{k+1}\ldots c_n\right] \tag{5a}$$

where for every i

$$\mathbf{a}_i=\mathbf{r}_i-\mathbf{r}_k \qquad c_i=\frac{\|\mathbf{r}_i\|^2-\|\mathbf{r}_k\|^2}{2} \tag{5b}$$

Thus, the dual problem can be used for finding extrema of a VR.

Depending on the choice of the dual objective vector $\mathbf{b}$, different extrema will be found and different properties of the VR will be investigated. Especially, if $\mathbf{b}$ is chosen first as a unit vector and next as the same vector negated, problem 1 above is solved by two linear programs. If this is repeated for all coordinates and all VRs, a circumscribed hyperrectangle will be found for each VR, which is the precomputed information required for encoding with the *Projection Method* [2].

Problem 2 is solved similarly by two linear programs, if $\mathbf{b}$ is set orthogonal to the given hyperplane, pointing in both directions. If the two extrema lie on the same side of the plane, so does the whole VR; otherwise it is intersected. The answer to this kind of questions is vital for the design of the decision tree used in the *Binary Hyperplane Testing Algorithm* [3].

To test the neighborship between VRs j and k (problem 3), $\mathbf{b}$ is chosen equal to $\mathbf{a}_j$, which is orthogonal to the common face of V_j and V_k, if such a face exists, whereas A and $\mathbf{c}$ as before denote V_k (5). With this input, a linear programming algorithm will return the point $\hat{\mathbf{w}}$ in V_k whose projection on $\mathbf{a}_j$ is closest to $\mathbf{r}_j$. If the two VRs have a common face, the dual optimum $\hat{\mathbf{w}}$ will inevitably lie on it. The primal optimum shows whether this has occurred: the face was reached if and only if the component of $\hat{\mathbf{x}}$ corresponding to $\mathbf{a}_j$ is greater than zero.

A VR is defined by $n-1$ linear inequalities as in (2) or (4). Some of them are in general redundant. Define the set N_k of neighbors to a codeword k as all codewords whose VRs have a face in common with V_k. The corresponding inequalities are the only ones needed to be considered in order to determine if a vector $\mathbf{w}$ belongs to a certain VR V_k:

$$V_k=\{\mathbf{w}: \mathbf{a}_i^T\mathbf{w}\le\mathbf{c}_i;\ i\in N_k\} \tag{6}$$

Now, solving problem 3 for all pairs of VRs in a VQ generates the complete neighborship table N_i; $i=1,\ldots,n$, which is a useful tool for analysis as well as in applications. Because the description (6) defines exactly the same region as (2), but more economically, the table speeds up other geometrical studies, such as the solution of problems 1 and 2. In addition, it makes a new approach to fast encoding possible, called *neighbor descent*.

III. THE "NEIGHBOR DESCENT" ENCODING METHOD

Suppose that a vector $\mathbf{w}$ is to be encoded and that there is reason to believe that $\mathbf{r}_k$ is a good reconstruction vector for $\mathbf{w}$. Calculate the distortions for all the neighbors of k, that is, the distances $d(\mathbf{w},\mathbf{r}_i)$; $i\in N_k$. Replace k with the neighbor that has the smallest distortion and restart. If no codeword in N_k is better than k itself, then stop.

Theorem of uniqueness: In any VQ, for any input $\mathbf{w}$, no more than one codeword can have a smaller distortion than all its neighbors.

A necessity for the success of the method is that a path through neighboring VRs, along which the distortion $d(\mathbf{w},\mathbf{r}_k)$ is monotonic decreasing, does not terminate in a suboptimal local minimum. The above theorem states that this can never be the case. Its proof follows as a consequence of (6) and the observation that a vector cannot belong to the interior of more than one VR.

The performance of the neighbor descent method was evaluated in experiments on VQs without an induced structure. The results show that most of the n distance calculations can be avoided with the neighbor descent method. The reduction is greatest for VQs with high bit rates, or, if the rate is kept constant, in many dimensions.

REFERENCES

[1] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*. New York: Wiley, 1990.
[2] D.-Y. Cheng, A. Gersho, B. Ramamurthi, and Y. Shoham, "Fast Search Algorithms for Vector Quantization and Pattern Matching," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, 1984, pp. 9.11.1–4.
[3] D.-Y. Cheng and A. Gersho, "A Fast Codebook Search Algorithm for Nearest-Neighbor Pattern Matching," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, 1986, pp. 265–268.
[4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

AN OPTIMAL DATA COMPRESSION CODE FOR MEMORYLESS GAUSSIAN SOURCE

Hiroki Koga and Suguru Arimoto
Faculty of Engineering, The University of Tokyo
7 - 3 - 1, Hongo, Bunkyo-ku, Tokyo 113, Japan

Abstract: An asymptotically optimal code with a fidelity criterion for any memoryless gaussian source is proposed. It is shown that the proposed code achieves the rate-distortion bound with probability $1-\varepsilon$ for any given distortion level under the squared-error criterion. Asymptotical behaviors of the code performance are also analyzed in detail.

Introduction

To develop data compression schemes with a fidelity criterion is essential especially for continuous sources. It is therefore important to devise a general method to encode an output sequence of practical analogue sources.

In this manuscript, any memoryless gaussian source is considered, but it is sufficient to treat one with zero mean and unit variance. Let $\boldsymbol{x} = x_1 \cdots x_n$ be an n-tuple of source symbols that satisfies $x_i \sim N(0,1^2)$ for all $i = 1, \cdots, n$. The probability density function can be written as

$$p(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} \exp\Big[-\frac{1}{2}(x_1^2 + x_2^2 + \cdots + x_n^2)\Big], \tag{1}$$

and each $\boldsymbol{x}$ can be treated as an element of $\boldsymbol{R}^n$. An original word $\boldsymbol{x} = x_1 \cdots x_n$ is encoded into a reproduction word $\hat{\boldsymbol{x}} = \hat{x}_1 \cdots \hat{x}_n$ by a fixed-to-fixed length code. The distortion between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ is defined by

$$d(\boldsymbol{x}, \hat{\boldsymbol{x}}) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{x}_k)^2.$$

The rate-distortion function of the memoryless gaussian source with zero mean and unit variance under squared-error criterion is of the form

$$R(\Delta) = \frac{1}{2}\log_2\frac{1}{\Delta}, \qquad \Delta \in (0,1] \tag{2}$$

(See [1].) In the following section, it is shown that an asymptotically optimal code which achieves (2) can be generated with probability $1-\varepsilon$ for any given distortion level $\Delta \in (0,1]$.

Main Results

Since the probability density function (1) is sphere-symmetric for any word $\boldsymbol{x}$ of length n, it is natural to have an idea of encoding $\boldsymbol{x}$ by two separated steps, i.e., 1) to quantize the magnitude $\|\boldsymbol{x}\|$ and 2) quantize the shape $\tilde{\boldsymbol{x}} \stackrel{\text{def}}{=} \boldsymbol{x}/\|\boldsymbol{x}\|$, where $\|\cdot\|$ denotes the Euclidean norm. First, select an arbitrary set $\mathcal{A} = \{a_1, \cdots, a_L\}$ satisfying the following four conditions:

C1) All elements must satisfy $0 = a_1 < a_2 < \cdots < a_L < \infty$,
C2) a_L must satisfy $a_L \geq \sqrt{n}$,
C3) $\zeta \stackrel{\text{def}}{=} \max_{1\leq l<L}(a_{l+1} - a_l)$ must satisfy $\lim_{n\to\infty}(\zeta^2/n) = 0$,
C4) L must satisfy $\lim_{n\to\infty}[(\log_2 L)/n] = 0$.

Let $\mathcal{Y} = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M\}$ be a set satisfying $\|\boldsymbol{y}_m\| = 1$ for all $m = 1, \cdots, M$, that is, all the elements of $\mathcal{Y}$ belong to the n-dimensional unit hypersphere S^{n-1}. Encoding is defined as a mapping $\varphi : \boldsymbol{R}^n \to \mathcal{A} \times \mathcal{Y}$ specified by φ_1 and φ_2 such as

$$\varphi(\boldsymbol{x}) = \varphi_1(\boldsymbol{x}) \cdot \varphi_2(\boldsymbol{x}), \tag{3}$$

where $\varphi_1 : \boldsymbol{R}^n \to \mathcal{A}$ and $\varphi_2 : \boldsymbol{R}^n \to \mathcal{Y}$ are two mappings defined as follows:

$$\varphi_1(\boldsymbol{x}) = a_l \quad \text{if} \quad a_l \leq \|\boldsymbol{x}\| \leq a_{l+1},$$
$$\varphi_2(\boldsymbol{x}) = \boldsymbol{y}_m \quad \text{if} \quad \|\tilde{\boldsymbol{x}} - \boldsymbol{y}_m\| = \min_{\boldsymbol{y}\in\mathcal{Y}}\|\tilde{\boldsymbol{x}} - \boldsymbol{y}\|.$$

The rate R and the average distortion $\bar{\Delta}$ of this code are given by

$$R = \frac{1}{n}\log_2 LM, \tag{4}$$

$$\bar{\Delta} = \int_{\boldsymbol{R}^n} p(\boldsymbol{x}) d(\boldsymbol{x}, \varphi(\boldsymbol{x}))\, d\boldsymbol{x}, \tag{5}$$

respectively. The following theorem evaluates the average distortion caused by the proposed code.

Theorem Let $\Delta \in (0,1]$ be arbitrarily given. Let $M = M(n,\Delta) = \lceil \pi |S^{n-1}| / (|S^{n-2}| \Delta^{\frac{n}{2}-1}) \rceil$, and select an arbitrary set $\mathcal{A}$ satisfying C1) $\sim$ C4). If M points on S^{n-1} are chosen independently as elements of $\mathcal{Y}$, then for any $\delta > 0$ and $\varepsilon > 0$ there exists an integer $n_0 = n_0(\delta, \varepsilon)$ that satisfies following two relations:

$$\begin{aligned} R &< R(\Delta) + \delta \\ E[\bar{\Delta}] &< \Delta + \varepsilon \end{aligned} \tag{6}$$

for all $n \geq n_0$, where $E[\bar{\Delta}]$ denotes the expectation of $\bar{\Delta}$ with respect to the choice of $\mathcal{Y}$. Moreover, for any $\varepsilon' > 0$ there exists an integer $n_1 = n_1(\varepsilon')$ that satisfies

$$V[\bar{\Delta}] < \varepsilon' \tag{7}$$

for all $n \geq n_1$, where $V[\bar{\Delta}]$ denotes the variance of $\bar{\Delta}$ with respect to the choice of $\mathcal{Y}$.

For any given $\Delta \in (0,1]$, the asymptotical behavior of this code is characterized in the following way:

$$R = R(\Delta) + \mathcal{O}\Big(\frac{1}{n}\log_2 L\Big), \tag{8}$$

$$\bar{\Delta} = \Delta + \mathcal{O}\Big(\frac{\zeta^2}{n}\Big). \tag{9}$$

It is easy to construct $\mathcal{A}$ that satisfies C1) $\sim$ C4). For instance, by setting $L = n+1$ and $a_l = (l-1)/\sqrt{n}$ $(l = 1, \cdots, L)$, it is easy to check that this example satisfies C1) $\sim$ C4) with $\zeta = 1/\sqrt{n}$. Hence, R converges to $R(\Delta)$ of order $\mathcal{O}(\log_2 n/n)$ and $\bar{\Delta}$ converges to Δ of order $\mathcal{O}(1/n^2)$. The paper [2] indicates a conjecture that the convergence of $\mathcal{O}(\log_2 n/n)$ in R and $\mathcal{O}(1/n)$ in $\bar{\Delta}$ would be the tightest possible. Our result in (8) and (9), however, not only reveals that there exists a code that has a better asymptotical behavior, but also represents a more general trade-off relationship between the rate and the average distortion.

References

[1] T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, 1971.
[2] D. J. Sakrison, "A Geometric Treatment of the Source Encoding of a Gaussian Random Variable," IEEE Trans. on Inform. Theory, vol. IT-14, No. 3, pp. 481 - 486, 1968.

A GENERAL THEORY OF INFORMATION TRANSFER

Rudolf Ahlswede
Fakultät für Mathematik
Universität Bielefeld
Postfach 8640
D-4800 Bielefeld 1
GERMANY

Summary

We present a unified theory of information, which naturally incorporates Shannon's theory of information transmission and the theory of identification in the presence of noise as extremal cases. It provides several novel coding theorems.

ON THE PROBABILITY OF UNDETECTED ERROR FOR ITERATED CODES*

Toshihisa NISHIJIMA† Osamu NAGATA‡ Shigeichi HIRASAWA§

Abstract

In practical applications of coding theory, e.g., ARQ system, Signature analysis, etc., the probability of undetected error, $P_u(\varepsilon)$, for linear block codes plays an improtant part. The exact value of $P_u(\varepsilon)$ can be calculated by using the weight distribution. However for almost all codes, the weight distribution is not exactly known. Hence, some methods calculating the upper bound of $P_u(\varepsilon)$ have been proposed. It has been known by deriving the average probability of undetected error for the ensemble of all (n,k,d)[1] linear codes, that there exists (n,k,d) linear code having $P_u(\varepsilon)$ upper bounded by $2^{-(1-r)n}$, where $r = k/n$[1]. Some codes, e.g., Hamming codes, satisfying $P_u(\varepsilon) < 2^{-(1-r)n}$ have been found out for finite code length n[3]. In these codes, $r \to 1$, and $P_u(\varepsilon)$ does not converge to zero for $n \to \infty$. The codes with $P_u(\varepsilon)$ converging exponentially to zero with order n for $n \to \infty$ is not also explicitly constructed. Of course, if $0 < r < 1$, and the asymptotic distance ratio $\delta, \delta > 0$, for $n \to \infty$, it is trivially shown that $P_u(\varepsilon)$ converges to zero. For example, Justesen codes[4] satisfy the above conditions. However, if $\delta = 0$, it is hardly to shown that $P_u(\varepsilon) \to 0$ as $n \to \infty$.

In this paper, it is shown from the theoretical viewpoint that under the code rate $R_O, 0 < R_O < 1$, iterated codes with $P_u(\varepsilon)$ converging to zero can be explicitly constructed, although the asymptotic distance ratio $\Delta_O, \Delta_O = 0$, for the code length $N_O, N_O \to \infty$. It is also shown that there exist $P_u(\varepsilon)$ of these explicitly constructed codes converging exponentially to zero with order N_O, for $N_O \to \infty$.

Throughout this paper, we assume that codes are the binary linear block codes, and channel, the binary symmetric channel with cross-over probability $p, 0 < p < 1/2$.

1. CODING AND DECODING METHODS

A. Coding Method

Let $\otimes$ be direct product. Then (N_O, K_O, D_O) iterated codes $C_O^{(s)}$ are constructed by $c_1 \otimes c_2 \otimes \cdots \otimes c_s$, where c_i is the i-th stage (n_i, k_i, d_i) code, i=1,2,...,s.

B. Decoding Method

Let $G_i(X), Y_i(X)$, and $S_i(X)$, be the generator polynomial of the i-th stage code c_i, the polynomial of the subsequence with length n_i of the recieved sequence at the *step* i, and the polynomial of the syndrome of the subsequence with length n_i at the *step* i, respectively. Then, decoding method of *step* i is as follows.

Step i : After partitioning the subsequence with length $n_s n_{s-1} \cdots n_i k_{i-1} \cdots k_1$ of the recieved sequence with length N_O into $n_s n_{s-1} \cdots n_{i+1} k_{i-1} \cdots k_1$ sequences with length n_i, $Y_i(X)$ of partitioned each sequence is sucssesivelly divided by $G_i(X)$. If $S_i(X) \neq 0$, a error is detected. If all of $n_s n_{s-1} \cdots n_{i+1} k_{i-1} \cdots k_1 S_i(X)$ are zeros, go to the *step* i+1.

2. THE PROBABILITY OF UNDETECTED ERROR

Lemma1 : The upper bound of the probability of undetected error, $P_u^{(s)}(\varepsilon)$ for iterated codes $C_O^{(s)}$ is given by

$$P_u^{(s)}(\varepsilon) < [\prod_{i=1}^{s} \{\sum_{j=1}^{n_i} A_{i,j} p^j (1-p)^{n_i - j}\}]^{k_s k_{s-1} \cdots k_1}, \quad (1)$$

where $A_{i,j}$ is the number of codedwords of Hamming weight j for the code c_i.

Lemma 2 : The sufficient condition to construct iterated codes C_O whose code rate $R_O, 0 < R_O < 1$ for $N_O \to \infty$, that is, $s \to \infty$, is $r_i < r_{i+1}$, where $r_i = k_i/n_i$, and $i = 1, 2, \ldots, \infty$.

Lemma 3 : The asymptotic distance ratio Δ_O of iterated codes C_O satisffying lemma 2 is

$$\Delta_O = \lim_{s \to \infty} (D_O/N_O) = 0. \quad (2)$$

From lemmas 1, 2, and 3, we have the following theorem.

Theorem 1 : The probability of undetected error, $P_u(\varepsilon)$ for iterated codes C_O is given by

$$P_u(\varepsilon) = \lim_{s \to \infty} P_u^{(s)}(\varepsilon) = 0, \quad (3)$$

where $0 < R_O < 1$, and $\Delta_O = 0$ as $N_O \to \infty$.

3. SOME EXAMPLES

Example 1 : Iterated codes $C_O^{(s)}$ constructed by applying the i-th stage code c_i with $(2^{m+i-1}, 2^{m+i-1} - (m+i), 4)$extended Hamming code, where m=2,3,.... Note that these iterated codes are error-free codes proposed by P. Elias[5].

Example 2 : Iterated codes $C_O^{(s)}$ constructed by applying the i-th stage code c_i with $(2^{m+i-1}, 2^{m+i-1} - 1, 2)$ even parity check code, where m=1,2,....

Example 3 : Iterated codes $C_O^{(s)}$ constructed by applying the i-th stage code c_i with $(2^{m+i-1} - 1, 2^{m+i-1} - (m+i+1), 3)$Hamming code, where m=2,3,.... Note that $P_u(\varepsilon)$ of these iterated codes satisfy $P_u(\varepsilon) < 2^{(1-R_O)N_O}$.

References

[1] S. Lin and D. J. Costello, Jr., *Error Control Coding : Fundamentals and Applications*. Englewood Cliffs, NJ : JPrentice-Hall, 1983.

[2] J. K. Wolf, A. M. Michelson, and A. H. Levesque,"On the probabiity of undetected error for linear block codes," *IEEE Trans. Commun.*, vol. COM-30, pp. 317-324, 1982.

[3] J.Justesen,"A class of constructive asymptotically good algebraic codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 652-656, Sept. 1972.

[4] P. Elias, "Error-free coding," *IRE Trans.* PGITE-4, pp. 29-37, 1954.

*The research leading to this paper was partially supported by the Minisutry of Education under Grant-in-Aids 04750364 for scientific Research, and by Waseda University Grant for Special Research Project No. 92A-188.

†Kanagawa Institute of Technology, JAPAN

‡Sony Corporation, JAPAN

§Waseda University, JAPAN

†The (n,k,d) code denotes the code of length n, the number of information symbols k, and minimum distance d.

ON THE KEY EQUATION FOR n-DIMENSIONAL CYCLIC CODES

Hervé CHABANNE [1] and Graham H. NORTON [2]

Abstract

Let $R = K[X_1,\ldots,X_n]/(X_1^{N_1}-1,\ldots,X_n^{N_n}-1)$ be a semisimple algebra. Ideals in R are known as n-dimensional cyclic codes or abelian codes.
Let F be the smallest extension of K containing an N_v^{th} primitive root of unity α_v for $v = 1,\ldots,n$.
Let $e \in \mathsf{F}[X_1,\ldots,X_n]$ be a non-zero polynomial.
We consider the series in $\mathsf{F}[[X_1^{-1},\ldots,X_n^{-1}]]$
$\Gamma_{\tilde{e}}(X_1^{-1},\ldots,X_n^{-1}) = \sum_{i_1}\ldots\sum_{i_n} e(\alpha_1^{-i_1},\ldots,\alpha_v^{-i_n})X_1^{i_1}\ldots X_n^{i_n}$.

We first introduce univariate polynomials $\sigma_v \in \mathsf{F}[X_v]$, $v = 1,\ldots,n$ and a multivariate polynomial $\omega \in \mathsf{F}[X_1,\ldots,X_n]$ such that $\begin{cases} \sigma_1\ldots\sigma_n\Gamma_{\tilde{e}} &= X_1\ldots X_n\omega \\ gcd(\omega,\sigma_1\ldots\sigma_n) &= 1 \end{cases}$
Thus, we show that the spectral behaviour of $\sigma = \sigma_1\ldots\sigma_n$ and ω allows us to recover e.

In the second part, we reinterpret the polynomials σ and ω, regarding $\Gamma_{\tilde{e}}$ as the generating function of the n-dimensional linear recurring sequence $\tilde{e} = \left(e(\alpha_1^{-i_1},\ldots,\alpha_n^{-i_n})\right)$.
Then we show how to obtain σ_v.

Hence, we deduce a new method for decoding abelian codes.

Notation

$\mathbf{X}$	:=	$X_1\ldots X_n$
$\mathsf{F}[\mathbf{X}]$	:=	$\mathsf{F}[X_1,X_2,\ldots,X_n]$.
$\mathsf{F}((\mathbf{X}^{-1}))$	:=	Laurent series in $\mathbf{X}^{-1}$ over F.
$\mathbf{i}$	:=	$(i_1,\ldots,i_n)$
$\pi_v(\mathbf{i})$	:=	i_v.
$\widehat{\pi_v}(\mathbf{i})$	:=	$(i_1,\ldots,i_{v-1},i_{v+1},\ldots,i_n)$
$\mathbf{i} \preceq \mathbf{k}$	$\Leftrightarrow$	$i_v \le k_v$, $v \in [1,n]$.
$\mathrm{supp}(\sum_{\mathbf{i}} p_{\mathbf{i}}\mathbf{X}^{\mathbf{i}})$	:=	$\{\mathbf{i} : p_{\mathbf{i}} \neq 0 \in \mathsf{F}\}$.
$\delta_v(p)$	:=	The degree of $p \in (\mathsf{F}[\frac{\mathbf{X}}{X_v}])[X_v]$.
$e(\alpha^{\mathbf{i}})$	:=	$e(\alpha_1^{i_1},\ldots,\alpha_n^{i_n})$
$\delta(p)$	:=	$(\delta_1(p),\ldots,\delta_n(p))$

1 The key equation.

Let $e \in \mathsf{F}[\mathbf{X}]$ be a non-zero polynomial. Our goal is to show how the series $\Gamma_{\tilde{e}}(\mathbf{X}^{-1}) = \sum_{\mathbf{i}\succeq\mathbf{0}} e(\alpha^{-\mathbf{i}})\mathbf{X}^{\mathbf{i}} \in \mathsf{F}[[\mathbf{X}^{-1}]]$ may be written as a quotient of two relatively prime polynomials.

Let Ξ_e be the smallest cartesian product containing $\mathrm{supp}(e)$.

Definition *For $v \in [1,n]$, define the* **error-locator X_v-polynomial** *by $\sigma_v(X_v) = \prod_{i_v\in\pi_v(\mathrm{supp}(e))}(X_v - \alpha_v^{i_v}) \in \mathsf{F}[X_v]$. We call $\sigma = \sigma(\mathbf{X}) = \prod_{v=1}^n \sigma_v(X_v)$ the* **error-locator product polynomial** *of e. Finally, we call*

$$\omega = \omega(\mathbf{X}) = \sum_{\mathbf{i}\in\mathrm{supp}(e)} e_{\mathbf{i}}\{\prod_{v=1}^n \prod_{j\in\pi_v(\mathrm{supp}(e)),\ j\neq i_v}(X_v - \alpha_v^j)\}$$

the **error-evaluator polynomial** *of e.*

[1] INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153, Le Chesnay Cedex, FRANCE

[2] Centre for Communication Research, Faculty of Engineering, Univ. of Bristol, UNITED KINGDOM (Research supported by Science and Engineering Research Council grant GR/H15141)

We are now ready to state the key equation.

Theorem 1 *In $\mathsf{F}((\mathbf{X}^{-1}))$, we have $\sigma\Gamma_{\tilde{e}} = \mathbf{X}\omega$, $gcd(\sigma,\omega) = 1$.*

Remark : We can consider Theorem 1 as generalization of the key equation for BCH codes to n-dimensional cyclic codes. As for the cyclic case, the spectral behaviour of ω and σ allow us to recover e :
- σ equals 0 only on Ξ_e,
- ω equals 0 only on $\Xi_e \setminus \mathrm{supp}(e)$.

2 The linear recurring sequence context.

As in [2], we let $S^n_{\preceq\mathbf{0}}(\mathsf{F})$ denote the commutative F-algebra of $(-\mathbf{N})^n$-indexed sequences $s : (-\mathbf{N})^n \to \mathsf{F}$.

The **generating function** of s is $\Gamma_s(\mathbf{X}^{-1}) = \sum_{\mathbf{i}\preceq\mathbf{0}} s_{\mathbf{i}}\mathbf{X}^{\mathbf{i}}$.

The **characteristic ideal** of $s \in S^n_{\preceq\mathbf{0}}(\mathsf{F})$ is
$Ann(s) = \{\sum_{\mathbf{i}\in\mathrm{supp}(f)} f_{\mathbf{i}}\mathbf{X}^{\mathbf{i}} : \forall \mathbf{j} \in (-\mathbf{N})^n,\ \sum_{\mathbf{i}\in\mathrm{supp}(f)} f_{\mathbf{i}}s_{\mathbf{j}-\mathbf{i}} = 0\}$.

If $Ann(s) \neq \{0\}$ then s is called a **linear recurring sequence**.

In section 1 we studied the sequence $\tilde{e}$ given by $\tilde{e}_{\mathbf{i}} = e(\alpha^{-\mathbf{i}})$.

Clearly $\tilde{e}$ is linear recurring sequence and $X_v^{N_v} - 1$ belongs to $Ann(\tilde{e}) \cap \mathsf{F}[X_v]$, $v \in [1,n]$. In fact, we can say more :

Theorem 2 *For $v \in [1,n]$, σ_v is the monic polynomial of minimal (positive) degree in $Ann(\tilde{e}) \cap \mathsf{F}[X_v]$.*

We conclude this section by recalling how σ_v may be computed.

Theorem 3 *[2] Let $f_v \in Ann(\tilde{e}) \cap \mathsf{F}[X_v]$, $\delta_v f_v \ge 1$, for $v \in [1,n]$. Put $\mathbf{d_v} = \delta(\prod_{u=1,u\neq v}^n f_u)$. For $\mathbf{i} \in (-\mathbf{N})^{n-1}$ define the $\mathbf{i}$-section of $\tilde{e}$, $\tilde{e}_{\mathbf{i}}^{\S(v)} \in S^1_{\preceq 0}(\mathsf{F})$ by $(\tilde{e}_{\mathbf{i}}^{\S(v)})_j = \tilde{e}_{\mathbf{k}}$, where $\widehat{\pi_v}(\mathbf{k}) = \mathbf{i}$ and $\pi_v(\mathbf{k}) = j$.*
Then $\sigma_v = lcm\{g_{\mathbf{i}}^{\S(v)} : (g_{\mathbf{i}}^{\S(v)}) = Ann(\tilde{e}_{\mathbf{i}}^{\S(v)}), \mathbf{0} \preceq \mathbf{i} \preceq \mathbf{d_v} - \mathbf{1}\}$.

3 Decoding algorithm.

If C is an ideal of R, we know that

$$c \in C \Leftrightarrow c(\alpha_1^{i_1},\ldots,\alpha_n^{i_n}) = 0,\ \forall (i_1,\ldots,i_n) \in Z_C$$

where Z_C is a set which only depends on C [1].

From $m = c+e$, $c \in C$, we can get the parts of $\Gamma_{\tilde{e}}$ corresponding to Z_C.

This yields the following algorithm

(1) From the known terms of $\tilde{e}_{\mathbf{i}}^{\S(v)}$ find
- $g_{\mathbf{i}}^{\S(v)}$ such that $(g_{\mathbf{i}}^{\S(v)}) = Ann(\tilde{e}_{\mathbf{i}}^{\S(v)})$ (used in point 2),
- the missing terms of $\tilde{e}_{\mathbf{i}}^{\S(v)}$ (used in point 3).

(2) For $v \in [1,n]$, put $\sigma_v = lcm(g_{\mathbf{i}}^{\S(v)})$ (Theo. 3), and $\sigma = \prod_{v=1}^n \sigma_v$

(3) Put $\omega = \dfrac{\sigma\Gamma_{\tilde{e}}}{X_1\ldots X_n}$ (Theo. 1)

(4) From σ and ω deduce e (Remark on p. 1)

References

[1] R. Blahut " *Theory and Practice of Error Control Codes* ", Reading, MA : Addison-Wesley, 1983

[2] G. H. Norton " *On n-dimensional sequences, II. Characteristic Ideals* ", submitted to J. Symbolic Computation

ON THE EQUIVALENCE OF SOME GENERALIZED CONCATENATED CODES AND EXTENDED CYCLIC CODES

B.Liesenfeld and B.G.Dorsch
Institut für Netzwerk- und Signaltheorie, Technical University Darmstadt
Merckstraße 25, D–6100 Darmstadt, Germany

ABSTRACT

It is known that cyclic codes of composite length are equivalent to generalized concatenated codes with inner cyclic codes and outer cyclic or constacyclic codes. Here it is shown that there is a large class of extended cyclic codes of length Q^m, that can be constructed by generalized concatenation of shorter extended cyclic codes. This class includes the generalized Reed-Muller codes and the Euclidean geometry codes. In many cases a simple multistage decoder corrects all errors of weight less than half of the true minimum distance of the code.

SUMMARY

We start with some notations. Let $(a_{i_1,i_2,\ldots,i_m})$ denote an $Q \times Q \times \cdots \times Q$-array over $GF(q)$, where $Q = q^s$ for some integer $s > 0$. We associate a polynomial in m variables

$$A(x_1, x_2, \ldots, x_m) = \sum_{j_1=0}^{Q-1} \sum_{j_2=0}^{Q-1} \cdots \sum_{j_m=0}^{Q-1} A_{j_1,j_2,\ldots,j_m} x_1^{j_1} x_2^{j_2} \cdots x_m^{j_m}$$

with each array $(a_{i_1,i_2,\ldots,i_m})$. Suppose $\alpha_0, \alpha_1, \ldots, \alpha_{Q-1}$ are the elementes of $GF(Q)$. Then the associated polynomial is uniquely determined by the following equation:[1]

$$a_{i_1,i_2,\ldots,i_m} = \frac{1}{(Q-1)^m} A(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_m}) \qquad 0 \le i_1, i_2, \ldots, i_m < Q$$

Definition 1 *Let $\mathcal{J}$ be a subset of $\{(j_1, j_2, \ldots, j_m) | 0 \le j_1, j_2, \ldots, j_m < Q\}$. An array $(a_{i_1,i_2,\ldots,i_m})$ over $GF(q)$ is a codeword in the code $C(q, s, m, \mathcal{J})$, iff its associated polynomial has coefficients $A_{j_1,j_2,\ldots,j_m}$ equal to zero for each $(j_1, j_2, \ldots, j_m) \in \mathcal{J}$. The set $\mathcal{J}$ will be called the zero set of code $C(q, s, m, \mathcal{J})$.*

Considering the conjugacy constraints we get the complete zero set $\hat{\mathcal{J}}$:

$$\hat{\mathcal{J}} = \{(r_Q(q^\mu j_1), r_Q(q^\mu j_2), \ldots, r_Q(q^\mu j_m)) \,|\, (j_1, j_2, \ldots, j_m) \in \mathcal{J}, 0 \le \mu < s\},$$

where $r_Q(l)$ is given by

$$r_Q(l) = \begin{cases} \text{the number in } \{0, 1, \ldots, Q-2\}, \text{ which is congruent to} \\ \quad l \text{ modulo } Q-1, & \text{if } l \not\equiv 0 \bmod Q-1, \\ 0, & \text{if } l = 0, \\ Q-1, & \text{if } l \equiv 0 \bmod Q-1 \text{ and } l \neq 0. \end{cases}$$

If $(a_{i_1,i_2,\ldots,i_m})$ is a codeword in the code $C(q, s, m, \mathcal{J})$, all coefficients $A_{j_1,j_2,\ldots,j_m}$, $(j_1, j_2, \ldots, j_m) \in \hat{\mathcal{J}}$, of its associated polynomial are equal to zero. The number of information symbols K can be calculated by the formula:

$$K = Q^m - |\hat{\mathcal{J}}|.$$

It can be shown that, if the zero set $\mathcal{J}$ satisfies some condition, the dual code of a $C(q, s, m, \mathcal{J})$ code also belongs to the class of codes defined above. All $C(q, s, m, \mathcal{J})$ codes with $m > 1$ can be constructed by generalized concatenation of extended cyclic codes of length $Q = q^s$. If $m > 2$, the outer codes of the generalized concatenated code (GC code) again are GC codes.

It is readily seen for generalized Reed-Muller codes and it can be proved for Euclidean geometry codes that both classes belong to the $C(q, s, m, \mathcal{J})$ codes. We want to show that there are other extended cyclic codes which are equivalent to $C(q, s, m, \mathcal{J})$ codes.

Let $\{\xi_1, \xi_2, \ldots, \xi_m\}$ be a basis of $GF(Q^m)$ over $GF(Q)$. Then every element in $GF(Q^m)$ can be represented by a sum $x_1\xi_1 + x_2\xi_2 + \cdots + x_m\xi_m$, where $x_1, x_2, \ldots, x_m \in GF(Q)$. We denote the index sets $\mathcal{L}(h)$ and $\mathcal{L}^*(h)$:

$$\mathcal{L}(h) = \{(j_1, j_2, \ldots, j_m) \,|\, 0 \le j_1, j_2, \ldots, j_m < Q,\ j_1 + j_2 + \ldots + j_m > h\},$$
$$\mathcal{L}^*(h) = \{j_1 + j_2 Q + \ldots + j_m Q^{m-1} \,|\, (j_1, j_2, \ldots, j_m) \in \mathcal{L}(h)\}.$$

[1] We define $0^0 = 1$ and $0^j = 0$, if $j > 0$. Furthermore $(Q-1) = \sum_{i=0}^{Q-2} 1$ in $GF(Q)$.

Theorem 1 *Suppose $s = 1$ and $A(x_1, x_2, \ldots, x_m)$ is the polynomial associated with an array $(a_{i_1,i_2,\ldots,i_m})$ over $GF(q) = GF(q^s) = GF(Q)$. Let $B(x)$ be a polynomial over $GF(Q^m)$ with degree less than Q^m such that for all $x_1, x_2, \ldots, x_m \in GF(Q)$*

$$\frac{1}{Q^m - 1} B(x_1\xi_1 + x_2\xi_2 + \cdots + x_m\xi_m) = \frac{1}{(Q-1)^m} A(x_1, x_2, \ldots, x_m).$$

Then:

$$\forall (j_1, j_2, \ldots, j_m) \in \mathcal{L}(h) : A_{j_1,j_2,\ldots,j_m} = 0 \Leftrightarrow \forall j \in \mathcal{L}^*(h) : B_j = 0$$

Corollary 2 *The codes $C(q, sm, 1, \mathcal{L}^*(h))$ and $C(q, s, m, \mathcal{L}(h))$ are equivalent.*

Clearly, if two codes are equivalent, their dual codes are also equivalent.

Since the $C(q, s, m, \mathcal{J})$ codes, $m > 1$, are GC codes, a multistage decoder can be used and it corrects all errors of weight less than half of d_{GC}. d_{GC} denotes the well known lower bound (see for example [3], p.591) for the minimum distance of GC codes.

In table 1 some extended cyclic codes of length 64 are tabulated, which are the dual codes of some codes $C(q, sm, 1, \mathcal{L}^*(h))$. They can be constructed by generalized concatenation of codes of length 4 and 8. The number k of information symbols, the minimum distance d, the distance bound d_{GC}, the length q^s of the codes, which are used in the generalized concatenation, and the exponents of the roots of the generator polynomial are given. Especially interesting is the extended cyclic (64, 28, 16) code, since a modified multistage decoder can correct all errors of weight less than half of the true minimum distance.

k	d	d_{GC}	q^s	exponents of the roots	remark
48	6	6	4	15,27,31	EG-Code
45	8	8	8	15,23,31	BCH-Code
37	10	10	8	7,15,21,23,31	EG-Code
34	12	12	8	7,15,21,23,27,31	
28	16	14	8	7,13,15,21,23,27,31	
24	16	16	4	7,11,13,15,23,27,31	BCH-Code
13	24	24	4	3,7,9,11,13,15,21,23,27,31	EG-Code
10	28	28	8	3,5,7,11,13,15,21,23,27,31	BCH-Code

Table 1: Some binary extended cyclic codes of length 64 which are equivalent to GC codes

REFERENCES

[1] T. Kasami, S. Lin, W.W.Peterson, "Polynomial codes," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 807-814, Nov.1968.

[2] E.L.Blokh and V.V.Zyablov, "Coding of generalised concatenated codes," *Probl. Inform. Trans.*, vol. 10, no.3, pp. 218-222, 1974.

[3] F.J.McWilliams and N.J.A.Sloane, *The Theory of Error Correcting Codes*. Amsterdam, the Netherlands: North Holland, 1977.

[4] J.M.Jensen, "The concatenated structure of cyclic and abelian codes," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 788-793, Nov.1985.

[5] G.D.Forney, Jr., "Coset codes – Part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1152-1187, Sep.1988.

[6] J.M.Jensen, "Cyclic concatenated codes with constacyclic outer codes," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 950-959, May 1992.

ON CYCLIC PRODUCT CODES

B.S.Rajan
Department of Electrical Engineering
Indian Institute of Technology, Delhi
Hauz Khas, N.Delhi 110 016, India
email:bsrajan@ee.iitd.ernet.in

H.S.Madhusudhana and M.U.Siddiqi
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Kanpur 208 016, India

ABSTRACT

It is known that when the blocklengths of two cyclic codes are relatively prime their product code is cyclic when serialized using Chinese Remainder Theorem. When the blocklengths are equal we characterize product codes that are cyclic when serialized either rowwise or columnwise.

SUMMARY

Given two codes C_1 and C_2, their product code C is the code whose codewords are the two dimensional arrays for which columns are codewords in C_1 and rows are codewords in C_2. The component codes C_1 and C_2 may be cyclic but the product code is not necessarily cyclic. If it is cyclic, with appropriate serialization, then it is called cyclic product code. It is known that when the blocklengths n_1 and n_2 of C_1 and C_2 are relatively prime then C is cyclic when serialized using Chinese Remainder Theorem [1]. We identify cyclic product codes when block lengths of component cyclic codes are equal.

Consider a two dimensional $n_0 \times n_1$ array with entries from GF(q).

$$\begin{matrix} \alpha_{0,0} & \alpha_{0,1} & \cdots & \alpha_{0,n_1-1} \\ \alpha_{1,0} & \alpha_{1,1} & \cdots & \alpha_{1,n_1-1} \\ \cdot & \cdot & \cdots & \cdot \\ \alpha_{n_0-1,0} & \alpha_{n_0-1,1} & \cdots & \alpha_{n_0-1,n_1-1} \end{matrix}$$

By associating this array to an element $\sum_{i_1=0}^{n_1-1}\sum_{i_0=0}^{n_0-1}\alpha_{<i_0,i_1>}g_{n_0}^{i_0}g_{n_1}^{i_1}$ of the group algebra KG, where K=GF(q) and G is an Abelian group of order $n = n_0 n_1$, (n,q) = 1, which is a direct product of two cyclic subgroups of order n_0 and n_1 with generators g_{n_0} and g_{n_1}, *one can interpret Abelian codes which are ideals of KG as product codes.* With this interpretation the problem reduces to identifying the Abelian codes which are closed under cyclic shifts. This class of Abelian codes are same as separable Abelian codes [2].

Abelian codes are characterized in the transform domain as follows: Define a mapping I from $\Omega = \{0, 1, \ldots, n-1\}$ to $\Omega_0 \otimes \Omega_1$, where $\Omega_0 = \{0, 1, .., n_0 - 1\}$ and $\Omega_1 = \{0, 1, .., n_1 - 1\}$, by $I(i) = (i_0, i_1)$ where $i = i_0 + i_1 n_0 \quad \forall \quad i \in \Omega, i_0 \in \Omega_0, i_1 \in \Omega_1$. (i_0, i_1) is called the mixed-radix representation of i and the mapping I is called the mixed-radix serialization (MRS). (MRS corresponds to serializing the product code columnwise.) For a given (i_0, i_1) the subset $C_{(i_0,i_1)} = \{(i_0,i_1), 2(i_0,i_1), 2^2(i_0,i_1), \ldots, 2^{e-1}(i_0,i_1)\}$ where for any integer α, $\alpha(i_0,i_1)$ is defined to be $(\alpha i_0 (\mathrm{mod}\ n_0), \alpha i_1 (\mathrm{mod}\ n_1))$ and $2^e(i_0,i_1) = (i_0,i_1)$, is called the conjugacy class containing (i_0, i_1). The appropriate Discrete Fourier Transform (DFT) is given by

$$A_{<j_0,j_1>} = \sum_{i_0=0}^{n_0-1}\sum_{i_1=0}^{n_1-1}\alpha_{n_0}^{i_0 j_0}\alpha_{n_1}^{i_1 j_1}\alpha_{<i_0,i_1>}$$

where α_{n_0} and α_{n_1} are elements of orders n_0 and n_1 in the extension field. An Abelian code can be defined as the set of n-tuples whose DFT coefficients are zero in specified conjugacy classes [3].

For the case when $n_0 = n_1$ the expressions for inverse DFT respectively for Abelian and cyclic case are

$$\alpha_{<i_0,i_1>} = \sum_{j_0=0}^{n_0-1}\sum_{j_1=0}^{n_0-1}\alpha_{n_0}^{-(i_0 j_0 + i_1 j_1)}A_{<j_0,j_1>} \quad \text{and}$$

$$\alpha_i = \sum_{j_0=0}^{n_0-1}\sum_{j_1=0}^{n_0-1}\alpha^{-(i_0+i_1 n_0)(j_0+j_1 n_0)}A_{j_0+j_1 n_0} \quad \text{(using MRS).}$$

We can take $\alpha_{n_0} = \alpha^{n_0}$, by working in the same extension field for both cases. Now, starting with the transform vector of idempotent generator of a cyclic code in both the cases and then replacing the transform vector corresponding to the shift $< k_0, k_1 >$ in the Abelian case and k cyclic shifts in the cyclic case we can find conditions on j_0 and j_1 for which left hand side of both the expressions are same for all i_0, i_1, k_0 and k_1. This leads to

Theorem 1: A product code of two cyclic codes of equal length n is cyclic when serialized columnwise (rowwise) iff the Abelian code obtained by serializing rowwise (columnwise) has idempotent generator whose DFT vector is same as that of an idempotent generator of a cyclic code of length n^2.

Extending the above approach to r (r>2) dimensional product codes the following theorem can be proved.

Theorem 2: When codewords are written as monomials of the form

$$\sum_{i_0}\sum_{i_1}\cdots\sum_{i_{r-1}}\alpha_{<i_0,i_1,\ldots,i_{r-1}>}x_0^{i_0}x_1^{i_1}\cdots x_{r-1}^{i_{r-1}}$$

then the idempotent generators of cyclic product codes (under MRS) are the direct sums of the generators given by

$$\sum_{i_0\in\Omega}\sum_{i_1\in\Omega}\cdots\sum_{i_{k-1}\in\Omega}\sum_{i_k\in C_l}x_0^{i_0}x_1^{i_1}\cdots x_{k-1}^{i_{k-1}}x_k^{i_k} \quad 0 \le k \le r-1, \text{ for some } l \ne 0.$$

or equivalently [4],

$$\left(\sum_{i_0\in\Omega}x_0^{i_0}\right)\left(\sum_{i_1\in\Omega}x_1^{i_1}\right)\cdots\left(\sum_{i_{k-1}\in\Omega}x_{k-1}^{i_{k-1}}\right)\left(\sum_{i_k\in C_l}x_k^{i_k}\right) \text{ for some } l \ne 0.$$

REFERENCES

[1]. Burton H.O and E.J.Weldon Jr., "Cyclic product codes", IEEE Trans. Inform. Theory, Vol.IT-11, pp.433-439, 1965.

[2]. P.Camion, On Abelian Codes, MRC Technical Report, No.1059, The University of Wisconsin, 1971.

[3]. B.Sundar Rajan and M.U.Siddiqi, "Spectral Characterization of Abelian codes", to appear in IEEE Trans. Inform. Theory, Vol.38, No.6, 1992.

[4]. H.S.Madhusudhana, "Abelian codes which are closed under cyclic shifts", M.Tech thesis, I.I.T.Kanpur, India, 1987.

ALGEBRAIC STRUCTURE AND DECODING OF TWO-DIMENSIONAL CASCADE CODES

Keith Saints
Center for Applied Mathematics, Cornell University, Ithaca NY 14853
Chris Heegard
School of Electrical Engineering, Cornell University, Ithaca NY 14853

This paper discusses the algebraic structure of cascaded Reed-Solomon (CRS) codes, and presents an algorithm for decoding them. A CRS code is a cascade (or "generalized concatenated") code constructed using Reed-Solomon codes as component codes. In particular, we consider hyperbolic CRS (HCRS) codes: these are CRS codes designed to have the minimum distance given by the cascade code bound. Compared to Reed-Solomon codes over the same alphabet, HCRS codes have longer block-lengths. Compared to other two-dimensional cyclic codes (products of Reed-Solomon codes, duals of such products, and codes proposed by Sakata [1]) with the same minimum distance, HCRS codes have higher rates.

Consider the finite field $\mathbf{F}_q$ with q elements, and let $n = q - 1$, so that there is an element α of $\mathbf{F}_q$ which is a primitive n^{th} root of unity. We define $\mathbf{F}_q^{n\times n}[x,y] = \mathbf{F}_q[x,y]/(x^n - 1, y^n - 1)$. Then for each value of d we define the set of parity-check points:

$$Z_d = \left\{(\alpha^i, \alpha^j) \in \mathbf{Z}^2 : (i+1)(j+1) < d\right\}.$$

Finally, we define a code HCRS_d consisting of codewords f which vanish at each parity-check point: $\mathrm{HCRS}_d = \{f \in \mathbf{F}_q^{n\times n}[x,y] : f(x,y) = 0 \text{ for all } (x,y) \in Z_d\}$. The minimum distance of HCRS_d is the value d given by the cascade-code bound [2-4].

We present two encoding schemes for CRS codes: a transform-based frequency-domain encoder, and a systematic time-domain encoder which makes use of a Gröbner basis for the code.

We now describe a decoding algorithm which corrects t errors for the code HCRS_{2t+1}. The decoder receives a corrupted version $c(x,y) + e(x,y)$ of the codeword $c(x,y)$, and deduces the error polynomial $e(x,y)$ by determining its Fourier transform, E. Initially the decoder knows only the entries of E corresponding to syndromes, and calculates the remaining entries by finding two-dimensional linear recursion relations (2DR) which hold on all of E. The set of all 2DR relations valid on the error transform array form an ideal called the error locator ideal because its zeros determine the error locations. The algorithm recursively iterates through the entries of E according to the pure lexicographic order, maintaining a set G of 2DR relations known to be valid on processed part of the error transform array. Upon termination, G is a Gröbner basis for the error locator ideal. For each entry of E that is a syndrome, the algorithm performs a validation step, and for unknown entries of E the algorithm performs a calculation-validation step. The validation step is the same as the main step in Sakata's algorithm [1]: each 2DR of G is checked for validity at this entry, and replaced if it proves to be invalid. In the calculation-validation step, the algorithm first calculates the entry of E, then performs a validation step. Each relation in G predicts a value for the unknown entry. We show that only one of these predicted values is consistent with an error pattern $e(x,y)$ of weight t or less. Moreover, an incorrect prediction is detected immediately in the subsequent validation step, so the entry is effectively calculated by trying each of the predictions in turn until there is no inconsistency.

REFERENCES

[1] S. Sakata, "Decoding binary 2-D cyclic codes by the 2-D Berlekamp–Massey algorithm," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1200–1203, July 1991.

[2] E. L. Blokh and V. V. Zyablov, "Coding of generalized concatenated codes," *Probl. Info. Trans.*, vol. 10, pp. 45–50, 1974.

[3] J. Wu and D. J. Costello Jr., "New multi-level codes over GF(q)," *IEEE Trans. Inform. Theory*, vol. 38, pp. 933–939, May 1992.

[4] R. Krishnamoorthy and C. Heegard, "Structure and decoding of Reed-Solomon based cascade codes," in *Proc. 25th Ann. Conf. Inform. Sci. Syst.*, pp. 29–33, 1991.

This work was supported in part by NSF grants NCR-8903931 and NCR-9207331, and in part by an AASERT grant from the U.S. Army Research Office administered through the Mathematical Sciences Institute of Cornell University.

The Polynomial of correctable patterns of concatenated codes

Nicolas Sendrier [1]

Abstract

The polynomial of correctable patterns is defined in [1] as the weight enumerator of the set of error patterns correctable by a given decoding algorithm. The polynomials of uncorrectable and miscorrected patterns can be defined as well ([5] and [6]).

These polynomials allow a compact representation of a decoding algorithm which is sufficient to compute the correction probability and the miscorrection probability through a memoryless symetric channel. These results are generalised in [5] and [6] for erasure channels.

Our purpose here is to compute the polynomials of correctable patterns of concatenated codes for different decoding algorithms.

- We give the weight distribution of the error patterns correctable by the standard decoding algorithm.
- We give bounds for the weight distribution of the error patterns correctable by Block-Zyablov algorithm.

This new method for evaluating concatenated codes will thus provide an efficient way to evaluate the standard algorithm. It will also give a way to evaluate with precision the performances of the Block-Zyablov decoding algorithm which needed, up to now, a (much more expensive) simulation.

As an example, we will compute the decoding performances of the concatenation of the Nordstrom-Robinson ([4, p. 73]) inner code, and a Reed-Solomon $(255, 223, 33)$ outer code over $\mathbf{F}_{256}$.

1 Polynomials of correctable patterns

We denote by $\mathbf{F}$ the finite field $\mathbf{F}_q$. Let $C(n,k,d)$ be a linear code over $\mathbf{F}$ of length n, dimension k and minimum distance d.

We consider a transmission channel where error and erasure may occur simultaneously, we represent an erasure by the symbol ∞ and we denote by $\hat{\mathbf{F}}$ the set $\mathbf{F} \cup \{\infty\}$. Let γ be a decoding algorithm for C for such a channel. The erasure weight $\rho_H(y)$ of an element y of $\hat{\mathbf{F}}^n$ is the number of its component equal to ∞, and its error weight $\nu_H(y)$ is equal the number of its components different from 0 and ∞. We call extended weight enumerator of $E \subset \hat{\mathbf{F}}^n$ the polynomial $\sum_{y \in E} X^{\nu_H(y)} Y^{\rho_H(y)} Z^{n-\nu_H(y)-\rho_H(y)}$.

The **polynomials of correctable, uncorrectable** and **miscorrected patterns** of γ [1, 5] are respectively the extended weight enumerators of $\{y \in \hat{\mathbf{F}}^n, \gamma(y) = 0\}$, $\{y \in \hat{\mathbf{F}}^n, \gamma(y) \text{ fails}\}$ and $\{y \in \hat{\mathbf{F}}^n, \gamma(y) \in C \setminus \{0\}\}$. They are denoted $P_0(X,Y,Z)$, $P_1(X,Y,Z)$ and $P_2(X,Y,Z)$.

If γ corrects errors alone these definitions apply with $Y = 0$.

Relationship with the error probability: if a codeword of C is transmitted through a q-ary symetric erasure channel with transition probability p_t and erasure probability p_∞ and then decoded by γ, the probability of correction is:

$$P_{corr} = P_0(p, p_\infty, 1 - (q-1)p - p_\infty).$$

Similar statements hold for failure probability and miscorrection probability and the polynomials $P_1(X,Y,Z)$ and $P_2(X,Y,Z)$.

[1] INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay CEDEX, FRANCE

2 Concatenated codes

Let $B(n,k,d)$ be a linear (inner) code over $\mathbf{F}_q$ and let $C(N,K,D)$ be a linear (outer) code over $\mathbf{F}_{q^k}$. Let $t = (d-1)/2$.

The concatenated code of B and C denoted $B \square C$ is the set of all the codewords of C whose components are replaced by codewords of B. It is a linear code over $\mathbf{F}_q$ code of length nN, of dimension kK and minimum distance $\geq dD$ [3]. A codeword can be seen as a succession of N inner codewords.

Let γ be a error correcting algorithm for B and let $P_0(X,Z)$, $P_1(X,Z)$ and $P_2(X,Z)$ be its polynomials of correctable, uncorrectable and miscorrected patterns. Let Φ be an error and erasure correcting algorithm for C and let $Q_0(X,Y,Z)$, $Q_1(X,Y,Z)$ and $Q_2(X,Y,Z)$ be its polynomials of correctable, uncorrectable and miscorrected patterns.

The **standard decoding algorithm** Γ consists in decoding all N inner codewords, then the outer codeword. Roughly we can denote it $\Gamma = \Phi \circ \gamma^N$.

Its polynomial of correctable patterns is equal to

$$Q_0(P_2(X,Y), P_1(X,Y), P_0(X,Y)).$$

The **Block-Zyablov algorithm** Ψ will use $t+1$ inner decoder, for $0 \leq i \leq t$, γ_i will only decode error patterns of weight $\leq i$ [2]. Roughly we can define $\Psi(y) = \text{best_of}_{0 \leq i \leq t}(\Gamma_i(y))$, where $\Gamma_i = \Phi \circ \gamma_i^N$.

Its polynomial of correctable patterns verifies

$$R_0(X,Y) \preceq \sum_{\substack{(s_0,\ldots,s_t) \in \mathbf{N}^{t+1} \\ \min_{0 \leq j \leq t}(s_j) < D}} [\prod_{j=0}^{t} Y_j^{s_j}] Q(X, Y_0, \ldots, Y_t), \qquad (1)$$

with $Q(X, Y_0, \ldots, Y_t) =$

$$\left(\sum_{i=0}^{t} P_{0,i}(X,Y) \prod_{j=0}^{i-1} y_j + P_1(X,Y) \prod_{j=0}^{t} Y_j + \sum_{i=0}^{t} P_{2,i}(X,Y) \prod_{j=0}^{i-1} Y_j \prod_{j=i}^{t} Y_j^2 \right)^N$$

where $P_{0,i}(X,Y)$ is the monomial of degree i of $P_0(X,Y)$ and $P_{2,i}(X,Y)$ is the weight enumerator of the miscorrected error patterns $y \in \mathbf{F}_q^n$ such that $d_H(y, \gamma(y)) = i$.

Notations for (1):

$[\prod_{j=0}^{t} Y_j^{s_j}] Q(\ldots)$ is the coefficient of $Y_0^{s_0} \ldots Y_t^{s_t}$ in $Q(X, \ldots)$.

$P(X,Y) \preceq Q(X,Y) \Leftrightarrow \forall i,j,\ [X^i Y^j] P(X,Y) \leq [X^i Y^j] Q(X,Y)$

References

[1] P. Camion and J-L. Politano. Evaluation of a coding design for a very noisy chanel. In *Coding Theory and Applications*. Springer Verlag, LNCS n° 388, 1988.

[2] T. Ericson. Concatenated codes - Principles and possibilities. In *AAECC-4*, 1986.

[3] G.D. Jr. Forney. *Concatenated Codes*. The M.I.T. Press, Cambridge, Massachussets, 1966.

[4] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error Correcting Codes*. North-Holland, 1977.

[5] N. Sendrier. *Codes Correcteurs d'Erreurs à Haut Pouvoir de Correction*. Thèse de doctorat, Université Paris 6, December 1991.

[6] N. Sendrier. The polynomial of error patterns. preprint, 1992.

Constructive codes for arbitrary DMC and the AGNC

Michael Steiner
Naval Research Laboratory

Summary

Shannon showed that arbitrarily low error (ALE) can be achieved on a channel when the information rate is less than the channel *capacity*. Two problems emerge in this respect. The first is concerned with the development of codes which can achieve ALE, which we refer to as an error-free code[1]. The second is concerned with the design of practical bandlimited efficient systems which perform with low probability of error while retaining a decoder which is of reasonable complexity. It was shown by Feinstein for discrete memoryless channels (DMC) that the average probability of error, $\overline{P_e}$, can be bounded exponentially in the length of the code, n, for rates r less than capacity. For such channels we define a good code as one which has an average probability of error less than a quantity $e^{-nE(r)}$, where $E(r)$ is a function of r and greater than 0. Note that an error-free code is not necessarily good, although a good code is error-free.

In 1953 Fano showed that orthogonal signaling could be used to assure a good coding system on the infinite bandwidth Gaussian channel for rates less than capacity. In 1954 Elias[1] presented an error-free coding system for the binary symmetric channel (BSC). In 1966 Schalkwijk and Kailath[2] presented a constructive coding scheme for the additive Gaussian noise channel (AGNC) when feedback is allowed for rates less than capacity. Justesen[3] in 1972 found constructive concatenated codes for the BSC in which $\lim \inf(d/n) > 0$, d being the minimum distance of the code. It can be shown that this also implies the codes are good codes in the sense defined above. In 1982 Delsarte and Piret[4] generalized Justesen's construction to demonstrate a good coding system for rates less than capacity for a class of channels called regular DMC, which include the class of symmetric DMC. In this paper we further generalize the Delsarte and Piret construction and prove that it is good for arbitrary DMC for rates less than capacity. Other channels for which the construction is good are discussed. The AGNC without feedback is examined. It is shown how to construct a good code which will signal with any rate less than $\frac{1}{2}\log(1+\frac{\mathcal{E}}{\sigma^2})$, where the average input power S is constrained $S \leq \mathcal{E}$ and σ^2 is the average noise power.

References

[1] P. Elias. "Error free coding," *IEEE Trans. Inform. Theory*, **IT-4**, pp. 29-37, 1954.

[2] J.P.M. Schalkwijk. "A coding scheme for additive noise channels with feedback-Part II: Band-Limited signals," *IEEE Trans. Inform. Theory*, **IT-12**, pp 183-189, April 1966.

[3] J. Justesen. "A class of constructive asymptotically good algebraic codes," *IEEE Trans. Inform. Theory*, **IT-19**, pp. 711-713, Sept. 1973.

[4] P. Delsarte & P. Piret. "Algebraic constructions of Shannon codes for regular channels," *IEEE Transactions on Inform. Theory*, **IT-28**, No. 4, pp 593-599, July 1982.

POLYPHASE PSEUDO-NOISE SEQUENCES WITH EQUIVALENT EVEN AND ODD CORRELATION PROPERTIES

Ryuji KOHNO, Hidenobu FUKUMASA and Hideki IMAI
Division of Electrical and Computer Engineering
Faculty of Engineering
Yokohama National University
156 Tokiwadai, Hodogaya-ku, Yokohama JAPAN 240
Phone:+81-45-338-1177, Fax:+81-45-338-1157, Email:kohno@kohnolab.dnj.ynu.ac.jp

Abstract: This paper proposes and investigates a method of designing pseudo-noise (PN) sequences having equivalently good properties of both even and odd correlations, i.e. EOE sequences, which are important for acquisition and demodulation in spread-spectrum (SS) communications. Odd correlation property of PN sequences should be designed as well as their even or periodic correlation property so that interference of data-modulated PN sequences can be reduced as small as possible. We describe a method of designing polyphase EOE sequences from biphase PN sequences with a good aperiodic correlation property under a certain condition. Absolute values of both even and odd correlation functions of these sequences at each shift can be equal. We evaluate properties of the derived sequences by peaks of crosscorrelations and out-of-phase autocorrelations and BER. It is shown that the polyphase sequences have lower peaks of crosscorrelations and out of phase autocorrelations and lower BER than the biphase ones by numerical evaluation. Furthermore, the generalized odd correlation function which is important in polyphase SS systems is defined, and a method to improve the generalized odd correlation properties are derived.

Introduction

In asynchronous code division multiple access (CDMA) based on spread spectrum techniques, spreading sequences should have low crosscorrelation to reduce co-channel interference and increase user capacity, as well as sharp autocorrelation to accomplish reliable acquisition. Moreover, the odd correlation function represents the output of the correlator in the case where the data symbol changes during the integration of the correlation operation while the even correlation function represents the output in the case where the data symbol remains constant over two consecutive symbols. Since the even and odd correlation functions appear with equal probability, both functions are of equal importance.

This paper proposes and investigates a method of designing pseudo-noise (PN) sequences having equivalently good properties of both even and odd correlations. Odd correlation property of PN sequences should be designed as well as the even or periodic correlation property so that interference of data-modulated PN sequences can be reduced as small as possible.

Definition of EOE Sequences

We define EOE sequecne as follows.

Let EOE sequecnes be the sequences, x and y, having even and odd (auto and cross) correlation functions whose absolute values at each shift are equal. That is,

$$|\theta(x)(l)| = |\theta(x)(l)|,$$
$$|\theta(y)(l)| = |\theta(y)(l)|,$$
$$|\theta(x,y)(l)| = |\theta(x,y)(l)|,$$

for every $l \in \{0, 1, \cdots, N-1\}$, where $\theta(x,y)(l)$, $\theta(x,y)(l)$ and $C(x,y)(l)$ are the even, odd and aperidic crosscorrelation functions of sequences, x and y, of period N respectively. They are defined by

$$\theta(x,y)(l) = C(x,y)(l) + C(x,y)(l-N), \tag{1}$$

$$\theta(x,y)(l) = C(x,y)(l) - C(x,y)(l-N), \tag{2}$$

$$C(x,y)(l) = \begin{cases} \sum_{n=0}^{N-l-1} x_n y^*_{n+l}, & 0 \le l < N, \\ \sum_{n=0}^{N+l-1} x_{n-l} y^*_n, & 1-N \le l < 0, \\ 0, & |l| \ge N. \end{cases} \tag{3}$$

A Method to Generate EOE Sequences

Polyphase PN sequences with such good correlation properties can be derived from biphase PN sequences with a good aperiodic correlation property under a certain condition. We propose a method of designing Equivalent Odd and Even correlation (EOE) sequences.

Method 1 *Let u and v be arbitrary real valued sequences of period N. Then, the complex valued sequences, x and y, given by*

$$x_n = u_n \exp j(\frac{\pi k n}{2N} + \beta), \qquad (n = 0, 1, \ldots, N-1) \tag{4}$$

$$y_n = v_n \exp j(\frac{\pi k n}{2N} + \beta), \qquad (n = 0, 1, \ldots, N-1) \tag{5}$$

are EOE sequcenes when k is an arbitrary odd integer and β, an arbitrary real constant satisfying $(0 \le \beta < 2\pi)$.

For instance, x and y are called EOE-Gold sequences when u and v are Gold sequences, or EOE-Bent sequences when u and v are Bent sequences.

Evaluation of Performance

We evaluate correlation and BER properties of EOE sequences. Table 1 shows the distribution of the maximum sidelobe, $\theta_m(\cdot)$, of EOE-Gold and Gold sequences' autocorrelation functions. Table 2 does the distribution of the maximum values of crosscorrelation functions, $\theta_m(\cdot,\cdot)$ in every pair of the set of EOE-Gold and Gold sequences.

It is confirmed that BER of systems using EOE sequences is improved over that of systems using original biphase sequences except for the circumstance when the E_b/N_0 is low.

EOE sequences are generally useful for biphase SS systems but not always useful for polyphase SS systems. Moreover, we consider modifying EOE sequences to improve the performance for polyphase SS system. Hence we define the generalized odd correlation function for M-phase SS system and descibe a method of improving the generalized odd correlation function.

References

[1] M. B. Pursley and D. V. Sarwate, "Performance Evaluation for Phase Coded Spread Spectrum Multiple Access Communication – Part II: Code Sequence Analysis," *IEEE Trans. Commun.*, Vol. COM-25, pp. 800–803, 1977.

[2] M. B. Pursley and H. F. A. Roefs, "Numerical Evaluation of Correlation Parameters for Optimal Phases of Binary Shift-Register Sequences," *IEEE Trans. Commun.*, Vol. COM-27, pp. 1597–1604, 1979.

[3] D. V. Sarwate and M. B. Pursley, "Crosscorrelation Properties of Pseudorandom and Related Sequences," *Proc. of IEEE*, Vol. 68, pp. 593–619, 1980.

[4] F. D. Garber and M. B. Pursley, "Performance of Offset Quadriphase Spread Spectrum Multiple Access Communications," *IEEE Trans. Commun.*, Vol. COM-29, No. 3, pp. 305–314, 1981.

[5] E. A. Geraniotis and M. B. Pursley, "Error Probability for Direct Sequence Spread Spectrum Multiple Access Communications – Part II: Approximations," *IEEE Trans. Commun.*, Vol. COM-30, No. 5, pp. 985–995, 1982.

[6] R. Kohno, T. Tanaka and H. Imai, "On Odd-Correlation Functions for the Class of Sequences Produced by Inversion of Alternate Values in Pseudonoise Sequences," *IEICE Trans.*, Vol. J65-A, No. 10, pp. 1029–1030, in Japanese, 1982.

[7] S. M. Krone and D. V. Sarwate, "Quadriphase Sequences for Spread Spectrum Multiple Access Communication," *IEEE Trans. Inform. Theory*, Vol. IT-30, No. 3, pp. 520–529, 1984.

Table 1: Comparison of autocorrelation properties EOE-Gold with Gold

value of autocorrelation	Gold		EOE-Gold
	even	odd	even = odd
1	2		
5		5	2
$\sqrt{45}$			3
7	1	25	4
$\sqrt{63}$			9
$\sqrt{65}$			13
9	30	3	2

Table 2: Comparison of crosscorrelation properties EOE-Gold with Gold

value of crosscorrelation	Gold		EOE-Gold
	even	odd	even = odd
$\sqrt{65}$			14
9	528	20	31
$\sqrt{85}$			83
$\sqrt{101}$			118
$\sqrt{109}$			69
$\sqrt{113}$			21
11		180	
$\sqrt{125}$			96
$\sqrt{137}$			24
$\sqrt{145}$			14
$\sqrt{153}$			27
13		204	1
$\sqrt{181}$			5
$\sqrt{185}$			18
$\sqrt{221}$			7
15		79	
17		33	
19		10	
21		2	

New Enumeration Results for Costas Arrays

Curtis P. Brown, Michal Cenkl,
Richard A. Games, & Joseph J. Rushanan[1]
The MITRE Corporation
202 Burlington Rd.
Bedford, MA 01730

Oscar Moreno
Pei Pei[2]
Department of Mathematics
University of Puerto Rico
Rio Piedras, PR 00931

Summary

An $n \times n$ Costas array is an $n \times n$ array of blanks and dots with exactly one dot in each row and column and with an optimum two-dimensional aperiodic autocorrelation function. In other words, if the Costas array is shifted vertically and/or horizontally (without wraparound) and then compared to a fixed copy of itself, at most a single pair of dots overlap. We denote the Costas array by the associated permutation $(r_0, \ldots, r_{n-1})$ of n elements where there is a dot in position (i, r_i). Costas arrays are used in a variety of ranging and synchronization applications [1], [2].

Costas arrays exist for arbitrarily large n since there are constructions for $n = p - 1$ and $n = q - 2$, where p is prime and q is a power of a prime [3]. They are conjectured to exist for all values of n. The smallest value of n for which there are no known Costas arrays is 32. This case is too large to search exhaustively with today's algorithms and computer technology. We present the enumeration of Costas arrays through $n = 32$ that satisfy, respectively, three different kinds of symmetries. Our aim was to discover a 32×32 Costas array or gather new evidence for its possible nonexistence.

The three types of symmetries considered are:

1. Diagonal: if there is a dot in position (i, r_i), then there is a dot in position (r_i, i). For example, $(0, 4, 6, 3, 1, 7, 2, 5)$. Costas arrays obtained from the Lempel or Golomb construction [3] have diagonal symmetry.
2. Anti-reflective: for n even, $r_i + r_{i+n/2} = n - 1$. For example, $(3, 0, 5, 6, 4, 7, 2, 1)$. Costas arrays obtained from the Welch construction [3] have anti-reflective symmetry.
3. Consecutive: for n even, r_i and r_{n-1-i} are consecutive. For example, $(6, 3, 5, 0, 1, 4, 2, 7)$.

Efficient search algorithms that take advantage of the assumed symmetry were developed and implemented using parallel computer processing. The enumeration through $n = 32$ was completed for each type of symmetry. Figure 1 lists the number of equivalence classes of Costas arrays found for each case. There are no diagonal symmetric Costas arrays for $n = 24$, 28, 31, and 32. This enumeration extends the enumeration described in [4] from size 22. There are no anti-reflective symmetric arrays for $n = 24$, 26, and 32. There are no consecutive symmetric arrays for even n with $22 \leq n \leq 32$.

All Costas arrays have been enumerated previously through size $n = 20$. They have been decreasing in number since $n = 17$. This fact and the above symmetric results contribute to a growing sense that there may not be a Costas array of size 32. We have extended this enumeration to $n = 21$: there are 3536 costas arrays of size 21.

We have constructed a database of all Costas arrays up to size $n = 21$ and have used this database to analyze the underlying structure of the arrays. Such an analysis could be used to expedite the enumeration for higher n, to find new constructions, and possibly to suggest an approach for proving the nonexistence of a 32×32 Costas array. As one example, we measured the size of the largest prefix that adjacent arrays had when the Costas arrays are ordered lexicographically. By $n = 20$, the largest common prefix size is only 6. Some of the other properties investigated include the distribution of the dots of the arrays (there is a tendency for the dots to form an annulus) and the number of dots by quadrants (there is a strong tendency of the dots to be distributed equally in the four quadrants when n is even). We also investigated the cycle structure and ascent properties of Costas arrays in comparison with random permutations.

array size	diagonal	anti-reflective	consecutive
1	1	-	-
2	1	-	-
3	1	-	-
4	1	2	-
5	2	-	-
6	5	4	4
7	10	-	-
8	9	4	10
9	10	-	-
10	14	24	6
11	18	-	-
12	17	44	4
13	25	-	-
14	23	31	5
15	31	-	-
16	20	77	8
17	19	-	-
18	10	29	10
19	6	-	-
20	4	3	3
21	8	-	-
22	5	55	0
23	10	-	-
24	0	0	0
25	2	-	-
26	2	0	0
27	7	-	-
28	0	84	0
29	5	-	-
30	4	60	0
31	0	-	-
32	0	0	0

Figure 1. Number of Symmetric Costas Arrays Equivalence Classes

References

1. J. P. Costas, "A Study of a Class of Detection Waveforms Having Nearly Ideal Rangle-Doppler Ambiguity Properties," *Proc. IEEE*, 72, August 1984, 996-1009.
2. S. W. Golomb and H. Taylor, "Two-dimensional Synchronization Patterns for Minimum Ambiguity," *IEEE Trans. Inform. Thy.*, IT-28, July 1982. 263-272.
3. S. W. Golomb and H. Taylor, "Constructions and Properties of Costas Arrays," *Proc. IEEE*, 72, September 1984, 1143-1163.
4. T. Etzion, "Combinatorial Designs Derived from Costas Arrays," *Sequences*, R. M. Capocelli, editor, New York, NY: Springer Verlag, 1989, pp. 208-227.

[1]The work of Brown, Cenkl, Games, and Rushanan was supported by the MITRE Sponsored Research Program.

[2]The work of Moreno and Pei was supported in part by the Army Research Office Cornell Mathematical Sciences Institute, by the Office of Naval Research under grant number N00014-90-J-1301, and by the NSF-EPSCoR of Puerto Rico Project.

A Partition of the Set of Permutations by the Monotone Subsequence Structure

KINGO KOBAYASHI AND HIROYOSHI MORITA

Department of Computer Science and Information Mathematics,
The University of Electro-Communications, Chofu, Tokyo 182, JAPAN

Abstract

We will study the distribution of longest upward and downward monotonic subsequences contained in sequences, or permutations of $1, 2, \ldots, n$. Thereby, a famous Ramsey-type existance theorem of sequence having a specific property is refined by the precise counting technique.

Summary

A *subsequence* of the permutation $\pi_1\pi_2\ldots\pi_n$ of $\{1,\ldots,n\}$ is a sequence considered in the same order as the numbers appear in the permutation. For example, 142 is a subsequence of permutation 7134652, but 753 is not. Given any permutation, a subsequence $\pi_{i_1}\pi_{i_2}\ldots\pi_{i_k}$ $(i_1 < i_2 < \ldots < i_k)$ is *upward monotonic* if it is always increasing, that is, $\pi_{i_1} < \pi_{i_2} < \ldots < \pi_{i_k}$. Similarly, a subsequence is *downward monotonic* if it is always decreasing. Then, we are interested in the longest length pair (d, u) of downward and upward monotonic subsequences contained in given permutation of order n. Our main concern is the determination of the number $c^{(n)}(d, u)$ of permutations of order n having the pair (d, u) for any $1 \le d, u \le n$.

This problem contains the famous theorem of Ramsey type as a special case. That theorem due to Erdös and Szekeres [1],[2] states that any sequence of distinct $n^2 + 1$ numbers containes a monotone (downword or upward) subsequence of length $n + 1$. This theorem can be expressed in our notation as

$$\lceil \sqrt{n} \rceil = \min_{c^{(n)}(d,u)>0} \max\{d, u\}$$

We can give the number $c^{(n)}(d, u)$ by making full use of the properties of Pascal triangle of binary coefficients and their combinatorial meanings. For example, we obtain the table of the number $c^{(n)}(d, u)$ for $n = 10$ as depicted in the following:

d \ u	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	1764	8100	5625	1225	81	0
3	0	0	0	107604	285444	149850	25600	1296	0	0
4	0	0	107604	769824	597114	122500	7056	0	0	0
5	0	1764	285444	597114	200704	15876	0	0	0	0
6	0	8100	149850	122500	15876	0	0	0	0	0
7	0	5625	25600	7056	0	0	0	0	0	0
8	0	1225	1296	0	0	0	0	0	0	0
9	0	81	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0	0

To reveal the riddle of the numbers in such tables, we must prepare some interesting recursion formula for auxiliary tables induced from Pascal triangle. As byproducts we have some formulas such as

$$\begin{aligned} c^{(n)}(d, n-d+1) &= \binom{n-2}{d-2}\binom{n-1}{d-1} + \binom{n-1}{d-1}\binom{n-2}{d-1} \\ &= \binom{n-1}{d-1}^2 \end{aligned}$$

$$\sum_{d=1}^{n} c^{(n)}(d, n-d+1) = \binom{2(n-1)}{n-1}$$

and

$$\sqrt{c^{(n)}(d,2)} = \begin{cases} 0 & \text{for } d < \lceil n/2 \rceil \text{ or } d = n \\ \sqrt{c^{(n-1)}(d-1,2)} + \sqrt{c^{(n-1)}(d,2)} & \text{for } \lceil n/2 \rceil \le d < n-1 \\ \sqrt{c^{(n-1)}(n-2,2)} + 1 & \text{for } d = n-1 \end{cases}$$

References

[1] Erdös and P. Szekeres, G., "On some extremum problems in elementary geometry," Ann. Univ. Sci. Budapest, 3-4, 53-62, 1960-61.

[2] Seidenberg, "A simple proof of a theorem of Erdös-Szekeres," J. London Math. Soc. 34, 352, 1959.

Optimal and Suboptimal Biphase Sequences of Period $2(2^r-1)$ and Linear Complexity $r(r+3)/2$

P.Udaya and M.U. Siddiqi
Department of Electrical Engineering
Indian Institute of Technology
Kanpur, 208016 (INDIA)

Introduction

This paper is concerned with construction of new families of biphase sequences which are obtained through a polynomial mapping from the ring Z_4 to GF(2). Such sequences are of interest in code division spread spectrum multi-user communication systems.

Optimal and suboptimal families of quadriphase sequences derived from maximal length sequences (m-sequences) and interleaved maximal length sequences (Im-sequences) over Z_4 are given in [1,2]. The period of m-sequences is 2^r-1 and that of Im-sequences is $2(2^r-1)$. The families of 2^r+1 biphase sequences of period 2^r-1 derived from families of m-sequences over Z_4 are optimal like Gold families and are reported in [6]. Linear complexity (LC) of these sequences is lower bounded by $r(r-1)/2$.

In this paper, we derive biphase families of size $2^{r-1}+1$, r a positive integer, from families of Im-sequences over Z_4. Most of the families satisfy Sidelnikov bound on θ_{max} ($\theta_{max} < \sqrt{2L}$, where L is equal to the period of the sequences) which is equal to the maximum magnitude of the periodic crosscorrelation and out of phase autocorrelation values. One of the families satisfies Welch bound on θ_{max} ($\theta_{max} < \sqrt{L}$), while rest of the families are suboptimal (θ_{max} is bounded by $2\sqrt{L}$). The linear complexity of all sequences is equal to $r(r+3)/2$ with the exception of the single m-sequence. Sequence imbalance and correlation distributions are also computed.

Main Results

We consider a non-linear polynomial ($\mathcal{NLP}$) mapping from Z_4 to its ideal $< 2 >$, given by $\varphi(x) = x^2-x$. The ideal $< 2 >$ is isomorphic to the binary field and the quadriphase mapping given by $\phi(x) = \omega^x$, where $\omega = \sqrt{-1}$, on the sequences over the ideal results in biphase sequences. Thus biphase families are constructed from families of Z_4 sequences by using the $\mathcal{NLP}$ mapping $\varphi(x)$ given above. The families of Z_4 sequences considered for biphase sequence construction in this paper are the families of Im-sequences over Z_4(or simply $\mathcal{IM}$ families), of period $2(2^r-1)$. Each family consists of $(2^{r-1}+1)$ sequences [1,2]. The definition of $\mathcal{IM}$ families depends on the structure of the Galois extension ring of Z_4, $GR(4,r)$, r a positive integer. Any group of units $GR^*(4,r)$ of a Galois ring $GR(4,r)$ is a direct product of two groups G_a and G_c, where G_a is Abelian group of order 2^r, and G_c is the cyclic component group of order 2^r-1. Associated with every element γ of $G_a \in GR^*(4,r)$, $\gamma \neq 1$, a $\mathcal{IM}$ family is defined. Three classes of $\mathcal{IM}$ families are identified depending on the nature of γ. They are

(a) $\mathcal{IM}^1$ families with $\text{trace}(\bar{\gamma}) = 1$

(b) $\mathcal{IM}^0$ families with $\text{trace}(\bar{\gamma}) = 0$, $\gamma \neq 1$

(c) $\mathcal{IM}^3$ family

where $\gamma = 1 + 2(\bar{\gamma})$, $\bar{\gamma} = \gamma' \bmod 2$, $\gamma' \in G_c$. A paper by Bostas, Hammons and Kumar [4] contains some of the Z_4 families given above. The families given in [4] correspond, as per the above classification, to $\mathcal{IM}^1$ families with $\text{tr}(\gamma) = 1$. The $\mathcal{IM}^1$ families with $\text{tr}(\bar{\gamma}) = 0$ and the $\mathcal{IM}^3$ family are additional sub-families among $\mathcal{IM}$ families of period $2(2^r-1)$. These families through $\mathcal{NLP}$ and quadriphase mappings yield biphase families. The biphase families thus constructed are named by prefixing the word $\mathcal{NLP}$ to their corresponding Z_4 families. The $\mathcal{NLP}$ mapping considered in this paper is equivalent to quarternary to binary transformation given in [3]. The correlation properties of binary families are computed from those of Z_4 families by using a method given in [3]. The bounds on the θ_{max} of binary families given [3] are improved by making use of specific correlation properties of $\mathcal{IM}$ families. The correlation distributions and sequence imbalance are computed by making use of the properties of Galois ring GR(4,r).The correlation properties of biphase families derived are tabulated in Table 1. The LC of resultant binary sequences is computed through the LC analysis of corresponding binary ideal sequences over Z_4. A generalised version of Blahut's theorem, which relates LC of a sequence over Z_4 to the number of non-zero positions in its Fourier transform, is used to compute the LC of ideal sequences. Since this ideal is isomorphic to the binary field, LC of binary ideal sequences thus computed via Blahut's theorem is indeed LC of the binary sequences. The LC of all sequences derived from $\mathcal{IM}$ families is equal to $r(r+3)/2$ with an exception of the single m-sequence.

Table 1 New Biphase Sequence Designs
Family Size: $2^{r-1}+1$; LC: $r(r+3)/2$; Period: $2(2^r-1)$

Family	r	C_{max}	Comment
$\mathcal{NLP}$-$\mathcal{IM}^1$ (γ=3)	odd	$2(1+2^{(r-1/2)})$	Optimal (Welch)
$\mathcal{NLP}$-$\mathcal{IM}^1$ $\text{tr}(\bar{\gamma})$=1 (γ=3)	even	$2(1+2^{(r/2)})$	Optimal (Sidelnikov)
$\mathcal{NLP}$-$\mathcal{IM}^1$ $\text{tr}(\bar{\gamma})$=0 $\text{tr}(\bar{\gamma})$=1	odd	$2(1+2^{(r+1/2)})$	Sub optimal (Sidelnikov)
$\mathcal{NLP}$-$\mathcal{IM}^1$ $\text{tr}(\bar{\gamma})$=0	even	$2(1+2^{(r+1/2)})$	Sub optimal (Sidelnikov)

REFERENCES:

[1] Udaya P, "Polyphase and Frequency Hopping Sequences obtained from Finite Rings", Ph.D Thesis, Department of Electrical Engineering, I.I.T Kanpur, 1992.

[2] Udaya P and M.U. Siddiqi, "Optimal Quadriphase Sequences Derived from Maximal Length Sequences over Z_4", Submitted to Journal of Applicable Algebra in Engineering, Communication and Computing, Springer-Verlag.

[3] S.M.Krone and D.V. Sarwate ," Quadriphase Sequences for Spread-Spectrum Multiple-Access Communication", Vol. IT-30, No. 3, May 1984, pp 520-529.

[4] S. Bostas, R. Hammons, and P. V. Kumar, 'Near-Optimal Sequences for CDMA', IEEE Trans Inform. Theory, Vol. IT-38, No. 3, May 1992, pp 1101-1113.

[5] J. L. Massey and T. Schaub, "Linear Complexity in Coding Theory", Coding Theory and Applications, Lecture Notes in Comp. Sc., Vol. 311, 1988

[6] S. Bostas and P. V. Kumar, "Binary Sequences with Gold like Correlation Properties but Larger Linear Span", 1991 IEEE international Symposium on Inform. Theory, Budapest, Hungary, June 23-29, pp 381.

Perfect Maps

Kenneth G. Paterson

Dept. of Mathematics
Royal Holloway
University of London
Egham, Surrey TW20 0EX

Abstract

Given positive integers r,s,u and v, an (r,s;u,v) Perfect Map (PM) is defined to be a periodic r x s binary array in which every u x v binary array appears exactly once as a subarray. Perfect Maps are the natural extention of the de Bruijn sequences to two dimensions.

In this paper we settle the existence question for Perfect Maps by proving the following result.

Let r,s,u,v be positive integers. Then there exists an (r,s;u,v) PM if and only if the following three conditions hold:

i) $rs = 2^{uv}$,

ii) $r > u$ or $r = u = 1$,

iii) $s > v$ or $s = u = 1$.

We make extensive use of previously known constructions by finding new conditions guaranteeing their repeated application. These conditions are expressed as bounds on the linear complexities of the periodic sequences formed from the rows and columns of Perfect Maps.

NEW BOUNDS FOR THE SIZE OF RADAR ARRAYS

Zhen Zhang and Chungming Tu*
Communication Sciences Institute, Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089-2565

Abstract

A "radar array" is a matrix of zeros and ones which has small one-dimensional autocorrection sidelobes. New general constructions and new upper bounds for the size of radar arrays are presented.

Summary

A radar array R is an $N \times M$ matrix of ones and zeros with a single one per column, such that the horizontal autocorrelation function only has values $0, 1, \ldots, k$, and M, where k is the maximal allowable sidelobe (generalized from [2] and [3]). We say a radar array is optimal if it has the maximum M for given values of N and k. Denote that maximum number as $GR_k(N)$. Previously the best known asymptotic upper and lower bounds for $GR_1(N)/N$ have a gap of about 0.463 ([4]). This gap is shrunk to about 0.101 in this paper.

We applied the Erdös–Turan ([1]) argument to obtain the upper bounds. Suppose a window of width K $(N \le K \le M)$ is superimposed on the radar array. Let's slide the window from the left end of the array to the right end and count the number of spacings (the distance between any two 1's in the same row) within the windows. By estimating the minimum number of spacings in each window and considering how many windows contain a particular spacing, we have the following theorem. (For details, see [5].)

Theorem 1: $GR_k(N) \le \min\{g_1(1 + \lceil \rho_c \rceil), \min_{2 \le \rho \le 1+\rho_c} g_2(\rho)\}$,

where

$$\rho_c = \sqrt{1 + 6k - \frac{6k}{N}}; \quad g_1(\rho) = \frac{2k+\rho}{2}N; \quad g_2(\rho) = \frac{X + \sqrt{Z}}{Y}$$

$$X = 2\binom{\rho}{2}N + k\rho N - k; \quad Y = 2(\rho - 1)$$

$$Z = k^2\rho^2N^2 - 4k\binom{\rho}{3}N^2 - 2k^2\rho N + k^2$$

New heuristic methods are used to search for radar arrays with smaller sizes. Table 1 summarizes new upper and lower bounds for the $k = 1$ case. Next we introduce a new construction for the radar arrays which gives asymptotic lower bounds.

Define two $L \times L$ (L odd) permutation matrices A, B to be *"properly centered"* if their row-by-row differences range exactly from $-\frac{L-1}{2}$ to $\frac{L-1}{2}$. Table 2 shows some examples of such permutation matrices. (The existence of such permutation matrices of other odd sizes is still unknown. It would also be useful if one can find more than 4 such matrices.)

Given any $k = 1$, size $N \times M$ radar array with at most three 1's in each row (such radar arrays exist for all small N), a new $k = 1$ radar array of size $NL \times ML$ can be constructed according to the following rules.

1. Choose a set of A, B, C from Table 2.
2. Substitute any zero in the rows with an $L \times L$ all zero matrix.
3. Substitute the first 1 in each row with A.
4. Substitute the second 1 in each row with B.
5. Substitute the third (if any) 1 in each row with C.

For a proof of this construction, see [5]. The best achievable $GR_k(N)/N$ ratios by using this method is 306/113 when $k = 1$ and 4 when $k = 2$. Combined with the result of Theorem 1, we have $2.708 \le \overline{\lim}_{N\to\infty} GR_1(N)/N \le 2.809$ and $4 \le \overline{\lim}_{N\to\infty} GR_2(N)/N \le 4.276$.

N	max M found	upper bound	N	max M found	upper bound
2	4	4	14	37	38
3	7	7	15	40	41
4	10	10	16	42	44
5	12	13	17	45	47
6	15	16	18	48	49
7	18	19	19	51	52
8	21	21	20	53	55
9	23	24	21	56	58
10	26	27	22	59	61
11	29	30	23	61	64
12	32	33	24	63	66
13	34	35	25	65	69

Table 1: Some Upper and Lower Bounds for the $k = 1$ Case

L = 5	A = $[1,2,3,4,5]^T$ B = $[1,3,5,2,4]^T$ C = $[2,1,5,4,3]^T$ D = $[3,1,4,2,5]^T$
L = 7	A = $[1,2,3,4,5,6,7]^T$ B = $[1,3,5,7,2,4,6]^T$ C = $[3,1,6,4,2,7,5]^T$
L = 13	A = $[1,2,3,4,5,6,7,8,9,10,11,12,13]^T$ B = $[1,3,2,7,11,10,4,13,5,12,9,6,8]^T$ C = $[5,3,8,2,7,12,1,11,4,6,10,9,13]^T$

Table 2: Some Permutation Matrices with Pairwise Properly Centered Sets of Differences

*This research is supported in partby NSF under Grant NCR-8905052

Reference

[1] H. Halberstam and K.F. Roth, Sequences, vol. I. Oxford: Clarendon, 1966.

[2] S.W. Golomb and H. Taylor, "Two-dimensional synchronization patterns for minimum ambiguity," IEEE Trans. Inform. Theory, vol. IT-28, no. 4, pp. 600–604, July 1982.

[3] J.P. Robinson, "Golomb Rectangles," IEEE Trans. Inform. Theory, vol. IT-31, no. 6, pp. 781–787, Nov. 1985.

[4] A. Blokhuis and H.J. Tiersma, "Bounds for the size of radar arrays," IEEE Trans. Inform. Theory, vol. 34, no. 1, pp. 164–167, Jan. 1988.

[5] Z. Zhang and C. Tu, "New bounds for the size of radar arrays," submitted to IEEE Trans. Inform. Theory.

FAMILIES OF FOUR-PHASE QUASI-ORTHOGONAL CODE ARRAYS

Serdar Boztaş
Department of Electrical and Computer Systems Engineering
Monash University, Clayton, Victoria 3168, Australia

Abstract

A method of construction is presented for rectangular quasi-orthogonal code arrays over the ring Z_4.

The proposed arrays are easy to generate: The four-phase linear recurring sequences constructed by Boztaş et. al. *are utilized to generate the arrays by modulo 4 subtraction. The periodic auto- and cross-correlation properties of these arrays are then derived in a straightforward manner. The maximum off-peak correlation magnitude for these arrays is lower by a factor of $\sqrt{2}$ when compared to the binary Gold code arrays constructed by Kuo and Rigas.*

The arrays can be used for 'add-on' data transmission, pattern synchronization, and code division image multiplexing.

INTRODUCTION

Kuo and Rigas introduced binary quasi m-arrays and quasi Gold arrays in [1]. These arrays were proposed to overcome some disadvantages associated with the m-arrays studied by Nomura *et. al.* [2] and MacWilliams and Sloane [3]. Their construction for quasi m-arrays yields $L_1 \times L_2$ binary arrays where $L_i = 2^{n_i} - 1$, with n_i positive integers for $i = 1, 2$.

Given two binary sequences, say $s_1(t)$, $t = 0, 1, \ldots, L_1 - 1$ and $s_2(t)$, $t = 0, 1, \ldots, L_2 - 1$ (these sequences can either be two m-sequences or two Gold sequences) Kuo and Rigas use the construction $a(t_1, t_2) = s_1(t_1) \oplus s_2(t_2)$ where $\oplus$ denotes modulo 2 addition.

The maximum off-peak auto-correlation for the quasi m-arrays is given by $max\{L_1, L_2\}$ while the maximum cross-correlation magnitude depends on the choice of the m-sequences. This is one reason for the introduction of Gold code arrays in [1]. The maximum off-peak auto- and cross-correlation magnitude for the Gold code arrays is given by $\Theta_{max} = max\{L_1 \cdot (\sqrt{2L_2} + 1), L_2 \cdot (\sqrt{2L_1} + 1)\}$.

THE NEW CONSTRUCTION

In this paper four-phase quasi orthogonal arrays are introduced. The method used for the construction of these arrays from four-phase sequences is similar to the method used in [1].

Boztaş, Hammons and Kumar [4] constructed families of four-phase linear recurring sequences with near optimum correlation properties. These sequences are used here to construct new families of four-phase quasi-orthogonal code arrays. The reader is referred to [4] for a tabulation of generating polynomials (hence recursion coefficients) for these sequences. The sequences that are used in the construction here are referred to as the *family $\mathcal{A}$* in that paper and are defined as all the nonzero sequences satisfying a given linear recursion over Z_4.

Given two four-phase sequences (say $s_1(t)$, $t = 0, 1, \ldots, L_1 - 1$ and $s_2(t)$, $t = 0, 1, \ldots, L_2 - 1$, where $L_i = 2^{n_i} - 1$, with n_i a positive integer for $i = 1, 2$) belonging to family $\mathcal{A}$ the four-phase quasi orthogonal array $a(t_1, t_2)$ of size $L_1 \times L_2$ can be constructed by

$$a(t_1, t_2) = s_1(t_1) \ominus s_2(t_2) \tag{1}$$

where $\ominus$ denotes modulo 4 subtraction.

Definition 1 *The cross-correlation between two four-phase arrays $a(t_1, t_2)$ and $b(t_1, t_2)$ of the same dimensions $L_1 \times L_2$ is given by*

$$\Theta_{ab}(\tau_1, \tau_2) = \sum_{t_1=0}^{L_1-1} \sum_{t_2=0}^{L_2-1} \omega^{a(t_1+\tau_1, t_2+\tau_2) \ominus b(t_1, t_2)} \tag{2}$$

where $0 \le \tau_i \le L_i$, and the sums $t_i + \tau_i$ are interpreted modulo L_i, for $i = 1, 2$, and ω is defined as $e^{2\pi i/4}$.

Theorem 1 *(a) The cross-correlation between two four-phase quasi-orthogonal arrays satisfies*

$$| \Theta_{ab}(\tau_1, \tau_2) | \le \theta_{max}(L_1)\theta_{max}(L_2), \tag{3}$$

where $\theta_{max}(L)$ is the maximum off-peak auto- and cross-correlation magnitude of the four-phase sequences in family $\mathcal{A}$ of length L.
(b) The auto-correlation of a four-phase quasi-orthogonal array satisfies

$$| \Theta_{aa}(\tau_1, \tau_2) | \le \max\{L_1\theta_{max}(L_2),\ L_2\theta_{max}(L_1)\}. \tag{4}$$

Proof The proof is straightforward. Denote the two sequences in family $\mathcal{A}$ used to generate $b(t_1, t_2)$ by s_1' and s_2', i.e., $b(t_1, t_2) = s_1'(t_1) \ominus s_2'(t_2)$ and $a(t_1, t_2) = s_1(t_1) \ominus s_2(t_2)$. Substituting this in

$$\Theta_{ab}(\tau_1, \tau_2) = \sum_{t_1=0}^{L_1-1} \sum_{t_2=0}^{L_2-1} \omega^{a(t_1+\tau_1, t_2+\tau_2) \ominus b(t_1, t_2)} \tag{5}$$

gives

$$\Theta_{ab}(\tau_1, \tau_2) = \sum_{t_1=0}^{L_1-1} \sum_{t_2=0}^{L_2-1} \omega^{s_1(t_1+\tau_1) \ominus s_2(t_2+\tau_2) \ominus (s_1'(t_1) \ominus s_2'(t_2))} \tag{6}$$

or

$$\Theta_{ab}(\tau_1, \tau_2) = \sum_{t_1=0}^{L_1-1} \omega^{s_1(t_1+\tau_1) \ominus s_1'(t_1)} \sum_{t_2=0}^{L_2-1} \omega^{s_2'(t_2) \ominus s_2(t_2+\tau_2)}. \tag{7}$$

Note that the right hand side is just a product of two cross-correlations between pairs of sequences from family $\mathcal{A}$. Case (a) follows directly from this.
In case (b), $s_1 \equiv s_1'$ and $s_2 \equiv s_2'$ as sequences and therefore when either $\tau_1 = 0$ or $\tau_2 = 0$, the corresponding sum yields L_1 or L_2 respectively which proves the claim for auto-correlation. □
The theorem yields the immediate corollary below.

Corollary 2 $\Theta_{max} = \max\{L_1 \cdot (\sqrt{L_2 + 1} + 1),\ L_2 \cdot (\sqrt{L_1 + 1} + 1)\}$ *for the quasi-orthogonal four-phase sequences constructed here, and is lower than that of the Gold code arrays in [1] by a factor of $\approx \sqrt{2}$.*

Proof $\theta_{max}(L) = 1 + \sqrt{L + 1}$ for family $\mathcal{A}$ (see [4]) while it is $1 + \sqrt{2L}$ for Gold sequences. □

REFERENCES

[1] C. J. Kuo and H. B. Rigas, '2-D Quasi m-arrays and Gold code arrays', *IEEE Transactions on Information Theory*, vol. IT-37, no. 2, March 1991.

[2] T. Nomura, H. Miyakawa, H. Imai, and A. Fukuda, 'A theory of two-dimensional linear recurring arrays', *IEEE Transactions on Information Theory*, vol. IT-18, no. 6, November 1972.

[3] F. J. MacWilliams and N. J. A. Sloane, 'Pseudo-random sequences and arrays,' *Proceedings of the IEEE*, vol. 64, December 1976.

[4] S. Boztaş, R. Hammons, and P. Vijay Kumar, 'Four-phase sequences with near optimum correlation properties', *IEEE Transactions on Information Theory*, vol. IT-38, no. 3, May 1992.

Crosscorrelation of GMW sequences

Markus Antweiler

Institut für Elektrische Nachrichtentechnik, RWTH Aachen, D-5100 Aachen, Germany *

We consider p-nary GMW-sequences of length $p^M - 1$ which are defined as

$$s(n) = \exp\left(\frac{j2\pi}{p} a(n)\right) \text{ with } a(n) = \mathrm{tr}_1^J(\mathrm{tr}_J^M(\alpha^{dn})^r) \tag{1}$$

and some restrictions on the parameters J, d and r (see [1, 2]). $\mathrm{tr}(\cdot)$ denotes the trace function from the finite field $\mathrm{GF}(p^M)$ onto $\mathrm{GF}(p^J)$, and α is a primitive element of $\mathrm{GF}(p^M)$. The periodic crosscorrelation function (PCF) of s and g with $g(n) = \exp\left(\frac{j2\pi}{p} \mathrm{tr}_1^J(\mathrm{tr}_J^M(\alpha^{en})^s)\right)$ is defined by

$$\tilde{\varphi}_{sg}(k) = \sum_{n=0}^{p^M-2} s^*(n) g(n+k),$$

where $n+k$ is taken modulo $p^M - 1$. The crosscorrelation function depends on the parameters d, r, e and s. Therefore we write $\tilde{\varphi}_{sg}(k) = \tilde{\varphi}_{d,r,e,s}(k)$. The periodic crosscorrelation function of two p-nary m-sequences becomes in this notation $\tilde{\varphi}_{d,1,e,1}$, because s and g are equal to m-sequences for $r = s = 1$.

The paper aims at the calculation of the correlation function of GMW-sequences by reducing it to the PCF of ordinary m-sequences, because results on the crosscorrelation functions of m-sequences are well known (see [3] for a compressed description for $p = 2$ and [4, 5] for $p > 2$). One possible way is the description by the crosscorrelation function of shorter m-sequences with length $p^J - 1$. This was done in the papers [6, 7] for $d = e$ and $r = 1$, so that the crosscorrelation of an m-sequence ($r = 1$) and a GMW-sequence with 'same primitive polynomial' ($d = e$) is known up to now. We generalize this result to the case $r \neq 1$, so that for the first time the crosscorrelation of two GMW sequences was investigated:

Theorem 1 *The crosscorrelation for $d = e = 1$ is* $\tilde{\varphi}_{1,r,1,s}(k) =$

$$= \begin{cases} p^{M-J}(\tilde{\varphi}_{r,s}(k/T) + 1) - 1, & \text{for } k \equiv 0 \bmod T \\ -1, & \text{else,} \end{cases}$$

where $\tilde{\varphi}_{r,s}$ denotes the crosscorrelation function of the m-sequences $\exp(j2\pi\,\mathrm{tr}_1^J(\gamma^{rn})/p)$ *and* $\exp(j2\pi\,\mathrm{tr}_1^J(\gamma^{sn})/p)$ *of length $p^J - 1$ ($\gamma = \alpha^T, T = (p^M - 1)/(p^J - 1)$).*

Another way is the reduction of the PCF of GMW-sequences to the PCF of m-sequences with the same length: $\tilde{\varphi}_{d,r,e,s}(k) = \tilde{\varphi}_{d,1,e,1}(k)$, whereby restrictions has to be fullfilled by the parameters d, r, e and s. For two cases we found a description of the PCF of GMW sequences in this form:

Theorem 2 *The crosscorrelation function for $r = s$ and $d \equiv ep^k \bmod p^J - 1$ is*

$$\tilde{\varphi}_{d,r,e,r}(k) = \tilde{\varphi}_{d,1,e,1}(k).$$

This theorem allows the calculation of PCF of GMW sequences having the same linear span, because the linear span depends only on r and s. (The linear span is the minimal degree of a linear recursion satisfied by $a(n)$ in eq.(1)).

Theorem 3 *The crosscorrelation function for $d \equiv -esp^k \bmod p^J - 1$ and $rs \equiv p^l \bmod p^J - 1$ is*

$$\tilde{\varphi}_{d,r,e,s}(k) = \tilde{\varphi}_{d,1,e,1}(k).$$

The meaning of the condition $rs \equiv p^l$ or $r \equiv p^l s^{-1}$ is that the nonlinear mappings which are performed by raising to the rth and sth power are inverse to one another.

With known results on the PCF of m-sequences these three theorems allow the calculation of the PCF of GMW sequences for many cases.

References

[1] R.A. Scholtz and L.R. Welch, "GMW sequences," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 548 - 553, 1984.

[2] M. Antweiler and L. Bömer, "Complex sequences over $\mathrm{GF}(p^M)$ with a two-level autocorrelation function and a large linear span," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 120 - 130, 1992.

[3] D.V. Sarwate and M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," *Proc. IEEE*, vol. 68, pp. 593–619, 1980.

[4] H.M. Trachtenberg, *On the Cross-Correlation Function of Maximal Linear Recurring Sequences*. PhD thesis, University of Southern California, Los Angeles, January 1970.

[5] T. Helleseth, "Some results about the cross-correlation function between two maximal linear sequences," *Discrete Math.*, vol. 16, pp. 209 - 232, 1976.

[6] R.A. Games, "Crosscorrelation of m-sequences and GMW-sequences with same primitive polynomial," *Discrete Applied Mathematics*, vol. 12, pp. 139 - 146, 1985.

[7] A.H. Chan, M. Goresky, and A. Klapper, "Correlation functions of geometric sequences," in *Advances in Cryptology — Eurocrypt '90*, pp. 214–221, Springer-Verlag, 1990.

*Mr. Antweiler in now with CADIS, Kaiserstr. 100, D-5120 Herzogenrath, Germany

NON-BINARY SEQUENCES WITH THE PERFECT PERIODIC AUTO-CORRELATION AND WITH OPTIMAL PERIODIC CROSS-CORRELATION

Ernst M. Gabidulin
Moscow Institute of Physics and Technology, Institutskii per., 9,
141700 Dolgoprudnyi, Moscow region, USSR, e-mail: gab@ippi.msk.su

ABSTRACT

We propose two families of complex sequences with components on the unit circle (PSK sequences). Each sequence of a family has the perfect auto-correlation, i.e., all "out-of-phase" correlation coefficients are equal to zero. Magnitudes of all cross-correlation coefficients of any couple of sequences in a family are equal to the square root of a sequence length n. Thus both families are asymptotically optimal with respect to the Sidelnikov-Welch's lower bound.

1. SUMMARY

Let $M = \{\mathbf{x}(m) = (x_0(m), x_1(m), \ldots, x_{n-1}(m)),\ m = 1, 2, \ldots, M\}$ denote a family of complex sequences of length n. Let

$$R_t(i,i) = R_t(i) = \sum_{s=0}^{n-1} x_s(i) x_{s+t}^*(i)\ , \quad \begin{matrix} t = 0, 1, \ldots, n-1 \\ i = 1, 2, \ldots, M, \end{matrix} \tag{1}$$

$$\text{and} R_t(i,j) = \sum_{s=0}^{n-1} x_s(i) x_{s+t}^*(j)\ , \quad \begin{matrix} t = 0, 1, \ldots, n-1 \\ i, j = 1, 2, \ldots, M,\ i \neq j, \end{matrix} \tag{2}$$

denote the periodic auto- and cross-correlation coefficients, respectively. x^* denote the complex conjugate of x, subscripts are calculated modulo n.

Let r denote the maximum nontrivial coefficient. If all sequences have the same energy, say n, i.e., $R_0(i) = n$, then the Sidelnikov-Welch's lower bound [1],[2] is as follows

$$r = n\sqrt{\frac{M-1}{Mn-1}}. \tag{3}$$

There are a lot of papers devoted to designing of families with near-optimum correlation properties. Among other well known are Gold and Kasami families. We propose two new one's.

The sequence is said to be a perfect one if all "out-of-phase" auto-correlation coefficients are equal to zero.

Lemma 1 [3]: A sequence $\mathbf{x} = (x_0, x_1, \ldots, x_{n-1})$ is a perfect sequence if and only if all components of a sequence $\mathbf{y} = (y_0, y_1, \ldots, y_{n-1})$ have the same magnitude $\sqrt{R_0(\mathbf{x})} = \sqrt{n}$, where $\mathbf{y}$ is the Discrete Fourier Transform (DFT) of $\mathbf{x}$, i.e.,

$$y_j = \frac{1}{\sqrt{n}} \sum_{s=0}^{n-1} x_s \zeta^{sj}, \qquad j = 0, 1, \ldots, n-1, \tag{4}$$

where ζ is a primitive root of unity of degree n.

The sequence is known as the phase shift keyed (PSK) sequence if all components of this sequence are on the unite circle.

The first family $\mathcal{M}_1$ consists of sequences $\mathbf{x}(m)$ of length $n = p^{2k}$, where p is an odd prime. Any integer s, $0 \leq s \leq p^{2k} - 1$, can be represented uniquely in a form

$$s = up^k + v, \tag{5}$$

where $0 \leq u \leq p^k - 1$, $0 \leq v \leq p^k - 1$.

Let ζ be a primitive root of unity of degree p^{2k} and let λ be a primitive root of unity of degree $p^k = \sqrt{n}$. Consider sequences $\mathbf{x}(m) = (x_0(m), x_1(m), \ldots, x_{n-1}(m))$, whose sth components are

$$x_s = z_v(m)\lambda^{muv}, s = 0, 1, \ldots, n-1, \tag{6}$$

where $(m, p) = 1$, u and v are integers from Eqn. (5), and $z_v(m)$, $0 \leq v \leq p^k - 1$, are arbitrary complex numbers with absolute values 1.

Lemma 2 [4]: Sequences (6) are perfect sequences.

(Note, that if $z_v(m) = 1$ for all v then these sequences are well known Frank's sequences [5].)

Theorem 1: Let $\mathcal{M}_1$ is a set of sequences (6), where $m = 1, 2, \ldots, p-1$. Then all cross-correlation coefficients have the same magnitude $p^k = \sqrt{n}$.

Proof. If an integer t has a representation $t = ap^k + b$, then an integer $s + t$ has a representation $s + t = (u + a + \epsilon)p^k + (v + b - \epsilon p^k)$, where $\epsilon = 0$, if $v + b \leq p^k - 1$, and $\epsilon = 1$, if $v + b \geq p^k$. Thus

$$R_t(i,j) = \sum_{v=0}^{p^k-1} \sum_{u=0}^{p^k-1} z_v(i) z_{v+b}^*(j) \lambda^{tuv} \lambda^{-j(u+a+\epsilon)(v+b)}$$

$$= \sum_{v=0}^{p^k-1} z_v(i) z_{v+b}^*(j) \lambda^{-j(a+\epsilon)(v+b)} \Big(\sum_{u=0}^{p^k-1} \lambda^{u((i-j)v - jb)}\Big) \tag{7}$$

The inner sum in (7) is equal to 0, if $(i-j)v - jb \neq 0$, and is equal to $p^k = \sqrt{n}$, if $(i-j)v - jb = 0$. This equation has a unique solution $v_1 = (i-j)^{-1} jb$, since $i - j \neq 0$ modulo p. Thus

$$R_t(i,j) = \sqrt{n} z_{v_1}(i) z_{v_1+b}^*(j) \lambda^{-j(a+\epsilon)(v_1+b)}. \tag{8}$$

Corollary: If $n = p^2$ then there exists a family $\mathcal{M}_1$ of size $M = \sqrt{n} - 1$ with near-optimum cross-correlation $\sqrt{n}$. It is comparable with parameters of the Kasami family, but the auto-correlation is perfect.

Now we describe sequences of length $n = p^{2k+1}$ with the perfect auto-correlation. Every integer s, $0 \leq s \leq n-1$, can be represented uniquely in a form $s = up^{k+1} + vp^k + c$, where $0 \leq u < p^k$, $0 \leq v < p$, $0 \leq c < p^k$. Let for any c a sequence $(x_{v,c}, 0 \leq v < p)$ be a sequence of length p with the perfect auto-correlation. Let ζ be a primitive root of unity of degree n, λ be a primitive root of unity of degree p^{k+1}, ρ be a primitive root of unity of degree p^k. **Theorem 2**: A sequence of length n whose sth component is equal to

$$x_s = x_{v,c} \lambda^{uc} \rho^{vc}, \qquad 0 \leq s < n, \tag{9}$$

has the perfect auto-correlation.

Proof. Straightforward calculation of the DFT of the sequence (9) shows that all Fourier coefficients have the same magnitude.

Consider a family $\mathcal{M} = \{\mathbf{x}(m)\}$, each element of which equals $(x_0(m), x_1(m), \ldots, x_{p-1}(m))$, where $x_s(m) = \mu^{ms^2}$, μ is a primitive root of unity of degree p. It is known that $\mathbf{x}(m)$, $m = 1, 2, \ldots, p-1$, are perfect sequences.

Lemma 3 [6]: All cross-correlation coefficients of sequences from the family $\mathcal{M}$ have the same magnitude $\sqrt{p}$.

Consider a set $\mathcal{M}_2$ of $p-1$ sequences of a form (9) where instead of $x_{v,c}$ one uses $x_v(m)$ from the family $\mathcal{M}$.

Theorem 3: All cross-correlation coefficients of the family $\mathcal{M}_2$ have the same magnitude $\sqrt{n}$.

REFERENCES

[1] V.M. Sidelnikov, "On mutual correlation of sequences," *Soviet Math Doklady*, vol. 12, pp. 197-201, 1971.

[2] L.R. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 397-399, May 1974.

[3] E.M. Gabidulin, "On Classification of Sequences with the Perfect Periodic Auto-Correlation Function," *Proceedings of the third International Colloquium on Coding Theory*, Sept. 25 - Oct. 2, 1990, Dilijan, pp. 24-30, Yerevan, 1991.

[4] E.M. Gabidulin, "A Family of PSK-Sequences with the Perfect Periodic Auto-Correlation Function", *Proceedings of the Fifth Soviet-Swedish Workshop on Information Theory "Convolutional Codes; Multi-User Communication"*, January 13-19, pp. 69-72, Moscow, USSR.

[5] R.L. Frank, "Polyphase Codes with Good Nonperiodic Correlation Properties," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 43-45, January 1963.

[6] D.V. Sarvate, "Bounds on Crosscorrelation and Autocorrelation of Sequences," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 720 - 724, November 1979.

Geometrically Uniform Multidimensional PSK Constellations

S.Benedetto, R.Garello, M.Mondin, G.Montorsi
Dipartimento di Elettronica, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract

The theory of geometrically uniform (GU) codes is applied to the case of multidimensional (MD) PSK constellations. The symmetry group of an $L \times MPSK$ is completely characterized. Conditions for rotational invariance of GU partitions of a signal constellation are illustrated. Through suitable algorithms, "good" GU partitions of $L \times MPSK$ (M=4,8,16 and L=1,2,3,4) constellations are found. They are used as starting points in the search for good GU trellis codes.

1 GU TCM SCHEMES

A signal set S is GU [1] if it has a transitive symmetry group $\Gamma(S)$, i.e. if for any two points s and s' in S, there exists a symmetry of S that sends s to s'. A *generating group* $G(S)$ of S is a subgroup of $\Gamma(S)$ which is minimally sufficient to generate S starting from an arbitrary initial point of it. The MPSK Constellation is GU, its symmetry group is isomorphic to the dihedral group D_M and, in the case of M even, the only two possible generating groups are isomorphic to Z_M and $D_{M/2}$. GU signal sets have the important property that the Voronoi regions are congruent, so that the error probability is independent of which signal was transmitted. In [1] this property was shown to hold for signal sequences too, through a suitable extension of the concept of geometrical uniformity. A normal subgroup G' of the generating group $G(S)$ induces a partition S/S' of the signal set S, in which each subset of the partition is GU and has G' as a common generating group. A one-to-one mapping is induced between the quotient group G/G' and the subsets of the partition S/S'. If we combine a linear code over the label group $A \simeq G/G'$, i.e. a subgroup of A^l (with l possibly infinite) with the mapping $G/G' \rightarrow S/S'$ we obtain a GU code over S. As an example, a linear rate k/n binary convolutional code may be used if $G/G' \simeq (Z_2)^n$. The basis for a GU TCM code with good properties in terms of minimum Euclidean distance is a GU partition with a minimum squared Euclidean distance within signal sets at a given partition level as large as possible.

2 GU PARTITIONS OF MD PSK CONSTELLATIONS

We denote a multidimensional PSK constellation obtained through the L-fold Cartesian product of a 2D MPSK signal set with itself by LxMPSK. It contains M^L waveforms formed by L consecutive MPSK signals. We prove that the symmetry group of Lx4PSK constellations is isomorphic to $S_{2L} \bullet (Z_2)^{2L}$ and that of LxMPSK, M even larger than 4, is isomorphic to $S_L \bullet (D_M)^L$. Starting from the symmetry group we develop an algorithm able to construct all the possible generating groups of the constellation. In this way we find generating groups which are not simple Cartesian products of the generating groups of the constituent MPSK constellation. We call $G = G_0/G_1/\ldots/G_{n-1}/G_n$ a *binary partition chain* of a group G with $|G| = 2^n$ if $G_n, \ldots, G_1$ are normal subgroups of G and $|G_p| = 2 \cdot |G_{p+1}| \ \forall p$. In order to select "good" (in some sense) GU partition chains of the constellation S, we need to associate to a given partition chain some important parameters like: the *minimum Euclidean intraset squared distance* δ_p^2 at the p-th partition level, the *isomorphism* of both the normal subgroup generating the partition and the quotient group, and the *rotational invariance* of the partition chain at its various levels. Given S =MPSK we denote by r_k the rotation by $k\frac{360}{M}$ degrees with respect to the origin and by r_k^L the simmetry of S^L =L×MPSK obtained through L Cartesian products of r_k by itself. Introducing the subgroup of $\Gamma(S^L)$ called the *Rotationally Invariant subGroup*: $RIG(S^L) = \{1, r_1^L, (r_1^L)^2, \cdots, (r_1^L)^{M-1}\} = < r_1^L > \cong Z_M$, we say that a partition is *congruent* with respect to $r_k^L \in RIG(S^L)$ if r_k^L induces a permutation among the partition subsets, and *(rotationally) invariant* with respect to r_k^L if this permutation reduces to the identity. Necessary and sufficient conditions for the congruence and the invariance of a partition are stated. When $RIG(S^L) \subseteq G(S^L)$ the partitions are automatically congruent with respect to all $r_k^L \in RIG(S^L)$ and invariant with respect to r_k^L iff $r_k^L \in G_i$. An algorithm is illustrated which scans all the possible binary partition chains starting from a given generating group G. It constructs the tree of all possible binary partition chains induced by normal subgroups of G, identifies each partition level through the parameters aforementioned (minimum Euclidean distance, isomorphisms and rotationally invariance), and chooses the best partition chains as paths through the subgroup tree according to optimality criteria related to the previous parameters. Every partition chain is identified like in Table 1.

3 SEARCH FOR GOOD GU TCM CODES

The partitions tables obtained are used to find "good" GU TCM schemes based on binary as well as more general group convolutional codes. The obtained codes, as well as their performance, are presented in [2] and [3]. As an example, in Table 2 the results of the search for binary 3x8PSK GU codes transmitting 2.33 bit/T for increasing complexity are presented. Some of them improve over known non-GU codes. As for more general group codes, in Table 3 GU TCM codes for 3x8PSK based on the group Z_8^3 and transmitting 2 bits/T are presented. They present good characteristics both in terms of Euclidean distance and rotational invariance. Error event probability curves for these codes are shown in Figure 1.

This work was supported by Italian National Research Council (CNR) under "Progetto Finalizzato Trasporti" (Prometheus)

References

[1] G.D. Forney, Jr., "Geometrically Uniform Codes", *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1241-1260, September 1991.

[2] S. Benedetto, R. Garello, M. Mondin and G. Montorsi, "Geometrically Uniform Partitions of Multidimensional PSK Constellations and Related Binary Codes", *submitted for publication*, October 1992.

[3] S. Benedetto, R. Garello, M. Mondin and G. Montorsi, "Geometrically Uniform TCM Codes over Groups Based on Multidimensional PSK Constellations", *submitted for publication*, October 1992.

Signal set 3x8PSK				Generating Group $G(S) = (Z_8)^3$	
level p	generator	δ_p^2	Rot. inv.	Isomorphism of G_p	Isomorphism of $G(S)/G_p$
9	000	-	360	I	$(Z_8)^3$
8	444	12.000	180	Z_2	$Z_4 \times (Z_8)^2$
7	222	6.000	90	Z_4	$(Z_2) \times (Z_8)^2$
6	151	4.586	90	Z_8	$(Z_8)^2$
5	004	4.000	90	$Z_2 \times Z_8$	$Z_4 \times Z_8$
4	002	2.000	90	$Z_4 \times Z_8$	$Z_2 \times Z_8$
3	040	1.757	45	$Z_2 \times Z_4 \times Z_8$	$Z_2 \times Z_4$
2	020	1.757	45	$(Z_4)^2 \times Z_8$	$(Z_2)^2$
1	001	0.586	45	$Z_4 \times (Z_8)^2$	Z_2
0	010	0.586	45	$(Z_8)^3$	I

Table 1:

3x8PSK TCM codes					$R_{eff} = 2.33$ bit/T		
N	k	Inv.	d_{free}^2	N_{free}	d_{next}^2	N_{next}	γ(dB)
2	1	90	2.000	6	2.929	32	0.56
4	2	180	2.929	16	3.172	12	2.22
8	2	180	3.757	24	4.000	15	3.30
16	2	180	4.000	15	4.343	24	3.57
32	3	180	4.000	7	4.343	20	3.57
64	4	180	4.000	3	4.343	14	3.57

Table 2:

3x8PSK TCM codes					$R_{eff} = 2$ bit/T		
N	k	Inv.	d_{free}^2	N_{free}	d_{next}^2	N_{next}	γ(dB)
2	1	90	3.172	12	4.000	7	2.00
4	2	45	4.000	3	4.343	8	3.01
8	3	45	4.586	2	4.929	2	3.60
16	4	90	5.757	8	6.000	2	4.59
32	4	45	6.000	2	6.343	14	4.77
64	5	45	6.101	2	6.343	2	4.84

Table 3:

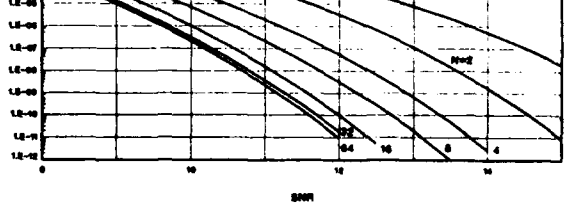


Figure 1:

HIGH-RATE PUNCTURED CONVOLUTIONAL CODES FOR TRELLIS-CODED MODULATION*

François Chan and David Haccoun
Département de génie électrique et de génie informatique
Ecole Polytechnique de Montréal
C.P. 6079, succ. A, Montréal, H3C 3A7

Abstract

The encoding and decoding advantages of high-rate punctured binary convolutional codes over memoryless channel are well known. The puncturing technique is applied to Trellis-Coded Modulation, resulting in simplified Viterbi decoding at the cost of a small reduction in coding gain compared to usual Viterbi decoding. Using computer search, short-memory rate 2/3 punctured codes with the same minimum free Euclidean distance as Ungerboeck's optimum codes have been found; these codes provide the same error performance when decoded in the usual manner. In addition to the decoding advantages, puncturing provides greater flexibility, allowing an easy implementation of variable bandwidth efficiency systems.

Summary

Trellis-Coded Modulation (TCM) by using an expanded signal set can yield significant coding gains of 3 to 6 dB over uncoded modulation without requiring more bandwidth [1]–[3]. A binary convolutional code of rate $R = m/(m+1)$ is used and the encoded symbols are mapped into channel signals by following a set of rules designed to maximize the Euclidean distance [1]. When decoding TCM signals with the Viterbi algorithm [3], at each trellis level, among the different paths merging into a given state, only the most likely path, or survivor, is kept. For a rate $R = m/(m+1)$ code, selecting the survivor among the 2^m paths merging at each state requires $(2^m - 1)$ binary comparisons per state. If the number of states is large and if the coding rate is high (i.e., $m > 3$), then clearly, Viterbi decoding in this usual manner may become impractical.

It is well known that for convolutional codes puncturing allows considerable simplifications of the encoding and decoding processes [4], [5]: decoding a rate $R = m/(m+1)$ punctured code requires only m binary comparisons instead of the $(2^m - 1)$ comparisons that are required by the usual decoder. As m increases, the savings are substantial while resulting in only a slight performance loss [4], [5] as compared to the best known $R = m/(m+1)$ codes.

In this paper we present an application of the same technique to Trellis-Coded Modulation. If the underlying convolutional code of the TCM scheme is a rate $R = m/(m+1)$ code and if there are no transmitted uncoded bits (i.e., no parallel transitions), a code of the same rate can be obtained by puncturing an original low-rate $R = 1/m$ code. Naturally, the original low-rate code and the puncturing pattern that will produce the TCM code with the maximum free Euclidean distance have to be determined. A computer search has provided codes with up to 64 states which, when punctured, result in rate 2/3 codes for 8–PSK modulation with the same free distance as Ungerboeck's codes. The advantage in using this technique is that by changing the puncturing pattern only, codes with coding rates R=1/2, 2/3 or 3/4 can be easily obtained from the same original code.

With punctured binary convolutional codes, simplified decoding is obtained because at each state, whether it is an "intermediate" state or a "true" state[1], a decision about the survivor can be made. Although it is not as straightforward as for binary codes, the same process can be applied to TCM with a non-Ungerboeck set partitioning method by using approximate metrics at intermediate states. This results in the same complexity savings as for binary convolutional codes but the coding gain is slightly lower than with usual decoding: about 0.15 dB degradation for 64 states, 8–PSK modulation. This degradation of the coding gain is caused by the metrics approximation at intermediate states and tends to decrease as the free Euclidean distance increases.

Reducing the decoding complexity of a high-rate code for TCM is particularly important when the number of states is large. The puncturing technique provides an attractive alternative to the usual approach, allowing a reduction in the number of binary comparisons by a factor of $((2^m - 1)/m)$ for a rate $R = m/(m+1)$ code. When the number of states becomes too large to be practical for the Viterbi algorithm, the use of the puncturing technique and suboptimum algorithms, such as sequential decoding or the Adaptive Viterbi Algorithm [6] can be combined to reduce further the complexity at a small cost to the error performance.

References

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[2] G. Ungerboeck, "Trellis-coded modulation with redundant signal sets — part II: State of the art," *IEEE Communications Mag.*, pp. 12–21, Feb. 1987.

[3] E. Biglieri, D. Divsalar, P. J. Mclane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. Macmillan Publishing Company, New York, 1991.

[4] J. B. Cain, G. C. Clark, and J. M. Geist, "Punctured convolutional codes of rate (n-1)/n and simplified maximum likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 97–100, Jan. 1979.

[5] D. Haccoun and G. Bégin, "High-rate punctured convolutional codes for Viterbi and sequential decoding," *IEEE Trans. Commun.*, vol. 37, pp.1113–1125, Nov. 1989.

[6] F. Chan and D. Haccoun, "Adaptive decoding of convolutional codes over memoryless channels," *submitted to IEEE Trans. Commun.*, Feb. 1992.

* This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada, the Fonds pour la formation des Chercheurs et l'Aide à la Recherche (FCAR) du Québec and by a grant from the Canadian Institute for Telecommunications Research under the National Centers of Excellence program of the Government of Canada.

[1] When decoding a rate R=b/v punctured code using the trellis of the low-rate original code, a state at level b (and any multiple of b), corresponding to b (or multiple of b) information bits is denoted a "true" state since it corresponds to a state of a true high-rate code. A state at any other trellis level is denoted an "intermediate" state.

ON THE DESIGN CRITERIA FOR TRELLIS CODES WITH SEQUENTIAL DECODING*

Fu-Quan Wang and Daniel J. Costello, Jr.

Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

ABSTRACT

Design criteria for trellis codes with sequential decoding are examined. A comparision of trellis codes with Optimum Distance Profile (ODP) and Optimum Free Distance (OFD) reveals that neither ODP nor OFD trellis codes result in the best trade-off between error performance and computational performance when sequential decoding is used. A new approach is proposed to construct robustly good trellis codes for use with sequential decoding. The new codes obtained using this approach achieve nearly the same free distances as the OFD codes and nearly the same distance profiles as the ODP codes.

SUMMARY

Most of the trellis codes constructed thus far have been for use with the Viterbi decoding algorithm[1,2]. However, the computational effort of the Viterbi algorithm grows exponentially with the code constraint length ν. This limits its application to codes with small values of ν and relatively modest coding gains. On the other hand, sequential decoding can perform almost as well as the Viterbi algorithm and its computational complexity is essentially independent of ν. Thus, more coding gain is possible when larger constraint length codes are used with sequential decoding. In [3,4], it has been shown that sequential decoding is a good alternative to the Viterbi algorithm for decoding trellis codes. However, no papers have addressed the problem of constructing trellis codes for use with sequential decoding. In this paper, trellis codes with Optimum Distance Profile (ODP) and Optimum Free Distance (OFD) are examined and design criteria for trellis codes with sequential decoding are discussed. We show that neither the ODP nor the OFD trellis codes provide the best trade-off between distance profile and free distance. Thus, a new algorithm is proposed to construct robustly good trellis codes for use with sequential decoding.

First, trellis codes with optimum distance profiles were constructed. In the construction algorithm, the free distance was used as a secondary criterion, i.e., the code having the larger free distance is retained whenever two codes have the same distance profile. Compared with the Ungerboeck codes, we found that the ODP trellis codes have much smaller free distances for some constraint lengths. For example, the free distance of ODP trellis coded 8-PSK with $\nu = 7$ is only 4.0 compared with 6.59 for the Ungerboeck code. This results in a reduction of more than 2.0 dB in asymptotic coding gain. Thus, it appears that ODP codes do not provide a good trade-off between free distance and distance profile.

We then conducted exhaustive searches for OFD trellis codes in which the distance profile was used as a secondary criterion. Our results indicate that the OFD trellis codes do not provide the best trade-off between distance profile and free distance, either. For example, the ODP and OFD trellis coded 8-PSK with $\nu = 7$ have distance profiles $(d_0^2, d_1^2, \cdots, d_7^2) = (2.0, 2.59, 2.59, 3.17, 3.17, 3.76, 3.76, 4.0)$ and $(2.0, 2.0, 2.59, 2.59, 2.59, 2.59, 3.17, 3.17)$, respectively. Note that the OFD code has a much worse distance profile than the ODP code.

Thus, we have constructed trellis codes which are neither optimum free distance nor optimum distance profile. We call the new codes robustly good trellis codes. Given that a robustly good trellis code of constraint length ν has been found, the approach used to find a constraint length $\nu + 1$ robustly good trellis code is to find the code that improves the free distance or the distance profile of the constraint length ν code, with priority given to improving the free distance. In other words, we try to find a longer code which has a free distance or a distance profile superior to or identical to the shorter one. Systematic feedback 8-PSK and 16-QAM robustly good trellis codes with ν up to 15 and asymptotic coding gains up to 6.66 dB are obtained using this approach. Compared to ODP and OFD trellis codes, the robustly good trellis codes provide a much better trade-off between free distance and distance profile. Indeed, the new codes achieve nearly the same free distances as the OFD codes and nearly the same distance profiles as the ODP codes.

References

[1] G. Ungerboeck, "Trellis Coded Modulation with Redundant Signal Sets, Part II: State of the Art", *IEEE Commun. Mag.*, **Vol. 25**, pp. 12–22, February 1987.

[2] J. Porath and T. Aulin, "Algorithmic Construction of Trellis Codes," submitted to the *IEEE Trans. Commun.*, November 1990.

[3] G. J. Pottie and D. P. Taylor, "A Comparison of Reduced Complexity Decoding Algorithms for Trellis Codes," *IEEE J. Sel. Areas Commun.*, **SAC–7**, pp. 1369–1380, December 1989.

[4] F. Q. Wang and D. J. Costello, Jr., "Erasurefree Sequential Decoding of Trellis Codes", submitted to the *IEEE Trans. Inform. Theory*, November 1992.

*This work was supported by NSF grant NCR 89-03429 and NASA grant NAG 5-557.

Practical Trellis Coded Modulation with Punctured Rate-2/3 Convolutional Codes

Stephen K. How

General Instrument, San Diego, CA

Abstract - *A trellis coded modulation scheme is described which uses a punctured rate-2/3 convolutional code to simplify decoder implementation. Simulations of a code based on a punctured 64-state rate-1/2 trellis show comparable performance to a 32-state Ungerboeck code (rate-2/3 trellis).*

Two-dimensional trellis coded modulation (TCM) achieves up to 6dB coding gain by partitioning the symbol constellation 8 ways, which increases uncoded symbol spacing by $2\sqrt{2}$. An Ungerboeck code maps two coded bits to the 8 subsets using a rate-2/3 convolutional code [1]. In high-speed decoders, the Viterbi algorithm (VA) is implemented in a parallel manner, and the complexity of the 2/3 trellis can limit the number of states to 16. In satellite applications 64-state rate-1/2 Viterbi decoders are commonly built as ASICs.

A rate-1/2 trellis with every other branch punctured specifies a rate-2/3 code. A trellis code with the same spectral efficiency as an Ungerboeck code needs to transmit 2 coded bits / symbol. Two bits are encoded in two steps through the rate-1/2 trellis, generating an unpunctured and punctured branch.

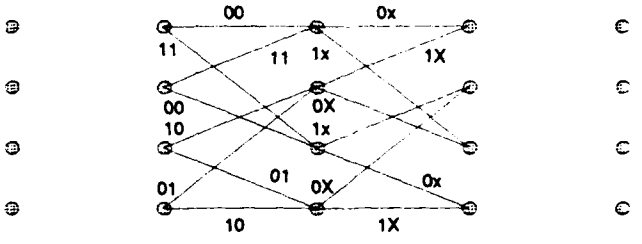


Figure 1. Punctured 4-state trellis

In the punctured TCM scheme, these 2 branches are mapped to a subset of an 8-way partitioned QAM constellation. A mapping and decoding method are needed to assign large Euclidean distances to the error events of the punctured trellis. In the example of the 4-state trellis in figure 1, the unpunctured branch output defines the 2 MSBs and the punctured branch defines the LSB of the symbol index in figure 2. In this manner, 2 coded bits define a subset. Uncoded bits define the subset member.

In decoding, the branch metrics for the unpunctured trellis step are first computed and applied to the VA. Using the same receive symbol, the punctured metrics are then computed to decode the second coded bit. The symbol mapping and decoding are an attempt to orthogonalize these two steps. The first set of branch metrics are computed approximately as the minimum distance^2 between the receive symbol and the subset points grouped by index MSBs. I.e., BM_{00} = min $\{d^2(A_n,s),\ s \in A \cup B\}$, BM_{01} = min $\{d^2(A_n,s),\ s \in C \cup D\}$, etc. where A_n is the receive symbol, and $d^2(\cdot,\cdot)$ is squared Euclidean distance. Actual metrics are based on the log of conditional probabilities. The punctured metrics used in the second trellis step are computed similarly, with the same receive symbol and different subset unions. In figure 1 two types of punctured branches are distinguished to enhance the Euclidean distance at this trellis step. The branches labeled by puncture "x" (lower-case) are "continuations" of the 3-bit branches 00- and 11-. Also puncture "X" implies the branch is the LSB of subset indexes 01- and 10-. Accordingly, BM_{0x} = min $\{d^2(A_n,s),\ s \in A \cup G\}$, BM_{1x} = min $\{d^2(A_n,s),\ s \in B \cup H\}$, BM_{0X} = min $\{d^2(A_n,s),\ s \in C \cup E\}$, and BM_{1X} = min $\{d^2(A_n,s),\ s \in F \cup D\}$. This distinction between punctured branches provides a larger Euclidean distance mapped to subset LSB, as shown by the shaded and unshaded sets in figure 2.

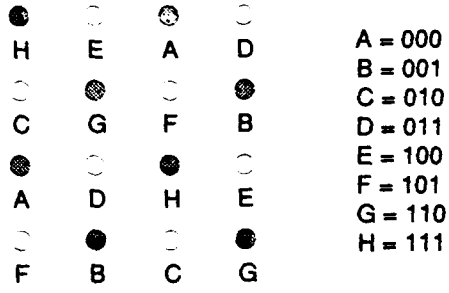


Figure 2. 16QAM partition

The described mapping approximates an assignment of $2\sqrt{2}\,\Delta_0$ to 3-bit branch 111 (from 000), $\sqrt{2}\,\Delta_0$ to 110 and 001, and Δ_0 to 010, 101, 100, and 011. From this assumption, an optimal punctured 64-state rate-1/2 code was found to be (101, 109, 101) (octal), yielding $d_e^2/\Delta_0^2 = 7$, or 5.45 dB asymptotic coding gain. Figure 3 shows simulated performance of the punctured TCM approach vs. Ungerboeck codes for 16QAM.

[1] G. Ungerboeck, "Trellis-coded Modulation with Redundant Signal Sets Part II", *IEEE Communications Magazine*, vol. 25, no. 2, Feb. 1987.

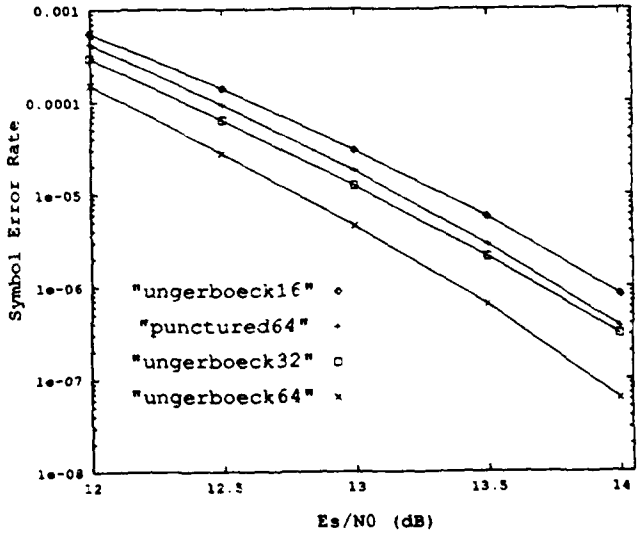


Figure 3. Comparison of punctured with Ungerboeck codes

Design of Optimal Filters for Use as Bandwidth-Efficient Coded Modulation

Amir Said* and John B. Anderson

Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180

Recent work has shown that in channels with intersymbol interference (ISI), whether finite or infinite response, it is possible to achieve almost maximum-likelihood detection performance with reduced-search algorithms. The number of operations required by those algorithms may be orders of magnitude smaller than that required by the Viterbi algorithm. Hence, controlled ISI, such as that introduced by a band-limitation filter, can be used to improve performance without an exponential increase in the detection complexity. This is actually a form of coding where, for a fixed noise immunity performance, the gains are measured in bandwidth reduction.

In a typical application of bandwidth-efficient coded modulation, ISI may be introduced by the non-ideal response of the channel and by intentional filtering at the modulator output to constraint the bandwidth. This is modeled in Fig. 1. The filter $f(n)$ may comprise an explicit coded modulation, for which we seek the optimal design. We propose a method that simultaneously constrains the bandwidth and maximizes the minimum Euclidean distance between signals. We show that it can be formulated as a linear program; and it allows uncoded or trellis coded data, filters with infinite impulse response, and many types of spectrum shaping constraints (e.g., zeros at $f = 0$ or Chebyshev filters). The proposed filter can also be considered a convolutional coder that *matches* the code, output and channel filters for better performance.

In Fig. 1, the discrete-time FIR filter is used for spectral shaping in the Nyquist frequency interval and to maximize the minimum Euclidean distance. The modulator output filter is used to steeply attenuate the frequencies outside the desired bandwidth; and $h_c(t)$ is the response of the linear channel. For now, we use a simple definition of the bandwidth W, where a fixed and small fraction of the modulator output power is outside the frequency interval $[-W, W]$.

Mathematical Formulation

Here we assume an ideal channel, i.e., $h_c(t) = \delta(t)$, but the generalization is straightforward. The impulse response

$$h(t) = \sum_{n=0}^{L-1} f(n) h_o(t - nT),$$

is used to define the modulator output

$$s(t, u) \doteq \sum_n u(n) h(t - nT),$$

where $u(n)$ is the complex data sequence.

We define the correlations

$$g_f(n) \doteq \sum_k f(k+n) f^*(k),$$

$$g_o(n) \doteq \int_{-\infty}^{\infty} h_o(t + nT) h_o^*(t)\, dt.$$

*This research was partially supported by CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasil.

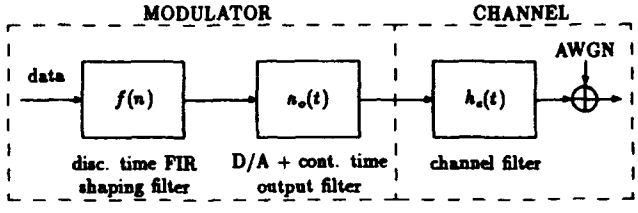


Figure 1: Modulator/channel model.

The objective is to find the filter taps, $f(n)$, that maximize the minimum Euclidean distance subject to the bandwidth constraint, but the problem is solved by finding the optimal correlation $g_f(n)$; the optimal $f(n)$ can be obtained from $g_f(n)$ via spectral decomposition.

It can be shown that the squared Euclidean distance between a transmitted and an erroneous sequence is

$$D_e^2 = \sum_n g_f(n) \mu_e^*(n),$$

where $e(n)$ is the difference of the two data sequences and

$$\mu_e(n) \doteq \sum_k g_o(n-k) \sum_m e(m+k) e^*(m).$$

The average energy per symbol is set by the linear constraint

$$\sum_m g_f(m) g_o^*(m) = 1,$$

and the fraction of the power inside the bandwidth is

$$E_w = \int_{-W}^{W} |H(f)|^2\, df = \sum_n g_f(n) \tau^*(n),$$

where

$$\tau(n) \doteq \int_{-W}^{W} |H_o(f)|^2 e^{j2\pi n f T}\, df.$$

Finally, $g_f(n)$ is a correlation sequence only if

$$\sum_n g_f(n) e^{-j2\pi n f} \geq 0, \text{ for all } f \in [0, 1).$$

In a practical solution method, we use sets with a small number of carefully chosen error sequences ($\mathcal{E}$) and frequency points ($\mathcal{F}$). The resulting linear program is:

$$\begin{aligned} D^2_{\text{min,opt}} = \quad & \max_{g_f, x} \quad x \\ & s.t. \quad \textstyle\sum_n g_f(n)\, \mu_e^*(n) \geq x, \quad \text{for all } \mathbf{e} \in \mathcal{E}, \\ & \textstyle\sum_n g_f(n)\, g_o^*(n) = 1, \\ & \textstyle\sum_n g_f(n)\, \tau^*(n) = E_w, \\ & \textstyle\sum_n g_f(n)\, e^{-2\pi n f} \geq 0, \quad \text{for all } f \in \mathcal{P}. \end{aligned}$$

We present results on a variety of code-filters designed by this procedure. The decoding complexity was measured by M-algorithm tests.

On a Class of Constant Envelope Continuous Phase Modulation Schemes, Obtained by Imposing Continuous Phase Transitions on Trellis Coded Asymmetric PSK

Johan Udén, Göran Lindell
Telecommunication Theory
Lunds University, Box 118, S-221 00 Lund, Sweden

Abstract

The problem of how to construct information carrying continuous phase functions, which yield power and bandwidth efficient schemes, is addressed. The additive white Gaussian noise channel and coherent maximum likelihood sequence detection are assumed. Our approach is to use trellis coded asymmetric PSK schemes of low complexity and with good distance properties. Consecutive phase values of the phase sequences generated by these schemes are interconnected by a continuous function. Thus, a continuous phase function is obtained. In conventional full response CPM, the Euclidean distance is bounded by an error event two symbols long. The continuous phase schemes obtained here, have a shift-register state trellis structure. This guarantees long error events, thus the schemes have a potential for large Euclidean distances related to the number of states in the trellis. There are schemes within this class with power and bandwidth efficiencies (d^2_{min} versus 99% bandwidth) which are very good, considering the low complexity of the schemes. They are, in fact, competetive with and sometimes better than, some of the best coded continuous phase modulated schemes, of comparable complexity, previously published.

System Description

Our approach is to start with good coded asymmetric PSK schemes, [1], having a shift register state trellis. By asymmetric is meant that the phase values used are nonuniformly spaced, i.e. for asymmetric 4-PSK the set $\{0, \phi, \pi, \pi+\phi\}$ is used. Consecutive phase values of the phase sequences generated by these schemes are interconnected by a continuous function. Thus, a continuous phase function is obtained, see fig. 1. The shape of the phase transitions is the same in every symbol interval, but the amount of change in the phase during a symbol interval depends on the current data and the state of the encoder. The continuous phase function can be written as $\Psi(t, \underline{U}) = 4\pi \sum_{n=-\infty}^{\infty} h(U_n, \sigma_n) q(t - nT_s)$. $U_n \in \{0, 1, \ldots, M-1\}$ is the data that arrives at the modulator at $t = nT_s$; σ_n is the state of the encoder at $t = nT_s$; $q(t)$ is the phase response and equals 0 when $t < 0$ and $1/2$ when $t \geq T_s$. The amount of change in the phase during a symbol interval is $2\pi h(U_n, \sigma_n)$. $h(U_n, \sigma_n)$ is the modulation index associated with the transition in the trellis caused by the data U_n when the encoder is in the state σ_n. The choice $h(U_n, \sigma_n) = U_n h$ renders a scheme in the traditional CPM class, but in general $h(U_n, \sigma_n)$ is a nonlinear function. The transmitted signal is $s(t, \underline{U}) = \sqrt{\frac{2E_s}{T_s}} \cos(2\pi f_0 t + \Psi(t, \underline{U}))$; f_0 is the carrier frequency assumed to be much larger than $1/T_s$.

When continuous phase is imposed on a coded PSK scheme, consecutive values of $h(U_n, \sigma_n)$ are chosen so that $\Psi(t, \underline{U})$ coincides modulo 2π with the values of the original phase sequence at the end of each symbol interval. There are several, in fact infinitely many, possible choices of the modulation index for a specific phase transition. An extra M-ary delay element is needed in the shift-register, and the number of states in the new trellis, S, is M times larger than in the original one. The extra delay element is necessary because continuous phase demands knowledge of the phase both at the beginning of the current symbol interval and at the beginning of the next symbol interval, see fig. 1 and 2.

Within the obtained class of continuous phase modulated schemes, schemes of low complexity having power and bandwidth efficiency comparable to, and sometimes better than, the schemes given in [2,3,4] can be found. Examples of results for the symmetric case ($\phi = \pi/2$) are given in figure 3.

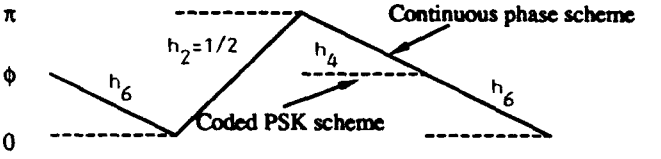


Figure 1. A continuous phase function obtained by interconnecting consecutive phase values of a coded PSK scheme.

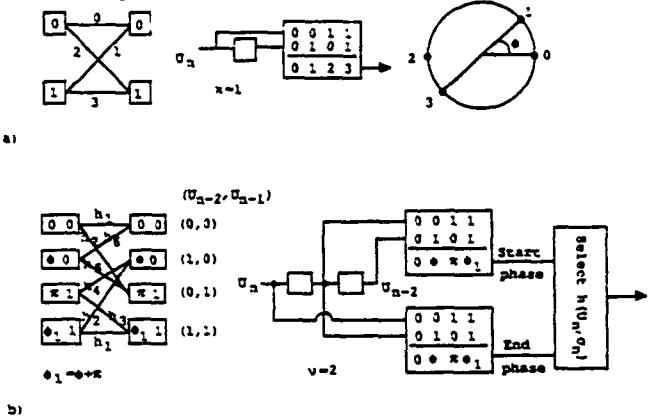


Figure 2. a) Trellis and a encoder of a coded asymmetric PSK scheme, see ref [1]. b) Trellis and encoder when continuous phase is imposed on the scheme in a). The label on a state transition is the modulation index used for that transition. The start phase is given as part of the state.

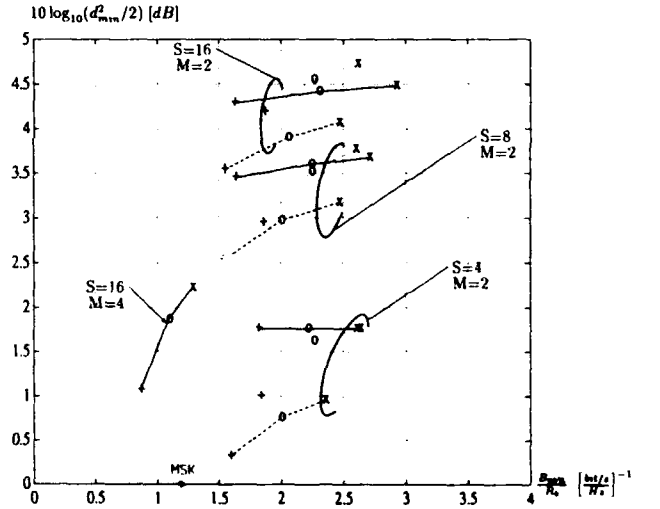


Figure 3. Asymptotic power gain over MSK plotted against the 99% in band power bandwidth. Schemes using the same set of modulation indices, but different frequency pulses, are connected with straight lines ('+' 1REC , 'o' 1HCS and 'x' 1RC). The 1REC schemes on the dashed lines have the same efficiency as schemes of the same complexity given in [2,3].

[1] D. Divsalar, M. K. Simon, and J. H. Yuen *Trellis Coding with Asymmetric Modulations* IEEE Transactions on communications, vol. COM-35, No. 2, February 1987

[2] B. Rimoldi *Design of Coded CPFSK Modulation Systems for Bandwidth and Energy Efficiency* IEEE Transactions on Communication, Vol. COM-37, No. 9, Sept.1989

[3] J. Huber, W. Liu, *Convolutional Codes for CPM Using the Memory of the Modulation Process*, IEEE Global Telecommunications Conference 1987 (GLOBECOM'87), Conference Record Vol. 3, pp. 43.1.1-43.1.5.

[4] J. P. Fonseka *Nonlinear Continuous Phase Frequency Shift Keying* IEEE Transactions on communications, vol. COM-39, No. 10, October 1991

A TRELLIS CODED MODULATION SCHEME CONSTRUCTED FROM BLOCK CODED MODULATION WITH INTERBLOCK MEMORY

Shang-Chih Ma Mao-Chao Lin
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan, Republic of China

Abstract

In this paper, we introduce a new Trellis Coded Modulation scheme with a two-fold dependency between signal points. In our coding scheme, in addition to the dependency among coded multi-dimensional signal points described by the trellis, each coded multi-dimensional signal point has another kind of dependency on one previously coded multi-dimensional signal point.

1 Preliminaries

Let C_a and C_{aa} be (n, k_a, d_a) and $(n, k_a + r, d_{aa})$ binary codes with generator matrices G_a and $[G_a^T G_{ar}^T]^T$ respectively. Also, let C_b and C_{bb} be (n, k_b, d_b) and $(n, k_b + r, d_{bb})$ binary codes with generator matrices G_b and $[G_b^T G_{br}^T]^T$ respectively. We can construct a $(2n, k_a + k_b + r)$ binary code C with generator matrix of the following form : $G = \begin{bmatrix} G_{br} & G_{ar} \\ 0 & G_a \\ G_b & 0 \end{bmatrix}$, where each 0 represents an all zero matrix.

Consider a BCM scheme with interblock memory[3]. Each two-dimensional signal symbol in the two-dimensional signal space W_0 is labeled by three bits (a, b, c) as shown in Figure 1. Let $\overline{v} = (a_1, b_1, c_1, \cdots, a_n, b_n, c_n)$ and $\overline{v}^* = (a_1^*, b_1^*, c_1^*, \cdots, a_n^*, b_n^*, c_n^*)$ represent two consecutively encoded $2n$-dimensional signal points in V_0. The combination of two adjacent $2n$-dimensional signal points, represented by $(\overline{v}, \overline{v}^*)$, may be called a superblock. In our scheme, $(c_1, \cdots, c_n)$ and $(c_1^*, \cdots, c_n^*)$ are codewords of an (n, k_c, d_c) binary linear code C_c. Moreover, $(b_1, \cdots, b_n, a_1^*, \cdots, a_n^*)$ is a codeword in C. Let $\overline{v}' = (a_1', b_1', c_1', \cdots, a_n', b_n', c_n')$ and $\overline{v}^{*'} = (a_1^{*'}, b_1^{*'}, c_1^{*'}, \cdots, a_n^{*'}, b_n^{*'}, c_n^{*'})$ be combined to represent another superblock. Suppose that $(a_1, \cdots, a_n) = (a_1', \cdots, a_n')$. If the condition of $min\{0.8 \cdot d_{aa} + 1.6 \cdot d_{bb}, 0.8 \cdot d_a, 1.6 \cdot d_b\} \geq 3.2 \cdot d_c$ is satisfied, the MSED between coded signal superblocks represented by $(\overline{v}, \overline{v}^*)$ and $(\overline{v}', \overline{v}^{*'})$ is $3.2 \cdot d_c$.

Example 1 : Let $n = 4$. Let C_a, C_{aa}, C_b, C_{bb} and C_c be $(4,1,4)$, $(4,3,2)$, $(4,2,2)$, $(4,4,1)$ and $(4,4,1)$ binary linear block codes respectively. As a result, $D_1^2 = D_2^2 = D_3^2 = D_4^2 = 3.2$. The average coding rate is 9/4 information bits per two-dimensional signal symbol. Compared to uncoded QPSK, the asymptotic coding gain is 2.55 dB.

2 The Proposed Coded Modulation Scheme

We now illustrate the procedure of introducing interblock memory to the TCM constructed from BCM by modifying example 1.

Let V_1 represent a 16-dimensional signal space, in which each 16-dimensional signal point is labeled by $(a_1, b_1, c_1, \cdots, a_4, b_4, c_4, a_1^*, b_1^*, c_1^*, \cdots, a_4^*, b_4^*, c_4^*)$, where a_1, a_2, a_3 and a_4 are fixed. Here the two blocks $(a_1, b_1, c_1, \cdots, a_4, b_4, c_4)$ and $(a_1^*, b_1^*, c_1^*, \cdots, a_4^*, b_4^*, c_4^*)$ are separated by 18 blocks. Hence, these two blocks are not adjacent. Since the 2-dimensional signal space W_0 is the 8-AMPM signal space, we see that V_1 is a (8,20,0.8) block modulation code Let V_3 be a subset of V_1, for which the partial labeling $(b_1, \cdots, b_4, a_1^*, \cdots, a_4^*)$ of each 16-dimensional signal point is a codeword of C. Thus, V_3 is a (8,17,3.2) block modulation code. We may partition V_1 into the disjoint union of 8 cosets of V_3.

Let V_2 be a subset of V_1. It can be constructed such that V_2 is a (8,19,1.6) block modulation code. The partition chain $V_1/V_2/V_3$ has increasing intraset MSED of 0.8, 1.6 and 3.2 respectively. With the partition chain of $V_1/V_2/V_3$, we can design an efficient TCM constructed from BCM with additional interblock memory. During the encoding, each time we encode an 11-bit message $\overline{m} = (m_1, m_2, \cdots, m_{11})$ into an 8-dimensional signal point represented by $(a_1, b_1, c_1, \cdots, a_4, b_4, c_4)$, where $(a_1, \cdots, a_4)$ was determined in an earlier encoding time. In the meantime, the $(a_1^*, \cdots, a_4^*)$ part of another 8-dimensional signal point is also determined for later usage. The message bits m_6 and m_7 are used as the input of a (3,2,3) convolutional code encoder and generate the output bits u_0, u_1, u_2, which are then used to select one of the eight cosets of C. The message bits $m_1, m_2, \cdots, m_5$ are used to choose a codeword $(b_1, \cdots, b_4, a_1^*, \cdots, a_4^*)$ from the selected coset of C. The message bits $m_8, \cdots, m_{11}$ are used to determine the codeword $(c_1, \cdots, c_4)$ of C_c. In the trellis, the branches emanating from the same state or merging into the same state all belong to the same coset of V_2. Thus, the MSED between any two code paths is 3.2. The coding rate of this coded modulation scheme is 11/4 bits per two-dimensional signal symbol. Compared to uncoded QPSK, the asymptotic coding gain is 3.42 dB.

References

[1] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. on Information Theory*, IT-28, No.1, pp.55-67, Jan. 1982.

[2] J.M. Wu and S.L. Su, "A Combination of Block Coded Modulation and Trellis Coded Modulation," presented at 1990 International Symposium on Information Theory and Its Applications, Hawaii, USA, November 27-30, 1990.

[3] M.C. Lin and S.C. Ma, "A Coded Modulation Scheme with Interblock Memory," to appear on IEEE Trans. on Communications.

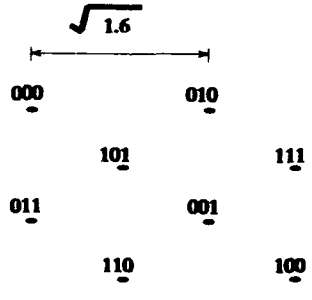


Figure 1. 8-AMPM Signal Set

UNIVERSAL SCHEMES FOR SEQUENTIAL DECISION FROM INDIVIDUAL DATA SEQUENCES

Neri Merhav and Meir Feder

Deptartment of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL

Department of Electrical Engineering - Systems
Tel Aviv University,
Tel Aviv 69978, ISRAEL

Abstract

Sequential decision algorithms are investigated for individual data sequences, with various application areas. Simple universal schemes are known to approach optimality as fast as $n^{-1}\log n$, where n is the sample size. For the finite-alphabet case, schemes that are implementable by finite-state machines (FSM's), are studied. It is shown that Markovian machines with sufficiently long memory are nearly as good as any randomized FSM. For the continuous-valued case, a useful class of parametric schemes is discussed with application to the recursive least squares (RLS) algorithm.

Summary

Various problems in information theory and signal processing are associated with selecting a good *strategy* b_t for minimizing an additive loss function $\sum_{t=1}^{n} l(b_t, x_t)$. While the data $x_1, x_2, \ldots$ normally flows sequentially, the best strategy (within some class) for this sequence depends on the *entire* sequence, and hence cannot be anticipated. Nevertheless, it has been observed in some applications, that applying the best strategy for the data observed *so far* is asymptotically as good as the best fixed strategy that could have been chosen in retrospect. Moreover, the performance of this dynamic policy is within $O(n^{-1}\log n)$ close to optimality, uniformly for every possible n-sequence.

One example is sequential universal data compression. Let $x_1, x_2, \ldots, x_n$ be a binary string. Let $n_t(0)$ and $n_t(1)$ denote counts of '0' and '1', respectively, along the t first symbols. Define $p_t(x) = (n_t(x) + 1/2)/(t+1)$, $x = 0,1$, as the respective empirical probabilities of '0' and '1'. Then, it is well known that

$$\frac{1}{n}\sum_{t=1}^{n} -\log p_{t-1}(x_t) \le \frac{1}{n}\sum_{t=1}^{n} -\log p_n(x_t) + \frac{1}{2}\frac{\log n}{n} + O(\frac{1}{n}). \quad (1)$$

The left hand side is the normalized length of a codeword associated with a sequential Shannon encoder based on current empirical letter probabilities from data observed so far. The first term on the right is the empirical entropy associated with x^n, which corresponds to the minimum normalized codeword length associated with a fixed codebook that one could have achieved if he knew a-priori $\{p_n(x)\}_{x=0,1}$. The $O(n^{-1}\log n)$ term is the loss in performance due to sequentiality. Eq. (1) can be formalized as a sequential minimization problem, where $l(b,x)$ is $-\log b$ for $x=0$ and $-\log(1-b)$ for $x=1$, and where $b \in (0,1]$.

Another application of (1) is sequential gambling where at each round t the player doubles the fraction of the current capital S_t wagered on the next outcome, i.e., $S_{t+1} = 2bS_t$ if $x_{t+1} = 0$ and $S_{t+1} = 2(1-b)S_t$ if $x_{t+1} = 1$. It is easy to see that the exponential growth rate $n^{-1}\log S_n$ of the capital is the average of $1 - l(b, x_t)$, where $l(\cdot,\cdot)$ is as before.

Portfolio selection for optimal investment is an extension of the above described gambling problem, where S_t is distributed over m investment opportunities according to some portfolio $b \in \mathbb{R}^m$, a vector of weights summing to unity. The stock market at day t is given by a vector $x_t \in \mathbb{R}^m$ with components, x_t^i, representing the return per monetary unit allocated to stock i at day t. The yield per unit invested is the weighted average of return ratios, i.e., $b^\theta x_t$, where θ denotes transposition. Thus, the exponential growth rate $n^{-1}\log S_n$ of the capital is the time-average of $l(b, x_t) = \log(b^\theta x_t)$. In [1] a sequential portfolio selection scheme is proposed for bounded market vector sequences, which is again as good as the optimal fixed investment policy up to a term of $O(n^{-1}\log n)$. The proof in [1], however, relies heavily on special properties of the function $\log(b^\theta x)$.

In [2] a result in the same spirit is established for prediction of binary sequences, where predictors are sought that uniformly minimize the fraction of errors. The strategy b_t is an estimate $\hat{x}_{t+1}$ of x_{t+1} and $l(\hat{x}_{t+1}, x_{t+1})$ is the indicator function of an error. Again, the techniques in [2] are specific to this particular loss function.

These examples are all special cases of the *sequential compound decision problem* (SCDP), which was first presented by Robbins [3] and has been thoroughly investigated since then by many researchers. The setup of the SCDP is more general because it assumes that the observer sees noisy versions of $\{x_t\}$. Upper bounds have been developed in the literature (see, e.g., references in [4]) on the decay rate of the difference between the average loss associated with the best sequential strategy and that of the best fixed strategy. The scope of these results has been later extended (see, e.g., [5]) and sequential decision procedures have been developed whose performance is nearly as good as that of the best kth order *Markovian* (rather than fixed) strategy, i.e., the best strategy that depends on the k preceding outcomes. While the Markovian strategy is plausible when the sequence has a "Markov structure" [5], it has not yet been justified rigorously for a general sequence.

Our first result serves as a step towards such a justification. For simplicity, we assume $\{x_t\}$ to be directly accessible (without noise) as in the above examples, and we consider strategies that are implementable by a deterministic M-state machine. We extend Theorem 2 of [6] and show that for a sufficiently large k (independently of the data) and any M-state machine, the best kth order Markov machine performs within o as good as the M-state machine. This means that in the limit as $k \to \infty$, a Markovian machine is as good as the best deterministic FSM. As a result, one can gradually increase the Markov order at a logarithmic rate independently of the particular sequence, and guarantee convergence to the limit as $M \to \infty$ of the minimum loss attainable by M-state machines for an infinite sequence. This result further extends and it turns out that deterministic Markovian machines compete successfully with every *randomized* FSM in the sense of minimizing the expected value of $n^{-1}\sum_{t=1}^{n} l(b_t, x_t)$ where the expectation is with respect to the randomization. For more general performance criteria, however, it is demonstrated that this principle does not necessarily hold.

This property of Markovian strategies is then utilized in order to relate the least asymptotic loss achievable by FSM's over individual sequences to that of the probabilistic case where any limitations on the allowed nonanticipating strategies are relaxed. Specifically, following Algoet [6], where the Shannon-McMill-Brieman theorem has been extended to a general sequential decision problem under a stationary ergodic regime, we show that these two quantities agree with probability one over an infinite sequence.

Markovian schemes are useful also in continuous alphabet applications. One familiar example is prediction under the mean squared error (MSE) criterion, i.e., $l(b_t, x_t) = (x_t - b_t)^2$, where the the predictor b_t is given by a function $f(x_{t-k}, \ldots, x_{t-1})$ of the k most recent outcomes, e.g., a linear predictor, where $f(x_{t-k}, \ldots, x_{t-1}) = \sum_{i=1}^{k} c_i x_{t-i}$. The sequential version of this linear predictor leads to the recursive least squares (RLS) algorithm, which is here shown to be universal in the above sense. Another example is vector quantization where $x \in \mathbb{R}^m$ and $l(b,x) = d(x, Q_b(x))$, $d(\cdot,\cdot)$ being a distortion measure and $Q_b(\cdot)$ a quantization function with quantization cells and centroids parametrized by b. Again, by allowing b to depend on the k preceding samples (or their quantized versions), we can implement a family of vector quantizers with memory, e.g., feedback quantizers, predictive quantizers, finite-state quantizers, etc.

References

[1] T. M. Cover, "Universal Portfolios," *Math. Finance*, Vol. 1, No. 1, pp. 1-29, January 1991.

[2] M. Feder, N. Merhav, and M. Gutman, "Universal Prediction of Individual Sequences," *IEEE Trans. Inform. Theory*, Vol. IT-38 , No. 4, pp. 1258-1270, July 1992. July 1992.

[3] H. Robbins, "Asymptotically Subminimax Solutions of Compound Statistical Decision Problems," *Proc. 2nd Berkeley Symp. Math. Statist. Prob.*, pp. 131-148, 1951.

[4] N. Merhav and M. Feder, "Universal Schemes for Sequential Decision from Individual Data Sequences," submitted for publication.

[5] T. M. Cover and A. Shenhar, "Compound Bayes Predictors for Sequences with Apparent Markov Structure," *IEEE Trans. Syst. Man. Cybern.*, Vol. SMC-7, pp. 421-424, May-June 1977.

[6] P. H. Algoet, "The Strong Law of Large Numbers for Sequential Decisions under Uncertainty," preprint.

Some Results on Sequential Detection of Weak Signals*

V.N.S.Samarasooriya and P.K.Varshney
Department of Electrical and Computer Engineering
Room 121, Link Hall
Syracuse University
Syracuse, New York 13244-1240.

Abstract:
In this paper we present a truncated sequential test for the detection of weak signals in additive noise. Performance evaluations for both truncated and untruncated tests are considered, and numerical results are presented. We also develop a sequential test for weak signal detection with M-level quantization of observed data. Numerical results are presented for the cases when the signal is deterministic, and when the signal is stochastic with known probability density. A performance comparison of the sequential tests using quantized and unquantized data is also provided.

Summary:

Detection of a signal in additive noise is formulated as the hypothesis testing problem stated as follows:

$$H_1 : X_i = S_i + N_i \quad , \quad i = 1,2,\ldots\ldots,N \qquad (1)$$
$$H_0 : X_i = N_i \quad , \quad i = 1,2,\ldots\ldots,N$$

where $\underline{S} = (s_1, s_2, \ldots\ldots, s_n)^T$ is the signal sample sequence, and $\underline{N} = (n_1, n_2, \ldots n_N)^T$ represents the additive noise. $\underline{X} = (x_1, x_2, \ldots\ldots, x_N)^T$ represents the observed data vector. Hypothesis testing is implemented either as a fixed-sample-size (FSS) test or as a sequential test. The FSS test, involves the comparison of a likelihood ratio against a single threshold, while deciding in favor of either H_0 or H_1, and uses N observed data samples in the process. Hence, a decision is reached only after N observations have been received. A FSS test can be implemented using several methods including the Neyman-Pearson, and the Bayesian techniques. The detector threshold is designed according to the required detector performance, namely the probabilities of detection and false alarm. In comparison with the FSS test, the sequential test requires on the average, a smaller number of samples to reach a decision.A sequential test can be designed to minimize the average detection time. The Sequential Probability Ratio Test (SPRT) derived by Wald [1] is known to be the optimum sequential test.

The Sequential Probability Ratio Test can be stated as follows:

$$LR_n(\underline{X}) = \frac{f_{\underline{X}|H1}(\underline{x}/H_1)}{f_{\underline{X}|H0}(\underline{x}/H_0)} \begin{cases} \geq A & \Rightarrow \text{ accept } H_1 \\ \leq B & \Rightarrow \text{ accept } H_0 \\ \text{otherwise} & \Rightarrow \text{ continue test} \end{cases} \qquad (2)$$

where $LR_n(\underline{X})$ the likelihood ratio at the n-th stage of the sequential test, with n being a random variable. $f_{\underline{X}|H1}(\underline{x}|H_1)$ and $f_{\underline{X}|H0}(\underline{x}|H_0)$ are the multivariate density functions, of $\underline{X}$, conditioned on H_1 and H_0 respectively. A and B are the thresholds of the sequential test. For prespecified probability of false alarm α, and probability of miss β, approximate expressions for A and B are obtained in [1] as; $A = (1-\beta)/\alpha$, and $B = \beta/(1-\alpha)$.

Consider the detection problem of a random signal sequence with known multivariate density $f_S(\underline{S})$. The likelihood ratio for this case can be expressed as follows [2]:

$$LR_n(\underline{X}) = \int_{\underline{S}} \frac{f_N(\underline{X} - \underline{S})}{f_N(\underline{X})} f_S(S)\, d\underline{S} \qquad (3)$$

where $f_N(.)$ is the density function of the additive noise.

Under weak signal conditions $f_{\underline{X}|H1}(\underline{x}|H_1)$ is approximately equal to $f_{\underline{X}|H0}(\underline{x}|H_0)$. The likelihood ratio in (3), is therefore approximately equal to one. This introduces difficulties in the implementation of a likelihood ratio test. Alternatively, $f_N(\underline{X} - \underline{S})$ in (3) can be expanded in a Taylor series around $\underline{S} = 0$. Assuming that the signal is always small, and keeping only the first and second order terms in $\underline{S}$, we can obtain a more manageable form of the Likelihood ratio to implement the hypothesis testing problem. i.e. :

$$LR_n(\underline{X}) - 1 = \frac{1}{f_N(\underline{X})}\left(-\sum_{i=1}^{n} y_i (s_i)_{av} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (y_i y_j + z_{ij})(s_i s_j)_{av}\right)$$

$$\text{with} \quad y_i = -\frac{\partial}{\partial x_i}\log(f_N(\underline{X})) \qquad (s_i)_{av} = \int_{\underline{S}} s_i f_S(S)\, d\underline{S} \qquad (4)$$

$$(s_i s_j)_{av} = \int_{\underline{S}} s_i s_j f_S(S)\, d\underline{S} \qquad z_{ij} = \frac{\partial^2}{\partial x_i \partial x_j}\log(f_N(\underline{X}))$$

Sequential detection of weak signals using the Taylor series approximation of the likelihood ratio given above has been considered to a certain extent in [3]. Numerical results are not presented, and a performance comparison with the corresponding FSS test was not considered.Also, in the sequential test considered in [3], the number of samples required to terminate the test (the actual detection time) can become large. We develop a truncated sequential test for weak signal detection, based on the series approximation of the likelihood ratio,in order to avoid prolonged test durations. At stage N_T the cumulative likelihood ratio is compared against a single threshold T, and a decision is reached. N_T and the threshold T become design parameters of the truncated sequential test. Performance evaluations of both untruncated and truncated sequential tests, in terms of the Average Sample Number(ASN) function, and the operating characteristic function, are presented. The probability that an untruncated test would terminate before the corresponding truncated test is also presented.

We also consider a sequential test for weak signal detection with M-level quantization of observed data, based on the series approximation of the likelihood ratio. Quantization for sequential signal detection for non-weak signal situations has been considered in [6]. The optimal set of quantization thresholds is obtained by minimizing a weighted sum of the ASN under each hypothesis. Numerical results are presented for the case when the signal is deterministic. A performance comparison between the sequential test for weak signals based on unquantized observed data, and the sequential test using quantized data, is also presented.

References:

[1] A. Wald, *Sequential Analysis* , New York: Wiley, 1947.
[2] P. Rudnick, 'Likelihood Detection of Small Signals in Stationary Noise', *J. Appl. Physics*, vol 32, pp. 140-143, 1961.
[3] R.F.Dwyer, 'Robust Detection of Weak Signals in Undefined Noise using Acoustical Arrays', *J. Acous. Soc. of America*, vol 67, March 1980.
[4] S.A.Kassam, *Signal Detection in Non-Gaussian Noise*, New York:Springer-Verlag, 1988.
[5] S.Tanataratana and J.B.Thomas, 'Truncated Sequential Probability Ratio Test', *Information Sciences*, vol 13, pp. 283-300, 1977.
[6] S.Tanataratana andJ.B.Thomas, 'Quantization for Sequential Signal Detection', *IEEE Trans. Communications*, pp. 696-703, July 1977

*This work was supported by Rome Laboratory under contract No. F30602-89-C-0082

REDUCED-COMPLEXITY ITERATIVE MAXIMUM-LIKELIHOOD SEQUENCE ESTIMATION ON CHANNELS WITH MEMORY

J.W. Modestino
Electrical, Computer and Systems Engineering Department
Rensselaer Polytechnic Institute
Troy, New York 12180

Abstract

Existing maximum-likelihood sequence estimation (MLSE) schemes for channels with memory, resulting in intersymbol interference (ISI), have typically been implemented using the Viterbi algorithm (VA). For memoryless modulation schemes the resulting search complexity is $O(M^L)$, where M is the alphabet size and L is the length of the ISI span in channel signaling intervals. This complexity renders the VA impractical for large M and/or L. In this paper we describe the structure and properties of a novel reduced-complexity iterative MLSE scheme based upon the expectation-maximization (EM) algorithm. This reduced-complexity iterative MLSE scheme is shown to have complexity $O(LM)$ at each iteration. The approach provides an attractive alternative to the VA for large signaling alphabets and/or ISI span.

Summary

Existing maximum-likelihood sequence estimation (MLSE) schemes for linear channels with memory, resulting in intersymbol interference (ISI), have typically been implemented using the Viterbi algorithm (VA). The VA provides a structured dynamic programming search of the underlying trellis defined by the modulator/channel cascade. For memoryless modulation schemes the resulting search complexity is of the order M^L where M is the alphabet size and L is the length of the ISI span, or the delay dispersion, measured in channel signaling intervals. For modulation schemes with memory, or for coded systems operating on ISI channels, the associated complexity can be considerably greater than this. Thus, it's of some interest to develop reduced-complexity MLSE techniques and a host of research efforts have been directed at this problem, all with varying degrees of success.

In the meantime, a fair amount of work has been done, mostly in the statistics literature, in developing iterative solutions to a variety of ML estimation problems which can be cast in terms of an incomplete data problem. Here, the observations, called the *incomplete* data, are related to another quantity, called the *complete* data, for which the ML estimation problem is simpler. The estimation-maximization (EM) algorithm [1] has been used in such situations to obtain an iterative solution to the original ML estimation problem, based on the incomplete data, which at each iteration is no more complex than obtaining a ML solution of the much simpler problem based on the complete data. The EM algorithm has found extensive applications in a variety of problem areas including: spectral estimation [2], image reconstruction [3], and image segmentation [4],[5]. Recently, the EM algorithm has been applied to several communications problems including: problems of carrier recovery [6] and channel state estimation[7]. Experience has generally demonstrated rapid convergence of the EM algorithm and, since each iteration is reasonably simple to implement, this generally leads to substantial computational savings relative to straightforward ML procedures.

In the present paper we apply the EM algorithm to the problem of MLSE on linear ISI channels. This results in an iterative algorithm with substantial computational savings over conventional full-search MLSE approaches since we exploit, at each stage, the rather simple structure of the ML solution based upon the associated complete data. While there are many ways to relate the observations to a corresponding complete data quantity, the particular formulation we consider is suggested by related work on the ML parameter estimation problem for superimposed signals [8] which is directly applicable to the MLSE problem treated here.

In this work we provide the formal development of the proposed reduced-complexity MLSE approach and describe some of its performance characteristics. The relative complexity advantages of this scheme depends, of course, on how many iterations are required for acceptable convergence. This is related to the resulting error probability and is best determined by simulation. We provide simulation results demonstrating the rapid convergence properties of this reduced-complexity iterative MLSE scheme.

References

[1] A.D. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood From Incomplete Data via the EM Algorithm," J. Roy. Stat. Soc., vol. 39, pp. 1-38, 1977.

[2] M.J. Miller and D.L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," IEEE Proc., vol. 75, pp. 892-907, July 1987.

[3] A.K. Katsaggelos and K.T. Lay, "Maximum-Likelihood Identification and Restoration of Images Using the Expectation-Maximization Algorithm," Chap. in *Digital Image Restoration*, Ed. A.K. Katsaggelos, Springer-Verlag, 1991.

[4] J. Zhang, J.W. Modestino and D.A. Langan, "Maximum-Likelihood Parameter Estimation for Unsupervised Model-Based Image Segmentation," to appear in IEEE Trans. Sig. Proc.

[5] D.A. Langan, K.J. Molnar, J.W. Modestino and J. Zhang, "Use of the Mean-Field Approximation in an EM-Based Approach to Unsupervised Stochastic Model-Based Image Segmentation," Proc. of IEEE ICASSP'92, San Francisco, CA, pp. 57-60, March 1992.

[6] C.N. Georghiades and J.C. Han, "Sequence Estimation in the Presence of Phase-Errors via the EM Algorithm," submitted to IEEE Trans. on Commun.

[7] J.W. Modestino, "Use of the EM Algorithm for Incorporating Channel State Information into Decoding Procedures," ECSE Dept. Report, Rensselaer Polytechnic Institute, Feb. 1992.

[8] M. Feder and E. Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," IEEE Trans. Acoust., Speech, Sig. Proc., vol. ASSP-36, pp. 477-489, April 1988.

On Sequential Delay Estimation in Wideband Digital Communication Systems *

Yossef Steinberg and H. Vincent Poor
Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA

ABSTRACT

The problem of estimating the symbol timing in wideband data communication signals is considered. Conventional approaches to this problem suffer from several drawbacks, including possible lack of consistency due to multiple extrema in the error surface, and very slow convergence due to exceedingly sharp waveform correlation functions. In this work, sequential estimation algorithms that alleviate these problems are constructed and analyzed. These algorithms are based on two techniques: the use of regularization (i.e., prefiltering) to produce a consistent initial estimate at the expense of higher mean-square error; and the coupling of recursive maximum-likelihood with this consistent estimator to produce the desired goal - a recursive consistent and efficient estimator.

1 Introduction and Summary

In this work we consider the problem of delay estimation in wideband binary digital communication systems. Let the received signal be modeled as

$$dr(t) = a\sum_{i\geq 0} b_i s(t - iT - \tau^*)dt + \sigma dw(t) \qquad (1)$$

where a is the received amplitude; b_i - i'th data bit ($\{b_i\}_{i\geq 0}$ is a sequence of iid equiprobable random variables in $\{-1,1\}$); T - duration of symbol interval; $s(t)$ - code waveform; τ^* - unknown delay, $\tau^* \in [0,T)$; $w(t)$ - standard Brownian motion; σ^2 - channel noise intensity. We assume that $s(t) = 0$ for $t \notin [0,T]$, and that it can be written as $s(t) = \sum_{i=0}^{N_c-1} \gamma_i \varphi_i(t - iT_c)$, where $\varphi_i(t)$ are basic "chip" waveforms, N_c is the number of chips in symbol interval and $\{\gamma_i\}$ is a maximal length sequence [1]. The problem addressed here is that of estimating the delay τ^* given the observation signal $\{r(t), t \geq 0\}$, where the receiver has a knowledge of a, σ, and the waveform $s(t)$. Denote by $c(\tau)$ the periodic autocorrelation function of $s(t)$. We assume throughout that $c(\tau)$ has no local maxima in the interval $[-T/2, T/2]$. Consider for a moment the idealized system where the iid data bits are known to the receiver (or estimator). In this case the maximum likelihood (ML) estimate of τ^* based on observations of the first n time units is the value of $\hat{\tau}$ that maximizes the output of a filter matched to the waveform $\sum_{i=0}^{n-1} b_i s(t - iT - \hat{\tau})$. In cases of high signal to noise ratio (or n large) this maximization can be viewed as that of a "close" estimate of the aperiodic autocorrelation function of the waveform $s(t)$. Classical approaches for obtaining $\hat{\tau}$ include serial search techniques (see [2] and the references therein) and gradient search methods. In general, the aperiodic autocorrelation function of $s(t)$ is not guaranteed to have a unique maximal point even when $\{\gamma_i\}$ is a maximal length sequence, and thus gradient search algorithms can result in a nonconsistent estimate for τ^*. Moreover, typical autocorrelations of code sequences are sharply peaked and have low sidelobes. Therefore, if the initial guess $\hat{\tau}$ is far from the exact delay, the output of the matched filter provides only little information (if any) on the direction and distance of τ^*, and gradient search algorithms can wander in the flat zone of the autocorrelation function for a long time before an initial lock is achieved.

More realistic is the situation where the sequence $\{b_i\}_{i\geq 0}$ is not known to the receiver, and the estimation of τ^* is performed due to the fact that knowledge of the delay is a prerequisite for reliable detection of the bits. Decision-directed procedures for estimating the delay in this setup are described in [3] and in the references there. These algorithms also suffer from the drawbacks described above.

*This work was supported in part by a Wolfson Postdoctoral Fellowship, and in part by the National Science Foundation under Grant NCR-90-02767.

In this work we construct and analyze sequential detection-estimation algorithms that alleviate these problems. We suggest a recursive scheme for estimating τ^*, which is based on maximization of a smoothed version of the periodic autocorrelation function instead of the aperiodic correlation itself, and has a good initial lock property at the expense of higher asymptotic mean square error. The construction of this procedure is based on the following observation. Fix some $0 < \Delta < T/2$ and define the symmetric smoothing kernel $\phi_\Delta(t)$ and the smoothed waveform $s_\Delta(t)$ as

$$\phi_\Delta(t) = \begin{cases} 0 & t \notin [-\Delta, \Delta] \\ (\Delta + t)/\Delta^2 & t \in [-\Delta, 0] \\ (\Delta - t)/\Delta^2 & t \in [0, \Delta] \end{cases}, \quad s_\Delta = s * \phi_\Delta,$$

where $f * g$ stands for the convolution of $f(\cdot)$ and $g(\cdot)$. For every integer $l \geq 0$, let $y_l(\tau, \Delta)$ stand for the output of a filter matched to $s_\Delta(t - lT - \tau)$ and driven by the observation process. Define

$$\rho_\Delta(l, \tau) \triangleq \int_{-\infty}^{\infty} s_\Delta(t - lT - \tau)s(t)dt, \quad c_\Delta \triangleq c * \phi_\Delta.$$

It is easy to verify that $c_\Delta(\tau) = \sum_{i=-1}^{1} \rho_\Delta(i, \tau)$ for $\Delta \geq 0$ and $\tau \in [-T/2, T/2]$. Since the data bits are independent and equiprobable, the following identity holds

$$E_{\tau^*}\left[y_l^2(\tau,\Delta) + 2y_l(\tau,\Delta)y_{l+1}(\tau,\Delta) + 2y_l(\tau,\Delta)y_{l+2}(\tau,\Delta)\right]$$
$$= c_\Delta^2(\tau - \tau^*) + \sigma^2\rho_\Delta(0,0) + 2\sigma^2\rho_\Delta(1,0). \qquad (2)$$

The right hand side of (2) depends on τ only via the periodic correlation function of the signal (or its smoothed version). This suggests that an algorithm with good initial lock properties can be constructed by choosing an appropriate Δ and performing recursive stochastic maximization (with respect to τ) on the left hand side of (2). Moreover, this scheme would not utilize any decisions on the data bits.

In most applications, recursive maximization of the log-likelihood function results in an algorithm which, under the assumption that it converges to the consistent root, can be made asymptotically efficient. Based on this, we construct a second algorithm by coupling a recursive ML scheme to the smoothed correlation scheme, which serves as a "guide" to the correct root of the likelihood function. Using known results, it can be shown that the resulting algorithm is consistent and asymptotically efficient; that is, the delay estimate converges w.p.1 to τ^* (as $n \to \infty$), and the asymptotic mean square error is the optimal one. This technique has been demonstrated in the related problem of multiuser amplitude estimation in [4].

References

[1] R. J. McEliece, *Finite Fields for Computer Scientists and Engineers*. (Kluwer: Norwell, MA, 1986.)

[2] V. M. Jovanovic, "Analysis of Strategies for Serial Search Spread-Spectrum Code Acquisition - Direct Approach," *IEEE Trans. Comm.*, Vol. 36, pp. 1208-1220, November 1988.

[3] D. D. Falconer and J. Salz, "Optimal Reception of Digital Data Over the Gaussian Channel with Unknown Delay and Phase Jitter," *IEEE Trans. Inform. Theory*, Vol. 23, pp. 117-126, January 1977.

[4] Y. Steinberg and H. V. Poor, "Sequential Amplitude Estimation in Multiuser Communications," submitted to *IEEE Trans. Inform. Theory*, 1992.

Performance Study of Maximum-Likelihood Receivers and Transversal Filters for the Detection of Direct-Sequence Spread-Spectrum Signal in Narrowband Interference

Arif Ansari and R. Viswanathan
Department of Electrical Engineering
Southern Illinois University
Carbondale, IL 62901

Abstract - Linear least squares estimation techniques can be used to enhance suppression of narrowband interference in direct-sequence spread-spectrum systems. Nonlinear techniques for this purpose have also been investigated recently. Here, we derive maximum-likelihood receivers for direct-sequence signal in Gaussian interference with known second order characteristics. It is shown that if the receiver uses samples from outside the bit interval, then the receiver structure (called *ML II*)is nonlinear. The bit error rate performances of these ML receivers are compared to those of linear receivers employing one-sided and two-sided least squares estimation filters, for the case of Gaussian autoregressive interference. It is shown that the ML II receiver outperforms the matched filter, the one sided and the two sided transversal filters.

I. INTRODUCTION

Direct-sequence spread-spectrum systems offer an inherent capability of rejecting narrowband interference. This is achieved by modulating the bit waveform with a PN signal before transmission and correlating the received signal with a replica of the PN signal. In this way, interfering signals, whose bandwidths are narrow compared to the spread signal, are attenuated by the receiver. Processing the received signal prior to correlating with the PN sequence has been employed to improve the suppression of narrowband interference. Linear least squares estimation techniques to estimate and subtract the narrowband interference have been studied [1]. Nonlinear techniques for interference suppression in spread-spectrum systems have been investigated in [2]. Here, we study the performance of maximum-likelihood receivers for direct sequence spread spectrum signals received in Gaussian interference with known second order statistics. When the receiver operates on the observations in the bit duration only, the receiver is the well known linear detector known as the matched filter. When the observation interval extends outside the bit interval, the receiver structure is shown to be nonlinear. The nonlinearity arises not due to the modeling of the binary chip sequence as as random as in [2], but due to the uncertainty on the bits adjacent to the bit being tested.

II. MAXIMUM-LIKELIHOOD RECEIVERS

We consider here the performance of maximum-likelihood receivers for the following problem . We shall restrict to the case where an entire maximal length PN code sequence is embedded in each bit (so called short PN sequences). A similar analysis can be easily done for the case of long PN sequences. Let the received signal be processed by a chip-matched-filter and sampled at the chip rate of the PN sequence to yield [2]:

$$z_k = s_k + n_k + j_k \qquad (1)$$

where $s_k = S\, b_k\, c_k$. Without loss of generality,the signal strength S is assumed to be 1.0. c_k is the kth chip of the PN sequence with chip interval τ_c. c_k for $k < 0$ or $k > L-1$ is taken modulo L. $b_k \in \{+1,-1\}$ is the binary information with bit duration $T_b = L\tau_c$, L is the processing gain given as the number of PN chips per message bit. Note that $b_k = b \in \{\pm 1\}$ for all k in the same bit interval. n_k is a sequence of zero mean i.i.d. Gaussian noise with known variance σ_n^2. j_k is a sequence of narrowband interference modeled as a zero mean Gaussian process with autocovariance $R_j(k)$. The detection problem is:

all b_k over the current bit (i.e. b) = $\begin{cases} -1: & H_0 \\ +1: & H_1 \end{cases}$ (2)

Let $v_k = n_k + j_k$ be the white noise plus the interference with autocovariance $R_v(m)=\sigma_n^2\,\delta(m) + R_j(m)$. Let Λ be the $L\times L$ covariance matrix of $\{v_k\}$. The maximum-likelihood detector for the detection problem in (2) is given by :

$$\underline{z}^T \Lambda^{-1} \underline{c} \;\underset{-1}{\overset{+1}{\gtrless}}\; 0 \qquad (3)$$

where $\underline{z}^T = [z_0, z_1, \ldots, z_{L-1}]$, $\underline{c}^T = [c_0, c_1, \ldots, c_{L-1}]$. Call this the *ML I* receiver.

2.1 ML II Receiver and its Bit Error Rate.

Now consider the observation vector to consist of the chips corresponding to the bit under test appended with some chips from the previous bit, i.e. the receiver has to test the present bit but uses observation samples from the present bit interval and a part of the previous bit interval. Let $\underline{r}^T = [\underline{\tilde{z}}^T \; \underline{z}^T]$ where $\underline{\tilde{z}}^T = [\tilde{z}_{L-i}, \tilde{z}_{L-i+1}, \ldots, \tilde{z}_{L-1}]$ is the vector of the last i chip samples from the previous bit, $i \le L$. The likelihood ratio, $\lambda(\underline{r})$, and the corresponding maximum-likelihood detector for the detection problem in (2) is then given by:

$$\lambda(\underline{r}) = \frac{\sum_{d\in\{\pm 1\}} \exp\{\underline{s}_{d,+1}^T \Lambda^{-1} (\underline{r} - \frac{1}{2}\,\underline{s}_{d,+1})\}}{\sum_{d\in\{\pm 1\}} \exp\{\underline{s}_{d,-1}^T \Lambda^{-1} (\underline{r} - \frac{1}{2}\,\underline{s}_{d,-1})\}} \;\underset{-1}{\overset{+1}{\gtrless}}\; 1 \qquad (4)$$

where Λ is the $(L+i)\times(L+i)$ covariance matrix of the sequence $\{v_k\}$, and $\underline{s}_{d,b}^T = [dc_{L-i}, dc_{L-i+1}, \ldots, dc_{L-1}, bc_0, bc_1, \ldots, bc_{L-1}]$, the subscript d indicates the previous bit, $d \in \{\pm 1\}$. Using straightforward calculations involving partitioned vectors and matrices, it can be shown that the bit error probability for the detector in (4) is given by:

$$P_e = Pr\{\sinh(\theta_1) > \gamma \sinh(\theta_2) \mid H_0\} \qquad (5)$$

where $\theta_1 = \underline{s}_{+1,+1}^T \Lambda^{-1} \underline{r}$, $\theta_2 = \underline{s}_{-1,+1}^T \Lambda^{-1} \underline{r}$, γ is a negative constant obtained from the entries in Λ^{-1} matrix and $\underline{s}_{+1,+1}$ vector The test statistic given by (4) is nonlinear in observations. The receiver based on (4) will be called ML II.

III. PERFORMANCE COMPARISON

The bit error rate performances of the ML I and ML II receivers are evaluated numerically and compared to the performances of the one-sided and two-sided transversal filters. The narrowband interference is modeled as a second order zero mean Gaussian autoregressive process with known parameters. As expected, both the maximum-likelihood receivers and the transversal filters perform better when the power spectral density is peaky. The nonlinear ML II receiver outperforms the matched filter receiver and the one-sided and two-sided transversal filters.

REFERENCES

[1] L. B. Milstein, "Interference rejection techniques in spread spectrum communications,"Proc. IEEE, vol. 76, No.6, June 1988, pp 657-671.

[2] R. Vijayan and H. Vincent Poor, "Nonlinear techniques for interference suppression in spread-spectrum systems," IEEE Trans. on Commun., Vol. 38, No. 7, July 1990, pp 1060-1065.

Optimal Detection of Discrete Markov Sources Over Discrete Memoryless Channels — Applications to Combined Source-Channel Coding †

Nam Phamdo and Nariman Farvardin
Electrical Engineering Department
and Systems Research Center
University of Maryland
College Park, Maryland 20742

Summary

In his celebrated paper [1], Shannon stated that in information transmission over a noisy channel, "redundancy must be introduced in the proper way to combat the particular noise structure involved. However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has a certain redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise."

This statement, though made more than forty years ago, forms the foundation of the present work. The principal assumption here is that the source to be transmitted has a certain redundancy and due to certain constraints (for example, on the complexity), the transmitter makes no attempt to "match" the source to the channel. Instead, the source is transmitted directly over the channel. The problem thus is to design a receiver which fully "utilizes" the source redundancy to combat the effect of channel noise.

It is hypothesized that the source is in the form of a discrete Markov chain and that the channel is a discrete memoryless channel. The receiver is a *maximum a posteriori* (MAP) receiver (detector). The redundancy between successive symbols of the Markov source is used by the MAP detector to provide some protection against channel errors.

The above formulation has been considered before by several authors. The most notable is the work by Drake [2], who provided the optimal instantaneous MAP decoding rule as well as bounds on the achievable probability of error. Drake also studied the special case of binary symmetric Markov source and binary symmetric channel and gave a necessary and sufficient condition for the optimality of the singlet ("believe-what-you-see") decoding rule. More recently, Sayood and Borkenhagen [3] considered the detection of a discrete Markov source over a discrete memoryless channel in a joint source-channel DPCM image coding system.

In this work, we consider two variations of this problem: (i) *sequence* MAP detection which is to determine the most probable transmitted *sequence* given an observed sequence and (ii) *instantaneous* MAP detection which is to determine the most probable transmitted *symbol* at a particular time given all the observations up to that time. The solution to the first problem results in a "Viterbi-like" implementation of the MAP detector (with large delay) while the latter problem results in a recursive implementation (with no delay). For the special case of binary symmetric Markov source and binary symmetric channel, we give a necessary and sufficient condition (similar to Drake) for the optimality of the "believe-what-you-see" sequence MAP decoding rule (see [4]). Extensive simulation results for this special case are given in [4].

The solutions to the above problems are applied to a combined source-channel coding problem. The source is assumed to be highly correlated and the source encoder is a small-block-size vector quantizer (VQ). Since the VQ input is correlated from block to block, its output is also correlated. This correlation is referred to as the "residual" redundancy [3]. For simplicity, we model the VQ output as a discrete Markov source. The MAP detectors described above are then used for error detection and correction of the VQ indexes. Simulation results for this system on a Gauss-Markov source are obtained and comparisons are made with Farvardin and Vaishampayan's channel-optimized VQ (COVQ) [5, 6] and the ordinary VQ designed for a noiseless channel. Table 1 shows a summary of our simulation results. More extensive results can be found in [4].

ϵ	VQ	VQ+ Inst. MAP	VQ+ Seq. MAP	COVQ [5, 6]	VQ	VQ+ Inst. MAP	VQ+ Seq. MAP	COVQ [5, 6]
	$k=1; R=1.0$				$k=1; R=3.0$			
0.005	4.24	4.24	4.24	4.25	11.56	12.91	13.69	12.04
0.010	4.09	4.09	4.09	4.11	9.83	11.86	13.03	10.50
0.050	3.09	3.08	3.82	3.15	4.26	7.90	10.03	6.47
0.100	2.09	2.09	3.29	2.27	1.56	5.31	7.34	4.67
	$k=2; R=1.0$				$k=4; R=1.0$			
0.005	7.37	7.61	7.70	7.31	9.08	9.40	9.64	9.15
0.010	6.88	7.30	7.49	6.83	8.21	8.76	9.18	8.37
0.050	4.21	5.65	6.31	4.37	4.31	5.51	6.62	6.23
0.100	2.27	4.01	5.04	2.76	1.95	3.27	4.49	4.65

Table 1: SNR (in dB) Performances of Combined Source-Channel Coding Schemes Using MAP Detection for a Gauss-Markov Source with $\rho = 0.9$; k = Dimension; R = Rate (Bits/Sample); ϵ = Channel Bit Error Rate.

References

1. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, Vol. 27, pp. 379-423 and 623-656, 1948.
2. A. W. Drake, "Observation of a Markov Source Through a Noisy Channel," *Sc. D. Thesis*, M.I.T., Jun. 1962.
3. K. Sayood and J. C. Borkenhagen, "Use of Residual Redundancy in the Design of Joint Source/Channel Coders," *IEEE Trans. Commun.*, Vol. 39, pp. 838-846, Jun. 1991.
4. N. Phamdo and N. Farvardin, "Optimal Detection of Discrete Markov Sources Over Discrete Memoryless Channels — Applications to Combined Source-Channel Coding," submitted to *IEEE Trans. Inform. Theory*, Mar. 1992.
5. N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding," *IEEE Trans. Inform. Theory*, Vol. 33, pp. 827-838, Nov. 1987.
6. N. Farvardin and V. Vaishampayan, "On the Performance and Complexity of Channel-Optimized Vector Quantizers," *IEEE Trans. Inform. Theory*, vol. 37, pp. 155-160, Jan. 1991.

†This work was supported in part by National Science Foundation grants NSFD MIP-86-57311 and NSFD CDR-85-00108. and in part by NTT Corporation and General Electric Co.

A Communication Channel Modeled by the Spread of Disease

Fady Alajaji and Tom Fuja
Electrical Engineering Department, Systems Research Center
University of Maryland, College Park, Maryland 20742

1. Overview

We consider a discrete channel with memory in which errors spread like the spread of a contagious disease through a population. Our motivation is the observation by Stapper *et. al.* that the Polya-Eggenberger (PE) distribution is a better "fit" to the distribution of defects in silicon than the commonly used Poisson distribution. The PE distribution is one of the distributions generated by Polya's urn model for the spread of contagion. We introduce a communication channel with noise modeled by Polya's process. We first present a maximum likelihood (ML) decoding algorithm; we then show that this channel is in fact an "averaged" channel in the sense of Ahlswede and others, and its capacity is zero. Finally, we consider a finite-memory version of the Polya-contagion model; this channel is (unlike the original) ergodic with a non-zero capacity.

2. Polya-Contagion Channel

Consider a discrete binary additive communication channel: $\mathbf{Y}_i = \mathbf{X}_i \oplus \mathbf{Z}_i$, where the random variables $\mathbf{X}_i$, $\mathbf{Z}_i$, and $\mathbf{Y}_i$ are, respectively, the i'th input, noise, and output of the channel. We assume that the input and noise sequences are independent. The noise sequence $\{\mathbf{Z}_i\}$ is generated according to Polya's contagion urn scheme, described as follows. An urn originally contains T balls, of which R are red and S are black. Let $\rho = R/T$ and $\sigma = 1 - \rho = S/T$. We make successive draws from the urn; after each draw, we return to the urn $1 + \Delta$ balls of the same color as was just drawn. In our problem we assume that $\Delta > 0$ (contagion case) and that $\rho < \sigma$. The noise sequence $\{\mathbf{Z}_i\}$ is generated by the draws: $\mathbf{Z}_i = 1$ if the i'th draw yields a red ball and $\mathbf{Z}_i = 0$ if the i'th draw yields a black ball.

For an input block $\underline{\mathbf{X}} = [\mathbf{X}_1, \ldots, \mathbf{X}_n]$ and an output block $\underline{\mathbf{Y}} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_n]$, the block transition probability of the channel is:

$$P(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x}) = \frac{\Gamma(\frac{1}{\delta})\,\Gamma(\frac{\rho}{\delta} + d)\,\Gamma(\frac{\sigma}{\delta} + n - d)}{\Gamma(\frac{\rho}{\delta})\,\Gamma(\frac{\sigma}{\delta})\,\Gamma(\frac{1}{\delta} + n)} \qquad (1)$$

where $d = d_H(\underline{y}, \underline{x})$, the Hamming distance between $\underline{y}$ and $\underline{x}$.

Channel Properties: Two important properties: (1) *Stationarity:* From equation (1) the noise $\{\mathbf{Z}_i\}$ forms an infinite sequence of exchangeable random variables. Therefore, the noise process is strictly stationary. (2) *Non-Ergodicity:* Let $\mathbf{S}_n \triangleq \mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_n$. It can be shown that $\mathbf{Z} = \lim_{n\to\infty} \mathbf{S}_n/n$ is (with probability one) a random variable drawn according to the beta distribution with parameters ρ/δ and σ/δ. Thus the noise process $\{\mathbf{Z}_i\}$ is not ergodic since its sample average does not converge to a constant.

Maximum Likelihood (ML) Decoding: Suppose M codewords are possible channel inputs: $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_M$, each of length n. Given an output $\underline{y}$, ML decoding selects as its estimate of the transmitted codeword the $\underline{x}_k$ that maximizes $P(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x}_k)$.

Now $g(d) \triangleq P(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x})$ is strictly log-convex in $d \in [0, n]$ with a unique minimum at $d_0 = n/2 + (1 - 2\rho)/2\delta$. Thus the ML decoding algorithm for the channel is given as follows:

1. Given the received vector $\underline{y}$, compute $d_i \triangleq d_H(\underline{y}, \underline{x}_i)$ for each i. Compute also $d_{\max} \triangleq \max\{d_i\}$ and $d_{min} \triangleq \min\{d_i\}$.
2. If $|d_{\max} - d_0| \le |d_{\min} - d_0|$, map $\underline{y}$ to the $\underline{x}_j$ for which $d_j = d_{\min}$. In this case ML decoding $\Longleftrightarrow$ minimum distance decoding.
3. If $|d_{\max} - d_0| > |d_{\min} - d_0|$, map $\underline{y}$ to the $\underline{x}_j$ for which $d_j = d_{\max}$. In this case ML decoding $\Longleftrightarrow$ maximum distance decoding.

Averaged Communication Channels: Consider a family of discrete memoryless channels parameterized by θ:

$$\{W_\theta^{(n)}(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x}) = \prod_{i=1}^{n} W_\theta^{(1)}(\mathbf{Y}_i = y_i \mid \mathbf{X}_i = x_i) : \theta \in \Theta\}_{n=1}^{\infty}.$$

A channel is "averaged" if its block transition probability is the expected value of the block transition probability taken with respect to some distribution on θ – i.e., if it's of the form

$$W_A^{(n)}(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x}) = \int_\Theta W_\theta^{(n)}(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x})\, dG(\theta) \qquad (2)$$

where $(\Theta, \sigma(\Theta), G)$ is a probability space for the random variable θ. Note that the averaged channel has memory and is stationary.

Claim: The binary Polya-contagion channel is an averaged channel; specifically, the Polya-contagion channel represents the class of binary symmetric channels with crossover probability θ, where θ is distributed according to the beta distribution with parameters ρ/δ and σ/δ. Furthermore, from the results of Ahlswede we can show that the capacity of this channel is zero.

3. Finite-Memory Contagion Channel

An unrealistic aspect of the Poly-contagion channel is its infinite memory. Consider, for instance, the millionth ball drawn from Polya's urn; the very *first* ball drawn from the urn and the 999,999'th ball drawn from the urn have the identical effect on the outcome of the millionth draw. We now consider a perhaps more realistic model for a contagion channel with finite memory.

As before, consider an urn with T balls, of which R are red and $S = T - R$ are black. At the j'th draw we select a ball from the urn and replace it with $1 + \Delta$ balls of the same color; then, M draws later – after the $(j + M)$'th draw – we retrieve from the urn Δ balls of the color picked at time j. As before, let $\mathbf{Z}_i = 1$ if the i'th draw yields a red ball and $\mathbf{Z}_i = 0$ if the i'th draw yields a black ball. This modification keeps the total number of balls in the urn constant ($T + M\Delta$ balls) after an initialization period of M draws; it also limits the effect of any draw to M draws in the future.

For blocklength $n \le M + 1$, the block transition probability of this new channel is given by (1). For $n \ge M + 2$, we obtain:

$$P(\underline{\mathbf{Y}} = \underline{y} \mid \underline{\mathbf{X}} = \underline{x}) = L \prod_{i=M+2}^{n} \left[\frac{\rho + s_{i-1}\delta}{1 + M\delta}\right]^{e_i} \left[\frac{\sigma + (M - s_{i-1})\delta}{1 + M\delta}\right]^{1-e_i}$$

where $L = [\prod_{i=0}^{k-1}(\rho + i\delta) \prod_{j=0}^{M-k}(\sigma + j\delta)] / \prod_{\ell=1}^{M}(1 + \ell\delta)$. Here, $e_i = x_i \oplus y_i$, $k = e_1 + \cdots + e_{M+1}$, and $s_{i-1} = e_{i-1} + \cdots + e_{i-M}$.

Claim: The new noise process $\{\mathbf{Z}_i\}$ is a stationary ergodic Markov process of order M, and thus the resulting channel is a Markov channel with memory M. The capacity C_M of the channel is given by:

$$C_M = 1 - \sum_{i=0}^{M} \binom{M}{i} L_i\, h_b\left(\frac{\rho + i\delta}{1 + M\delta}\right)$$

where $L_i = [\prod_{j=0}^{i-1}(\rho + j\delta) \prod_{\ell=0}^{M-i-1}(\sigma + \ell\delta)] / \prod_{m=1}^{M-1}(1 + m\delta)$, and $h_b(x)$ is the binary entropy function.

Finally if we let $M \to \infty$, $C_M \to 1 - \int_0^1 h_b(x)\, f_{\mathbf{Z}}(x) dx$ where $f_{\mathbf{Z}}(x)$ is the beta pdf with parameters ρ/δ and σ/δ. This result is identical to $\lim_{n\to\infty}(1/n) I(\mathbf{X}^n; \mathbf{Y}^n)$ if $\mathbf{X}^n$ and $\mathbf{Y}^n$ are blocks of length n joined by the original Polya-contagion channel (with equally likely inputs). Thus as $M \to \infty$, the stationary ergodic finite-memory contagion channel converges in distribution to the stationary non-ergodic Polya channel, but C_M does not converge to $C_{Polya} = 0$.

Demodulation of AM-FM Signals in Noise Using Multiband Energy Operators

Alan C. Bovik, *Laboratory for Vision Systems, University of Texas, Austin, TX 78712-1084*
Petros Maragos, *Division of Applied Sciences, Harvard University, Cambridge, MA 02138*
Thomas F. Quatieri, *MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02173*

I. INTRODUCTION

We study the extraction of AM-FM information in signals of the form

$$s(t) = a(t)\cos[\phi(t)], \tag{1}$$

with time-varying amplitude a and instantaneous frequency $\omega_i = \dot{\phi}$, using the operator $\Psi(s) = (\dot{s})^2 - s\ddot{s}$ developed by Teager [1] and Kaiser [2], shown to be highly effective for detecting AM-FM modulations [3]. For signals of the form (1), $\Psi(s) \approx a^2\omega_i^2$ and $\Psi(\dot{s}) \approx a^2\omega_i^4$, with small approximation error under realistic conditions [3]. This motivates the *energy separation algorithm* (ESA):

$$\hat{a}^2 = \Psi^2(s) / \Psi(\dot{s}), \qquad \hat{\omega}_i^2 = \Psi(\dot{s}) / \Psi(s).$$

II. ENERGY OF FILTERED NOISY AM-FM SIGNAL

Define a noisy AM-FM signal $f = s + n$, with s given by (1) and n a zero-mean WSS Gaussian random process with autocorrelation $R(\tau)$ and power spectral density $\Phi(\omega)$. Consider bandpass filters with scaled, translated frequency responses

$$G_\sigma(\omega) = (1/j)\,[H_\sigma(\omega-\omega_c) - H_\sigma(\omega+\omega_c)] \tag{2}$$

where $H_\sigma(\omega) = \sigma^{-1/2}H(\omega/\sigma)$ is low-pass and unit energy, and denote $f_\sigma = s_\sigma + n_\sigma = s*g_\sigma + n*g_\sigma$. An important approximation is often used here:

$$s_\sigma \approx \hat{s}_\sigma = a|G_\sigma(\omega_i)| \cdot \cos[\phi + \angle G_\sigma(\omega_i)]. \tag{3}$$

The error (zero for a monochromatic signal) is bounded as follows. First define

$$\Delta_p(g_\sigma) = \left[\int_R t^{2p}\,|g_\sigma(t)|^2\,dt\right]^{1/2}, \qquad \delta(a) = \left[\int_R |\dot{a}(t)|^2\,dt\right]^{1/2}.$$

Theorem 1 - Let $\varepsilon_s = |s_\sigma - \hat{s}_\sigma|$ and $a_{max} = \sup_t|a(t)|$. Then

$$\varepsilon_s \le \frac{4}{3}\,a_{max}\,\Delta_2(g_\sigma)\cdot\delta(\omega_i) + 2\,\Delta_1(g_\sigma)\cdot\delta(a). \quad \blacklozenge$$

We can also bound estimates of the energy Ψ:

$$\widehat{\Psi}(s_\sigma) = (a\omega_i)^2|G_\sigma(\omega_i)|^2, \qquad \widehat{\Psi}(\dot{s}_\sigma) = (a\omega_i^2)^2|G_\sigma(\omega_i)|^2. \tag{4}$$

Theorem 2 - Let $\varepsilon_\Psi = |\Psi(s_\sigma) - \widehat{\Psi}(s_\sigma)|$, $\bar{g}_\sigma = \int_R |g_\sigma(t)|\,dt$. Then

$$\varepsilon_\Psi \le \tfrac{4}{3}(a_{max})^2\cdot\delta(\omega_i)\cdot[\bar{g}_\sigma\Delta_2(\dot{g}_\sigma) + \bar{\dot{g}}_\sigma\Delta_2(g_\sigma) + 2\bar{\dot{g}}_\sigma\Delta_2(\dot{g}_\sigma)]$$
$$+ 2a_{max}\cdot\delta(a)\cdot[\bar{g}_\sigma\Delta_1(\dot{g}_\sigma) + \bar{\dot{g}}_\sigma\Delta_1(g_\sigma) + 2\,\bar{\dot{g}}_\sigma\Delta_1(\dot{g}_\sigma)]. \quad \blacklozenge$$

Approximations (3), (4) suggest minimum uncertainty filters minimize model errors: $H_\sigma(\omega) = \sqrt{2/\sigma\sqrt{2\pi}}\,\exp\{-(\omega/\sigma)^2\}$; then (2) are Gabor functions.

III. COMPUTING THE ESA IN NOISE

Define the *instantaneous signal-to-noise ratio*: $S_\sigma(t) = a^2(t) / \Gamma_\sigma$, where Γ_σ is the concentration of noise power in the passband of $g_\sigma(t)$:

$$\Gamma_\sigma = \frac{1}{2\pi}\int_R |G_\sigma(\omega)/G_\sigma(\omega_c)|^2\,\Phi(\omega)\,d\omega.$$

For S_σ sufficiently large it can be shown using (3), (4) that:

$$E[\hat{\omega}_i^2] \approx \omega_i^2\left\{1 + \frac{4\,S_\sigma}{(S_\sigma+2)^2}\right\} \approx \omega_i^2 \tag{5}$$

$$E[\hat{a}^2] \approx a^2\left\{1 + \frac{10\,S_\sigma+4}{S_\sigma\,(S_\sigma+2)}\right\}\cdot|G_\sigma(\omega_i)|^2 \approx a^2|G_\sigma(\omega_i)|^2. \tag{6}$$

$$\mathrm{Var}[\hat{\omega}_i^2] \approx \omega_i^4\,4(S_\sigma+1)/(S_\sigma+2)^2 \tag{7}$$

$$\mathrm{Var}[\hat{a}^2] \approx 4a^4\,[(5S_\sigma+1)/S_\sigma^2]\,|G_\sigma(\omega_i)|^4. \tag{8}$$

The ratios of (7), (8) to the squares of (5), (6) are negligible for reasonably high S_σ, in which case it may be asserted that $\hat{\omega}_i^2 \approx \omega_i^2$ and $\hat{a}^2 \approx a^2|G_\sigma(\omega_i)|^2$.

IV. MULTIBAND FILTERING AND ESA

Figure 1 diagrams a multiband energy operator - $f(t)$ is passed through multiple passbands $g_m \leftrightarrow G_m$ with center frequencies ω_m, producing outputs $f_m(t)$.

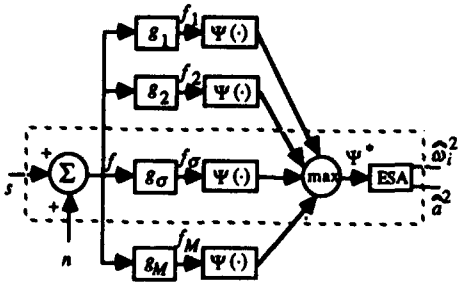


Fig. 1. Multiband filtering / ESA applied to noisy AM-FM signal.

At time t the maximum normalized energy response (dotted lines in Fig. 1)

$$\Psi^*(t) = \max_m\left\{\Psi[f_m(t)] \,/\, |G_m(\omega_m)|^2\right\}$$

is used by the ESA. Once $H(\omega)$ is selected, tesselate the frequency axis with translates/dilates of G_σ. Since the validity of (5)-(8) depend on functions of σ/ω_c - *in order to maintain consistent predicted performance across the filter channels* the error bound is made constant by taking σ_m/ω_m = constant.

V. EXAMPLE

The multiband ESA was applied to the noisy chirp (SNR = 15dB) with initial frequency 2480Hz, a 3000Hz/sec sweep rate, and a 20Hz amplitude modulation, shown in Fig. 2(*a*). The ESA results with multiband filtering are shown in Figs. 2(*b*) and 2(*c*), indicating excellent estimates of both AM and FM components; these could be improved even further by post-filtering.

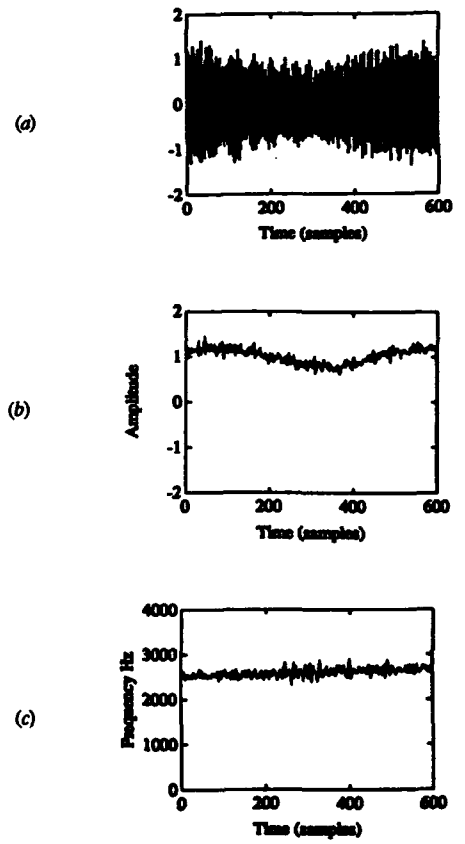


Fig. 2. (*a*) noisy AM chirp signal. (*b*), (*c*) Computed AM and FM.

ACKNOWLEDGEMENTS

ACB was supported by a University of Texas Faculty Research Assignment on sabbatical at Harvard University and a grant from Texas Instruments. PM was supported by NSF Grant MIP-9120624 and an NSF PYI Award. TFQ was supported in part by the Department of the Air Force.

REFERENCES

[1] H.M. Teager & S.M. Teager, "Evidence for nonlinear speech production mechanisms in the vocal tract," *NATO Adv. Study Inst. Speech Prod. Speech Model.*, Bonas, France, July, 1989.

[2] J.F. Kaiser, "On Teager's energy algorithm and its generalization to continuous signals," *Proc. DSP Wkshp.*, New Paltz, NY, Sept. 1990.

[3] P. Maragos, T.F. Quatieri & J.F. Kaiser, "Speech nonlinearities, modulations, and energy operators," *Proc. ICASSP*, Toronto, 1991.

Information Theory and Radar Waveform Design

Mark R. Bell
School of Electrical Engineering, Purdue University
West Lafayette, IN 47907

Abstract

The use of information theory to design waveforms for the measurement of extended radar targets exhibiting resonance phenomena is investigated. The target impulse response is introduced to model target scattering behavior. Two radar waveform design problems with constraints on waveform energy and duration are then solved. In the first, a deterministic target impulse response is used to design waveform/receiver-filter pairs for the optimal detection of extended targets in additive noise. In the second, a random target impulse response is used to design waveforms that maximize the mutual information between a target ensemble and the received signal in additive Gaussian noise. The two solutions are contrasted to show the difference between the characteristics of waveforms for extended target detection and information extraction. The optimal target detection solution places as much energy as possible in the largest target scattering mode under the imposed constraints on waveform duration and energy. The optimal information extraction solution distributes the energy among the target scattering modes in order to maximize the mutual information between the target ensemble and the received radar waveform.

Summary

The application of information theory to radar was originally considered by Woodward and Davies [1, 2], who used information theoretic ideas to formulate the a posteriori radar receiver. They also made the observation that, although radar system design that maximizes the signal-to-noise ratio at the receiver output achieves the best target detection performance, it does not necessarily provide the greatest "information gain" about the target. By "information gain," they were referring to the mutual information between a random target parameter to be determined and the measured radar observation of the target. They did not, however, pursue this idea further and investigate the design of radar waveforms and receiver filters that maximize the mutual information between the observed target and the radar measurement of the target. In this talk, we investigate the problem of optimal waveform and receiver filter design for both the detection and information extraction in the case of extended radar targets. Detailed treatments of these problems can be found in [3].

First we consider the design of the optimal waveform/receiver-filter pair for optimal detection of an extended target with a given target impulse response under constraints on waveform energy and time duration in the presence of wide-sense stationary additive noise. The receiver filter is seen to be a straightforward generalization of the matched filter. However, the overall signal-to-noise ratio is dependent on the transmitted waveform. The transmitted waveform that maximizes the signal-to noise ratio is that which places as much of the transmitted energy as possible into the largest scattering mode of the target under the imposed duration and energy constraints. This waveform can be found by solving a Fredholm integral equation whose kernel is a function of the target impulse response and the power spectral density of the additive noise.

Next we examine the problem of designing waveforms that maximize the mutual information between a random extended target ensemble and the associated radar measurement in the presence of additive Gaussian noise. Here, the random target ensemble is modeled by a target impulse response that is assumed to be a non-stationary finite-energy Gaussian random process whose spectral-mean and spectral-variance are known. We solve for the family of waveforms that maximize the mutual information between the target ensemble and the measurement under constraints on waveform energy and duration. The resulting family of optimal waveforms can be interpreted as spreading the energy in the transmitted waveform under the among the various target scattering modes in such a way that the mutual information is maximized. The solution has the spectral form of the "water-pouring problem" in continuous waveform design, with parameters given in terms of the target ensemble's spectral-variance and the power spectral density of the additive noise.

We then note the physical interpretation of radar waveform design in terms of distributing energy among the various scattering modes of the target and note the distinct difference between optimal detection waveforms and optimal information extraction waveforms when viewed in this context. This serves to illuminate the distinct differences between optimal waveforms for these two tasks when making measurements of extended radar targets.

References

[1] P. M. Woodward and I. L. Davies, "A Theory of Radar Information," *Phil. Mag.*, vol. 41, pp. 1101–17, Oct. 1951.

[2] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, London, England: Pergamon Press, 1953.

[3] M. R. Bell, "Information Theory and Radar Waveform Design," to appear in *IEEE Transactions on Information Theory*.

An Upper Bound of the Capacity of Hopfield Net with Perceptron Algorithm

Shiyi Shen
Department of Mathematics
Nankai University
Tianjin 300071 P.R.China

Zhongxing Ye
Department of Applied Math
Shanghai Jiao Tong University
Shanghai 200030 P.R.China

Abstract

The storage canacity of Hopfield net is discussed based upon Perceptron algorithm. We first prove a theorem of linear separability for perceptron using the technique of convex analysis. It is then applied to estimate the memory ratio c for Hopfield net with n neurons. We evaluate the probability $P(n,cn)$ that any cn randomly generated patterns are the attractors of the net. When $\lim_{n\to\infty} P(n,cn)=1$, the net is capable to memorize cn patterns in the form of its equilibria. The maximum of such $c's$ denoted by C_a is defined as the capacity of the net. We obtain an upper bound $C_a \le 1/p_0$, where p_0 is the solution of the following equation

$$-p\log p-(1-p)\log(1-p)+p-1=0,\quad 0\le p\le 1$$

Since $p_0 > 0.227$. we obtain $c_a < 4.41$.

Summary

The dynamics of Hopfield net with n neurons can be described by a operator $T: \{\pm\}^n \to \{\pm\}^n$

$$y^n = T(x^n) = sgn(\mathbf{W}\mathbf{X}-\mathbf{h}),$$

where h is the vector of thresholds, $\mathbf{W}=(w_{ij})$ is the connection matrix and sgn is operated componentwise. A state x is an equilibrium state when it satisfies

$$\mathbf{x} = T\mathbf{x}.$$

Given an arbitrary set of m desired memories $\xi(1), \xi(2), \cdots, \xi(m)$, these vectors should indeed be stable vectors. i.e..

$$\xi_i(s) = sgn\Big|\sum_{j=1}^{n} w_{ij}\xi_j(s) - h_i\Big|$$

for any $s = 1.2.\cdots;.m.\ i = 1.2.\cdots;.n$. Based upon Perceptron algorithm. W can be choosen to satisfy the following set of inequalities

$$\sum_{j\neq i} w_{ij}\xi_j(s)\xi_j(s) - h_i\xi_i(s) > 0. \qquad (1)$$

When $\{\xi_i(s),\ s=1,2,\cdots,m;\ i=1,2,\cdots,n\}$ are randomly generated patterns, or more precisely, $\xi_i(s)$ are independent random variables taking 1 and -1 with probability 1/2 each, we define by $p(n,m)$ the probability that the set of equations (1) has solution. We say a rate c is achievable if

$$P(n,cn)\to 1 \quad as \quad n\to\infty.$$

The capacity C_a of Hopfield net under Perceptron algorithm is defined as the supremum of the achievable rates.

There have been some works regarding the memory capacities of different types under different learning algorithms for Hopfield networks. Among them Gardner (1988) discussed another type of capacity for the Hopfield net with Perceptron algorithm and obtained a Upper bound $C_a \le 2$. In this work we first prove a theorem of linear separability for perceptron using the technique of convex analysis. It is then applied to estimate the capacity C_a for Hopfield net. we obtain an upper bound

$$C_a \le 1/p_0,$$

where p_0 is the solution of the following equation

$$-p\log p-(1-p)\log(1-p)+p-1=0,\quad 0\le p\le 1$$

since $p_0 > 0.227$. we get $C_a < 4.41$.

This work is supportted by Chinese Tian Yuan Foundation.

Corrective Memory by a Symmetric Sparsely Encoded Network

Y. Baram *Department of Computer Science, Technion, Israel Institute of Technology, Haifa 32000, Israel.*

A neural network that retrieves stored binary vectors, when probed by possibly corrupted versions of them, is presented. It employs sparse ternary internal coding and autocorrelation (Hebbian) storage. It is symmetrically structured and, consequently, can be folded into a feedback configuration. Bounds on the network parameters are derived from probabilistic considerations. The asymptotic storage capacity is shown to be arbitrarily close to linear in the network size, which is exponential in the input dimension. The performance of a finite-size symmetric network is examined by simulation and found to be substantially higher than that of Kanerva's seminal model, operating as a content addressable memory.

Strong Universal Consistency of Neural Network Classifiers

András Faragó and Gábor Lugosi

Department of Telecommunication and Telematics
Technical University of Budapest
1521 Stoczek u. 2, Budapest, Hungary

Department of Mathematics,
Technical University of Budapest
1521 Stoczek u. 2, Budapest, Hungary

ABSTRACT In statistical pattern recognition a classifier is called *universally consistent* if its error probability converges to the Bayes-risk as the size of the training data grows, for *all possible distributions* of the random variable pair of the observation vector and its class. We prove that if a one layered neural network is trained to minimize the empirical risk on the training data, then it results in a universally consistent classifier if the number of nodes k is chosen such that $k \to \infty$ and $k\log(n)/n \to 0$ as the size of the training data n grows to infinity. We show that if certain smoothness conditions on the distribution are satisfied, then by choosing $k = O(\sqrt{n/\log(n)})$, the exponent in the rate of convergence does not depend on the dimension.

I. INTRODUCTION

The pattern classification problem can be formulated as follows: Let the random variable pair (X, Y) take its values from $\mathcal{R}^d \times \{0,1\}$. $X \in \mathcal{R}^d$ is called the *observation (or feature) vector*, while $Y \in \{0,1\}$ is its *class*. Observing X one wants to guess the value of Y by a *classification rule* $g : \mathcal{R}^d \to \{0,1\}$ such that the *error probability* $\Pr\{g(X) \neq Y\}$ be small. The best possible classification rule is given by

$$g^*(x) = \begin{cases} 0 & \text{if } P_0(x) \geq 1/2 \\ 1 & \text{otherwise} \end{cases}$$

where $P_0(x) = \Pr\{Y = 0 | X = x\}$ is the *a posteriori probability* of class 0. The minimal error probability $L^* = \Pr\{g^*(X) \neq Y\}$ is called the *Bayes-risk*. In practice, the a posteriori probabilities are rarely known, instead, a *training sequence*

$$\xi_n = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)) \tag{1}$$

is available, where $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ are independent, identically distributed (i.i.d.) random variable pairs. Now, one can estimate Y by $g_n(X) = g_n(X, \xi_n)$, a measurable function of the observation and the training sequence. The *error probability* of g_n is denoted by

$$L(g_n) = \Pr\{g_n(X) \neq Y | \xi_n\}.$$

A sequence of rules is *strongly universally consistent* if

$$\lim_{n\to\infty} L(g_n) = L^* \quad \text{with probability one} \tag{2}$$

for any distribution of (X, Y).

A classification rule realized by a *feedforward neural network with one (hidden) layer* can be expressed as

$$g(x, \theta_k) = \begin{cases} 0 & \text{if } f(x, \theta_k) \geq 0 \\ 1 & \text{otherwise,} \end{cases} \tag{3}$$

where

$$f(x, \theta_k) = \sum_{i=1}^{k} c_i \sigma(a_i x + b_i) + c_0. \tag{4}$$

Here k is the number of *nodes (hidden neurons)*, and $\theta_k = (a_1, \ldots, a_k, b_1, \ldots, b_k, c_0, c_1, \ldots, c_k)$ is the *parameter vector* of the neural network ($a_i \in \mathcal{R}^d, c_0, b_i, c_i \in \mathcal{R}, i = 1, \ldots, k$). Here we assume that the sigmoid σ is the step function

$$\sigma(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases}$$

Our goal is to choose the number of nodes k and set the parameters such that the error probability $\Pr\{g(X, \theta_k) \neq Y\}$ be small. Our strategy is to minimize the *empirical error*, in other words, we choose a parameter vector $\theta^*_{k,n}$ for which the corresponding classification rule $g^*_{k,n}(x) = g(x, \theta^*_{k,n})$ commits the minimum number of errors on the training sequence:

$$\hat{L}(g^*_{k,n}) = \min_{\theta_k} \hat{L}(g(\cdot, \theta_k)), \tag{5}$$

where

$$\hat{L}(g(\cdot, \theta_k)) = \frac{1}{n} \sum_{j=1}^{n} I_{\{g(X_j, \theta_k) \neq Y_j\}}$$

is the empirical error probability of the classification rule $g(x, \theta_k)$. (I_A denotes the indicator of an event A).

Our main result is that such parameter selection has very good properties. In particular, we can show the following:

Theorem 1 *If the number of nodes k is chosen to satisfy*

$$k \to \infty \tag{6}$$

and

$$\frac{k\log(n)}{n} \to 0 \tag{7}$$

as $n \to \infty$, then

$$\lim_{n\to\infty} L(g^*_{k,n}) = L^* \quad \textit{with probability one,}$$

regardless of the distribution, that is, the sequence of rules $\{g^*_{k,n}\}$ is strongly universally consistent.

The proof is based on the celebrated result of Cybenko and Hornik, Stinchcombe and White, that functions realized by networks with one hidden layer are dense in the class of continuous functions, and the Vapnik-Chervonenkis inequality.

Applying Barron's results we can estimate the rate of convergence for smooth distributions:

Theorem 2 *Assume that there is a compact set $A \subset \mathcal{R}^d$ such that $\Pr\{X \in A\} = 1$ and the Fourier transform $\tilde{P}_0(\omega)$ of $P_0(x)$ satisfies $\int_{\mathcal{R}^d} |\omega| |\tilde{P}_0(\omega)| d\omega < \infty$. Then*

$$EL(g^*_{k,n}) - L^* = O\left(\sqrt{\frac{kd\log(n)}{n}} + \frac{1}{\sqrt{k}}\right).$$

If, further, the number of neurons is chosen to be $k = O(\sqrt{n/d\log(n)})$, then

$$EL(g^*_{k,n}) - L^* = O\left(\left(\frac{d\log(n)}{n}\right)^{1/4}\right).$$

Sample Size Requirements of Feedforward Neural Network Pattern Classifiers*

Terrence L. Fine and Michael J. Turmon
E&TC 388
Cornell University School of Electrical Engineering
Ithaca, NY 14853

We investigate the tradeoffs among network complexity, training set size, and statistical performance of feedforward neural networks.

Nets, labeled as functions $\eta : R^d \to \{0,1\}$, classify input points $\underline{x} \in R^d$ as either type 0 or type 1. The architecture of all nets under consideration is $\mathcal{N}$, whose size is gauged by its VC dimension v, the size of the largest set of points the architecture can classify in any desired way. Nets $\eta \in \mathcal{N}$ are chosen on the basis of a training set $\mathcal{T} = \{(\underline{x}_i, t_i)\}_{i=1}^n$. These n samples are i.i.d. according to an unknown probability law P. Performance of a network is measured by the error probability

$$\mathcal{E}(\eta) = P(\eta(\underline{x}) \neq t).$$

and a good (perhaps not unique) net in the architecture is

$$\eta^0 = \arg\min_{\eta \in \mathcal{N}} \mathcal{E}(\eta).$$

To select a net using the training set we employ the empirical error frequency

$$\nu_{\mathcal{T}}(\eta) = \frac{1}{n} \sum_{i=1}^{n} |\eta(\underline{x}_i) - t_i|$$

sustained by η on the training set $\mathcal{T}$. A good choice for a classifier is then

$$\eta^* = \arg\min_{\eta \in \mathcal{N}} \nu_{\mathcal{T}}(\eta).$$

By definition $\mathcal{E}(\eta^*) \geq \mathcal{E}(\eta^0)$, and in fact arguments in Vapnik [5] can be adapted to yield the VC upper bound

$$P(\mathcal{E}(\eta^*) - \mathcal{E}(\eta^0) \geq \epsilon) \leq 6 \frac{(2n)^v}{v!} \epsilon^{-n\epsilon^2/8}.$$

This inequality shows that sample sizes of about

$$n_c = \frac{16v}{\epsilon^2} \log(\frac{6}{\epsilon})$$

are sufficient to obtain a small probability of a discrepancy of more than ϵ between $\mathcal{E}(\eta^*)$ and $\mathcal{E}(\eta^0)$. If for purposes of illustration we take $\epsilon = .1, v = 50$, we find that $n_c = 328\,000$, which disagrees by orders of magnitude with the experience of practitioners who train such low-complexity networks (about 50 connections).

One way to close this gap between theoretical guidelines and practical experience is to obtain a tighter upper bound. One source of the discrepancy is the union bound employed in the VC development, a tighter version of which is given by Naiman and Wynn [3]:

$$\sum_{1 \leq i \leq N} P(A_i) - \sum_{1 \leq i < j \leq N} P(A_i \cap A_j) \leq P(\bigcup_{1 \leq i \leq N} A_i) \leq$$

$$\sum_{1 \leq i \leq N} P(A_i) - \sum_{1 < i \leq N} P(A_i \cap A_{i-1}).$$

However, we have shown that these pairwise corrections reduce the upper bound by at most a multiplicative factor of n, which is insignificant compared to other factors entering exponentially, while the lower bound becomes trivial.

The number n_c obtained via VC theory represents a sufficient condition on sample size to obtain reliable classification. To supplement this we have obtained a lower bound or a necessary condition on the training set size needed to obtain reliable classification by examining in detail the error terms for a perceptron under multivariate normal input. Suppose the observed data $\underline{x}$ has equal prior probability of being $N(\mu_0, I_d)$ or $N(\mu_1, I_d)$, and that $n/2$ correctly classified samples are gathered from each prior. When the means are known, the classifier η^0 minimizing error probability is

$$\eta^0(\underline{x}) = 1/2 - \mathrm{sgn}((\underline{x} - (\mu_0 + \mu_1)/2)^T(\mu_0 - \mu_1))/2,$$

and $\mathcal{E}(\eta^0) = \Phi(-\Delta/2)$ where $\Delta^2 = (\mu_0 - \mu_1)^T(\mu_0 - \mu_1)$ and Φ is the distribution of $N(0,1)$. The empirically chosen classifier when the means are unknown is formed by substituting the sample means under each hypothesis, $\bar{x}_0$ and $\bar{x}_1$, into η^0:

$$\eta^*(\underline{x}) = 1/2 - \mathrm{sgn}((\underline{x} - (\bar{x}_0 + \bar{x}_1)/2)^T(\bar{x}_0 - \bar{x}_1))/2.$$

$\mathcal{E}(\eta^*)$ is hard to find (see [1], sec. 6.6), but it can be approximated using arguments in the spirit of Raudys [4]. The condition necessary for reliable classification becomes

$$\mathcal{E}(\eta^*) - \mathcal{E}(\eta^0) \approx \Phi(-(\Delta/2)(1 + 4d/n\Delta^2)^{-1/2}) - \Phi(-\Delta/2) < \epsilon,$$

uniformly over all values of Δ. Analysis reveals that meeting the above condition requires

$$n \geq \frac{v}{33\epsilon^2}.$$

lower than the VC sufficient condition by a factor of order just $\log(1/\epsilon)$. For this special case, it also improves on the necessary condition $n > v/32\epsilon$ obtained by Baum and Haussler [6]. This result confirms that the VC bound is relatively tight, and demonstrates that practitioners are overly optimistic when using small sample sizes.

References

[1] Anderson, T., *An Introduction to Multivariate Statistical Analysis*, second ed., New York: Wiley, 1984.

[2] Baum, E. and D. Haussler, "What size net gives valid generalization?," in D. S. Touretzky, ed., *Advances in Neural Information Processing Systems 1*. 81–90. 1989.

[3] Naiman, D. and Wynn, H., "Inclusion–exclusion–Bonferroni identities...", *Annals of Statistics*, **20**, 43–76.

[4] Raudys, Sh., "On the amount of a priori information in construction of a classification algorithm," *Engineering Cybernetics*, no. 4, 1972. (Russian trans.)

[5] Vapnik, V., *Estimation of Dependences Based on Empirical Data*, New York: Springer, 1982.

* Prepared with partial support of DARPA under grant number AFOSR-90-0016A.

Constructions of Depth-2 Majority Circuits for Comparison and Addition using Linear Block Codes

Noga Alon
Department of Mathematics
Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv, Israel

Jehoshua Bruck
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

We address the problem of computing the COMPARISON and ADDITION functions of two n-bit numbers using circuits of (non-monotone) MAJORITY gates.

Given n Boolean variables $x_1, \ldots, x_n \in \{-1, 1\}$, a non-monotone MAJORITY gate (in the variables x_i) is a Boolean function whose value is the sign of $\sum_{i=1}^{n} \epsilon_i x_i$, where each ϵ_i is either 1 or -1. We construct an explicit sparse polynomial whose sign computes the COMPARISON function of two integers. Similar polynomials are constructed for computing all the bits of the summation of the two integers. This supplies explicit constructions of depth-2 polynomial-size circuits computing these functions, which use only non-monotone MAJORITY gates. These constructions are optimal in terms of the depth and can be used to obtain the best known explicit constructions of MAJORITY circuits for other functions like the product of two n-bit numbers and the maximum of n n-bit numbers (see [3] and [6]). A crucial ingredient in our approach is the construction of a discrete version of a sparse "delta polynomial"—one that has a large absolute value for a single assignment and extremely small absolute values for all other assignments. We construct sparse delta polynomials using generator matrices of certain linear block codes.

In the rest of this summary we sketch the ideas related to the construction for the COMPARISON function. More details and related results appear in [1].

Let $X = (x_n, x_{n-1}, \ldots, x_1)$ and $Y = (y_n, y_{n-1}, \ldots, y_1)$ be two vectors in $\{1, -1\}^n$. Let a and b be the integers that correspond to X and Y, respectively. Since our convention is that a logical 0 is represented by $+1$ and a logical 1 is represented by -1 this means that $a = \sum_{i=1}^{n} \frac{1-x_i}{2} 2^{i-1}$ and $b = \sum_{i=1}^{n} \frac{1-y_i}{2} 2^{i-1}$. The COMPARISON function, $C(X, Y)$, is the Boolean function which is -1 iff $a \geq b$.

Next we introduce the concept of a sparse delta polynomial. A polynomial is called *sparse* if it is the sum of at most $n^{O(1)}$ monomials. For a vector $\epsilon = \{\epsilon_1, \ldots, \epsilon_n\}$, where $\epsilon_i \in \{-1, 1\}$, and for a positive real c, we call a polynomial $P(x_1, \ldots, x_n)$ a *delta polynomial* for ϵ and c if there are two positive constants d and e satisfying $\frac{d}{e} \geq c$ such that:

(i) $P(\epsilon_1, \ldots, \epsilon_n) = d$ and

(ii) For all $(x_1, \ldots, x_n) \in \{-1, 1\}^n$ which satisfies $(x_1, \ldots, x_n) \neq \epsilon$, $|P(x_1, \ldots, x_n)| \leq e$.

Our construction of delta polynomials can be obtained by using linear error-correcting codes over $GF(2)$ with length which is polynomial in the dimension and with the property that the Hamming weight of any non-zero codeword is sufficiently close to half the length. Let $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq t}$ be the generator 0, 1-matrix of a linear error-correcting code of length t and dimension n, and suppose that the Hamming weight of each non-zero codeword is between $(1 - \epsilon)\frac{t}{2}$ and $(1 + \epsilon)\frac{t}{2}$. Let $P_A = P_A(x_1, \ldots, x_n)$ be the polynomial defined by $P_A(x_1, \ldots, x_n) = \sum_{j=1}^{t} \prod_{i; a_{ij}=1} x_i$. Clearly $P_A(1, \ldots, 1) = t$, and it is not difficult to check that for every $(x_1, \ldots, x_n) \in \{-1, 1\}^n$ which is not $(1, \ldots, 1)$, $|P_A(x_1, \ldots, x_n)| \leq \epsilon t$, since $P_A(x_1, \ldots, x_n)$ is precisely the difference between the number of 0's and the number of 1's in the codeword defined by the sum (in $GF(2)$) of all rows i of A such that $x_i = -1$. Linear codes as above with length t polynomial in the dimension n and with ϵ inverse polynomial in the dimension are the duals of BCH codes [4], as well as other more recent constructions that have applications in derandomization of algorithms [2, 5].

The following theorem gives the construction for COMPARISON which is based on a sparse delta polynomial with n variables denoted by $\hat{P}(\cdot)$.

Theorem 1 *Let* $m_k(X, Y) = \hat{P}(x_n y_n, x_{n-1} y_{n-1}, \cdots, x_{k+1} y_{k+1})$. *Define* $\hat{C}(X, Y) = m_0(X, Y) + \sum_{i=1}^{n} (y_i - x_i) m_i(X, Y)$. *Then* $C(X, Y) = sign(-\hat{C}(X, Y))$.

References

[1] N. Alon and J. Bruck, *Explicit Constructions of Depth-2 Majority Circuits for Comparison and Addition*, IBM Research report, RJ8300, August 1991. To appear in SIAM J. on Disc. Math..

[2] N. Alon, O. Goldreich, J. Hastad and R. Peralta, *Simple constructions of almost k-wise independent random variables*, Proc. 31st IEEE FOCS, St. Louis, Missouri, IEEE (1990), 544-553.

[3] J. Bruck and R. Smolensky, *Polynomial Threshold Functions, AC^0 Functions and Spectral Norms*, SIAM J. on Computing, Vol. 21, No. 1, pp. 33–42, February 1992.

[4] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, North-Holland, 1977.

[5] J. Naor and M. Naor, *Small-bias probability spaces: efficient constructions and applications*, Proc. 22nd Annual ACM STOC, 1990, ACM Press, 213-223.

[6] K. Y. Siu and J. Bruck, *On the Power of Threshold Circuits with Small Weights*, SIAM J. on Disc. Math., Vol. 4, No. 3. pp. 423–435, August 1991.

Capacity of Two-Layer Networks with Binary Weights

Chuanyi Ji* and Demetri Psaltis**
*Department of Electrical Computer and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
**Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

Abstract

The capacity C_b of two layer ($N-2L-1$) feedforward neural networks is shown to satisfy the relation $O(\frac{W}{lnW}) \leq C \leq O(W)$. Here $N-2L-1$ stands for the networks with N input units, $2L$ hidden units and one output unit. W is the total number of weights of the networks. The weights take only binary values and the hidden units have integer thresholds.

Summary

The motivation for this work comes from hardware implementation of neural networks. When weights of neural networks are implemented, both their accuracy and magnitude have to be limited. Then a natural question to ask is whether the learning capability of neural networks will thus be affected.

Learning capability of neural networks can be characterized by their information capacity[2], which is defined as the total number of dichotomies implementable by a class of networks of the same architecture. The capacity C of two layer $N-L-1$ feedforward networks with *analog* weights has been shown to satisfy the relation $O(W) \leq C \leq O(WlnL)$[1]. Here W, L and N are the total number of weights, the number of hidden units and the input dimension, respectively. It remains an open question, however, what the capacity of multilayer networks would be if their weights can only take discrete values. In this work we answer this question by evaluating the capacity of two layer $N-2L-1$ feedforward networks (N inputs, $2L$ hidden units and 1 output) with binary weights and integer thresholds for the hidden units.

Specifically, upper and lower bounds for the capacity C_b of such networks are established in two steps. First, the statistical capacity[3] of a specifically constructed network is evaluated and found to be $O(\frac{W}{ln\,W})$, where W is the total number of weights of the network. It is used as a lower bound for the capacity C_b. Then an upper bound is obtained through a simple counting argument, and shown to be $O(W)$. Therefore, we have $O(\frac{W}{lnW}) \leq C_b \leq O(W)$.

This result shows that reducing the analog weights to only binary values, the capacity of two-layer networks is reduced by at most a log factor. This is consistent to what has been found for a single neuron with binary weights[4]. Therefore, even with binary weights only, multilayer neural networks still have strong learning capability.

References

[1] E. Baum, " On the Capacity of Multilayer Perceptron," *J. of Complexity*, 1988.

[2] T.M. Cover, " Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Trans. Elec. Comp.*, EC-14, pp 326-334, June, 1965.

[3] R.J . McEliece, E.C . Posner, E.R . Rodemich, S.S . Venkatesh, "The Capacity of the Hopfield Associative Memory," *IEEE Trans. Inform. Theory*, Vol. IT-33, No. 4, pp 461-482, July 1987.

[4] S. Venkatesh, "Directed Drift: A New Linear Threshold Algorithm for Learning Binary Weights On-Line," *Journal of Computer and Systems Sciences*, in press.

A NEW FIXED-RATE QUANTIZATION SCHEME BASED ON ARITHMETIC CODING*

Ahmed S. Balamesh and David L. Neuhoff
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, Michigan 48109

Abstract

The construction of fixed-rate vector quantizers based on entropy-coded scalar quantizers has been suggested in [1]. In [2], a particular case of such quantizers based on prefix-coded scalar quantizers was suggested and a very simple binary encoding scheme was presented as well as several reduced-complexity search methods. In this paper, we use arithmetic coding as the binary encoding scheme and show that good performance and low complexity are attained.

Summary

The basic idea is as follows : Let $\underline{q} = (q_1, q_2, \ldots, q_m)$ be a vector of levels such that $q_1 < q_2 < \cdots < q_m$. Let f_e be an encoding scheme that encodes n-dimensional vectors from $\mathcal{Q}^n = \{q_1, q_2, \ldots, q_m\}^n$ into binary strings, and for any such vector $\underline{y}$, let $\ell(\underline{y}) = \ell(f_e(\underline{y}))$ be the length of the corresponding binary string. We define the quantizer codebook as $\mathcal{C} = \{\underline{y} \in \mathcal{Q}^n : \ell(\underline{y}) \leq nr\}$, where r is the desired rate. We assume that the produced binary string can be completed to length $\lfloor nr \rfloor$ with properly-chosen, dummy bits without affecting the decodability. In such a case, using f_e, $\mathcal{C}$ can be encoded with $\lfloor nr \rfloor$ bits, i.e. with a rate of $\lfloor nr \rfloor / n \approx r$ bits per dimension.

With proper choices of f_e, one can easily see that the quantizers in [1] and [2] fit the above description.

An important issue here is quantizing a given n-dimensional, source vector $\underline{x}$, i.e. finding a vector in $\mathcal{C}$ nearest to $\underline{x}$. The complexity of this task is dependent on the structure of $\mathcal{C}$ which is determined by f_e. In the case of [1, 2], the condition $\ell(\underline{y}) \leq nr$ is replaced by a condition of the form $\sum_i l(y_i) \leq L$, where $l(\cdot)$ is a length function defined on $\mathcal{Q}$ that assumes positive integer values. In this case, it was shown in [1] that quantization can be performed using a dynamic programming search. In [2], several reduced-complexity, suboptimal search methods have been proposed. In particular, the Lagrange-multiplier-based (LM-based) method of [2] can be applied to more general cases like the case of arithmetic coding, discussed below.

Now, we consider the case when f_e is an arithmetic encoding rule. The use of arithmetic codes is motivated by their simplicity and the fact that arithmetic coding has a rate very close to the entropy of the encoded source.

The major problem with arithmetic coding is that it is hard to precisely calculate the number of bits produced unless the actual encoding is done. Therefore, we modify the definition of $\mathcal{C}$ above so that only a simple upper bound to $\ell(\underline{y})$ is constrained. For this purpose, we use Pasco's arithmetic codes [3], which permit a *tighter upper bound to* $\ell(\underline{y})$.

In detail, let $\hat{p}(q_j)$ be a J-bit, positive fraction corresponding to q_j. For a given positive integer K, an arithmetic code based on Pasco's method [3] will encode $\underline{y}$, in such a way that $\ell(\underline{y}) \leq \lceil -\sum_i \log_2 \hat{p}(y_i) + nV(K) \rceil \leq \lfloor -\sum_i \log_2 \hat{p}(y_i) + 1 + nV(K) \rfloor$ bits where $V(K) = -\log_2(1 - 2^{1-K})$. The quantity $nV(K)$ can be made as small as desired by increasing K; however, this will increase the complexity of the arithmetic code, since K together with J determine the precision of the arithmetic operations involved. Therefore, we replace the condition $\ell(\underline{y}) \leq nr$ in the definition of $\mathcal{C}$ by the condition $\lfloor -\sum_i \log_2 \hat{p}(y_i) + 1 + nV(K) \rfloor \leq nr$. We call the resulting systems, *arithmetic-encoded, block-constrained quantizers* (ae-BCQ).

The major disadvantage of Pasco's implementation is that it uses multiple-precision arithmetic. We solve this problem by showing that Pasco's codes can be implemented with fixed-point arithmetic using $(J + K + 1)$-bit registers, in a way very similar to the implementation of Jones [4].

Table 1 shows the performance of ae-BCQ for various values of n and r obtained by simulation for IID Gaussian and Laplacian sources and mean-squared-error distortion. The systems are optimized using variations of the methods described in [1, 2] and the initial parameters are based on the optimal, entropy-constrained, scalar quantizer (ECSQ) of the given rate. The quantization is performed using the LM-based method of [2]. From the table, we also see that the SNR of ae-BCQ with $n = 128$ is comparable to the SNR of the pe-BCQ scheme of [2] with $n = 192$ (a little better at low rates, a little worse at high rates). The latter has much higher quantization complexity, and somewhat less encoding complexity. Also, with very little loss in SNR, LM-based method can be used with pe-BCQ with the same complexity advantages [2]. The original SVQ of [1] with $n = 32$ has almost the same search complexity as pe-BCQ with $n = 192$ and significantly larger encoding complexity.

References

[1] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantizer derived from variable-length encoded scalar quantizers," *IEEE Trans. Inform. Theory*, to appear.

[2] A. S. Balamesh and D. L. Neuhoff, "Block-constrained methods of fixed-rate, entropy-coded, scalar quantization," *submitted to IEEE Trans. Inform. Theory.*

[3] R. C. Pasco, "Source coding algorithms for fast data compression," *Ph.D. dissertation*, Dept. Elec. Eng., Stanford Univ., CA, 1976.

[4] C. B. Jones, "An efficient coding system for long source sequences," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 280-291, May 1981.

Table 1: SNR (dB) for ae-BCQ and other methods

Rate	ae-BCQ					pe-BCQ	SVQ	ECSQ
r	$n = 32$	48	64	96	128	$n = 192$	$n = 32$	
IID Gaussian Source								
1.0	4.18	4.11	4.31	4.36	4.50		4.67	4.64
2.0	9.69	9.93	10.06	10.19	10.26	10.16	10.43	10.55
3.0	15.28	15.57	15.74	15.92	16.02	16.08	16.00	16.56
IID Laplacian Source								
1.0	4.18	5.33	5.33	5.52	5.60		5.61	5.76
2.0	10.25	10.52	10.71	10.91	11.00	10.80	10.73	11.31
3.0	15.51	15.85	16.08	16.31	16.42	16.52	16.05	17.20

*This work was supported in part by a scholarship from King Abdul-Aziz University, Jeddah, Saudi Arabia and by NSF grant NCR-9105647.

Statistics of the Binary Quantizer Error in Sigma-Delta Modulation with I.I.D. Gaussian Input

Timo Koski
Department of Mathematics
Luleå Institute of Technology
S-95187 Luleå SWEDEN

ABSTRACT

Representations and statistical properties of the process $\overline{z}$ defined by

$$z_{n+1} = \lambda(z_n + \xi_n),$$

where

$$\lambda(u) := u - b \cdot sign(u) + m$$

are given, when ξ_n is a Gaussian white noise. The process $\overline{z}$ represents the binary quantizer error in a model for (single loop) Sigma Delta modulation, see [3, 6]. The existence and uniqueness of an invariant probability measure, ergodicity properties as well as the existence of moments w.r.t. the invariant probability are proved using Markov process theory. Considering $\overline{z}$ as a random perturbation of the orbits of

$$s_{n+1} = \lambda(s_n)$$

the structure of the power spectrum of the quantizer error is studied approximately for small values of the white noise variance using the deterministic signal s_n under a uniform invariant distribution.

Summary

The binary quantiser error process $\overline{z} := \{z_n\}$ *as defined* above is a real valued, *discrete time Markov process.* An invariant probability measure denoted by π, is a probability measure satisfying $\pi(A) = \int_R \pi(dv)\mathrm{P}^1(v; A)$, for any (Borel) set A, where $\mathrm{P}^1(v; A)$ designates the one step transition probability of the binary quantizer error.Then we have:

Let $\lambda(u) = u - Q(u) + m$, ξ_n be a Gaussian white noise with the variance σ^2 and let $U_{n+1} = \lambda(U_n) + \xi_n$. Then the binary quantizer error, $z_n = \lambda(U_n)$ has a unique invariant probability if and only if $|\, m \,| < b$.

The process $\overline{z}$ is thus *positively recurrent.* We also show that if $b - |\, m \,| > \frac{2\sigma}{\sqrt{2\pi}}$, then the process $\overline{z}$ has moments of all orders w.r.t. the invariant probability measure. The exponential moments $E_\pi[e^{\kappa|z_n|}]$ do not exist for large κ, which shows that the invariant probability π differs substantially from the normal distribution.

A study of the detail in [1] and the representations in [2, 3] yields the stationary characteristic function $\Psi_{\overline{z}}(\omega)$ of the binary quantizer error process $\overline{z}$ as

$$\Psi_{\overline{z}}(\omega) := E\left[e^{j\omega\overline{z}_n}\right] = e^{-j\cdot m\cdot\omega} \cdot \frac{\sin(b\cdot\omega)}{b\cdot\omega} \cdot \Theta(\omega)$$

with

$$\Theta(\omega) := e^{\left(2\sum_{l=1}^{\infty} \int_0^{\infty} (\cos(\omega z) - 1)\cdot\varphi(z + l\cdot b; \sqrt{l}\cdot\sigma)dz\right)},$$

where $j := \sqrt{-1}$ and $\varphi(x;\sigma) := \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$. Evidently this shows that z_n can in the stationary state be split into a sum of two independent random variables,

$$z_n \overset{distr}{=} g_n + m + o_n \quad \textit{(Fine's Decomposition)},$$

where g_n ("granular noise") is uniformly distributed in $[-b, +b]$ and o_n ("slope overload noise") has the distribution determined by $\Theta(\omega)$ and $\overset{distr}{=}$ means equality in distribution. Another important formula here is the following *explicit solution* of $e^*_{n+1} = \lambda(e^*_n)$ due to [2, Thm.1, eq. (3.6), p. 485]:

$$e^*_n = 2b\langle x + n\beta\rangle + m - b, \textit{(Gray's formula)},$$

where $\langle x\rangle := x - \lfloor x \rfloor$ is the *fractional part of* x, $x := \frac{(e^*_0 + b - m)}{2b}$, and $\beta := \frac{b+m}{2b}$. Here we must assume $|\, m \,| < b$.

Using these decompositions of z_n and *Hermite expansions of the nonlinearities* involved we show that the power spectrum of the binary quantizer error has distinct peaks at integer multiples of $\pm\frac{b-m}{2b}$ (frequency scaled to $[-1/2, 1/2]$) for small values of σ. This follows by investigating the autocorrelation

$$r_{\overline{z}}(n) := E[z_n \cdot z_0] = -\sum_{i=0}^{n-1} E[Q(z_i + \xi_i)\cdot z_0] + \\ + n\cdot m\cdot E[z_0] + E\left[z_0^2\right].$$

Here we calculate using e^*_i, which is a function of e^*_0 in view of Gray's formula and since $e^*_i \overset{distr}{=} g_i + m$, and obtain after conditioning on e^*_0, as the Gaussian noise is independent of the sequence e^*_i that

$$E[Q(e^*_i + \xi_i)\cdot g_0] = b\cdot E_{g_0}\left[erf\left(e^*_i/\sqrt{2}\sigma\right)\cdot g_0\right],$$

where E_{g_0} denotes the expectation with respect to the uniform and invariant distribution of the random variable g_0. *Inserting Gray's formula and using* $e^*_0 - m \overset{distr}{=} g_0$ this gives

$$E[Q(e^*_i + \xi_i)\, g_0] = \frac{1}{2b}\int_{-b}^{b} x f(x + i(m+b))dx,$$

where we have introduced the auxiliary function $f(x) := b\cdot erf\left(\left\{2b\langle\frac{x+b}{2b}\rangle + m - b\right\}/\sqrt{2}\sigma\right)$. Since $\langle x+1\rangle = \langle x\rangle$ for every $x \in R$, $f(\cdot)$ has the period $2b$. Using the (complex) Fourier coefficients corresponding to $f(\cdot)$ we derive that if $\frac{b-m}{2b}$ be a nonrational number then $r_{\overline{z}}(n) := E[z_n \cdot z_0]$ can be expressed as

$$r_{\overline{z}}(n) \approx b^2 \sum_{k=-\infty, k\neq 0}^{\infty} \frac{1}{(\pi k)^2}\cdot\left[e^{-j\pi k\cdot n\left(\frac{b-m}{2b}\right)} - 1\right]$$

for $n > 0$. This agrees with the studies of the deterministic spectrum of the binary quantizer error, see [5], as well as with the simulated results.

References

[1] T.L. Fine: The Response of a Particular Nonlinear System with Feedback to Each of Two Random Processes. *IEEE Trans. Inf. Th.*, 14, 1968, pp. 255 - 264.

[2] R.M. Gray: Oversampled Sigma-Delta Modulation. *IEEE Trans. Comm.*, 35, 1987, pp. 481 - 489.

[3] R.M. Gray: *Source Coding Theory.* Kluwer Academic Publishers, Boston, 1990.

[4] R.M. Gray: Quantization Noise Spectra. *IEEE Trans. Inf. Th.*, 36, 1990, pp. 1220 - 1244.

[5] J.C. Kieffer: Analysis of DC Response for a Class of One-Bit Feedback Encoders. *IEEE Trans. Comm.*, 38, 1990, pp.337 - 341.

[6] P.W. Wong and R.M. Gray: Sigma -Delta Modulation with i.i.d. Gaussian Inputs. *IEEE Trans. Inf. Th.*, 36, 1990, pp.784 - 778.

Design of Entropy Constrained Multiple-Decription Scalar Quantizers*

J. Domaszewicz and V. Vaishampayan
Department of Electrical Engineering
Texas A&M University
College Station, TX 77843

The multiple descriptions problem is a generalization of the problem of source coding subject to a fidelity criterion [1]. In its simplest form, two channels, each with their own rate constraints, connect the source to the user. Either channel may be broken at any given time. The objective is to design a source code that minimizes the average distortion when both channels work, subject to constraints on the average distortion when only one channel works. The rate distortion region for a memoryless Gaussian source and squared-error distortion measure, for the multiple description problem has been derived in [2], [3]. Surprisingly, despite strong potential applications to speech and video transmission over packet switched networks and to digital mobile telephony, the design of such codes has received little attention. Jayant [4] considers the design of a system based on subsampling, for packetized speech.

We wish to encode the output of a stationary, ergodic and memoryless source which is represented by the random process $\{X_n,\ n \in \mathbb{Z}\}$ with zero mean, variance σ^2 and known probability density function (pdf). The entropy-coded multiple description scalar quantizer (ECMDSQ) is illustrated in Fig. 1. Let $\mathcal{I} = \{1, 2, \ldots, N\}$, $\mathcal{I}_1 = \{1, 2, \ldots, M_1\}$, $\mathcal{I}_2 = \{1, 2, \ldots, M_2\}$ and assume that $N \leq M_1 M_2$. The source sample x is mapped by $q(\cdot)$ to the index n that takes values in $\mathcal{I}$. The operation of $q(\cdot)$ can be described in terms of a vector of thresholds $\mathbf{t} = (t_1, t_2, \ldots, t_{N-1})$, $t_1 \leq t_2 \leq \ldots \leq t_{N-1}$ by the equation $q(x) = n$ if $x \in [t_{n-1}, t_n)$, $n = 1, 2, \ldots, N$, where $[t_0, t_N)$ is the support of the source pdf. The index n is mapped to indices $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_2$ by $a_1(\cdot)$ and $a_2(\cdot)$, respectively. The mapping (a_1, a_2) is called the index assignment. We associate with each channel a variable length code $C_m = \{c_{mi}, i = 1, 2, \ldots, M_m\}$, $m = 1, 2$, where each codword c_{mi} is a binary string of length l_{mi}, $i = 1, 2, \ldots, M_m$, $m = 1, 2$. Indices i and j are mapped by variable length encoders γ_1 and γ_2 to codewords c_{1i} and c_{2j}, and transmitted over Channel 1 and Channel 2, respectively. If only Channel 1 (Channel 2) works, the index i (j) is recovered by the variable length decoder and mapped by side decoder g_1 (g_2) to real number y_1 (y_2) which takes values in the reconstruction codebook $\mathcal{Y}_1 = \{y_{1i},\ i \in \mathcal{I}_1\}$ ($\mathcal{Y}_2 = \{y_{2j},\ j \in \mathcal{I}_2\}$). If both channels work, central decoder g_0 maps the index pair (i, j) to a real number y_0 which takes values in the reconstruction codebook $\mathcal{Y}_0 = \{y_{0n},\ n \in \mathcal{I}\}$.

Let $d_m(x, y_m)$ be the per-sample distortion between the source sample and the output of the mth decoder, $m = 0, 1, 2$. We refer to d_0 as the central distortion and to d_1 and d_2 as the side distortions. The average central and side distortions are denoted by $\bar{d}_0$ and $\bar{d}_1$ and $\bar{d}_2$.

For given values M_1, M_2, D_1, D_2, R_1 and R_2 a multiple description scalar quantizer is said to be optimal subject to entropy constraints, if it minimizes $\bar{d}_0$ subject to $\bar{d}_1 \leq D_1$, $\bar{d}_2 \leq D_2$, $H_1 \leq R_1$ and $H_2 \leq R_2$.

We derive necessary conditions for optimality, and present an iterative design algorithm for locally optimal ECMDSQ's. The assignment of index pairs to the quantizer bins—a crucial step in the design—is also addressed. Convergence of the algorithm is proved. As a reference system, we consider a multiple description scalar quantizer (MDSQ) system in which fixed length binary codes are used to transmit an index pair [5]. We also make comparisons against the optimum performance theoretically attainable (OPTA) [2], [3]. Our results indicate that significant performance improvements are obtained over the MDSQ. For example, with $R_1 = R_2 = 4.0$ bits/sample/channel and $M_1 = M_2 = 24$, ECMDSQ achieves a given side distortion at a central distortion which is 4.5 dB better than that of the MDSQ. Comparisons against the OPTA indicate that for a given side distortion further gains of 3 dB for the central distortion can be achieved over ECMDSQ.

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. part 4, pp. 142–163, March 1959.

[2] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Th.*, vol. IT-28, pp. 851–857, November 1982.

[3] L. Ozarow, "On a source coding problem with two channels and three receivers," *The Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, December 1980.

[4] N. S. Jayant, "Subsampling of a DPCM speech channel to provide two "self-contained" half-rate channels," *Bell Syst. Tech. J.*, vol. 60, pp. 501–509, April 1981.

[5] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, accepted for publication.

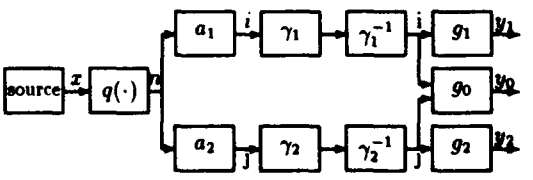


Figure 1: Block diagram of the ECMDSQ system.

*This work was supported in part by NSF Grant Number NCR-9104566.

ENUMERATION ENCODING AND DECODING ALGORITHMS FOR PYRAMID TRELLIS CODES†

Thomas R. Fischer and Jianping Pan
School of Electrical Engineering and Computer Science
Washington State University
Pullman, WA 99164-2752

ABSTRACT

A pyramid source code is a code that assigns equal-length binary strings to all reproduction codewords of equal (weighted) ℓ_1 norm, and finds application in the encoding of Laplacian-distributed data. A pyramid source encoding is partitioned into two concatenated mappings; the first from source word to reproduction codeword within a codebook; the second from the reproduction codeword to a binary string. The first mapping allows distortion and is accomplished using trellis coded quantization. The second mapping is noiseless and is denoted as enumeration. Efficient pyramid enumeration encoding and decoding algorithms are presented, for use with fixed-rate or variable-rate pyramid trellis codes.

SUMMARY

A pyramid source code is a code that assigns equal-length binary strings to all reproduction codewords of equal (weighted) ℓ_1 norm. Such codes are well-suited for encoding Laplacian data [1]-[3] and find application in transform and sub-band image coding [4]-[7].

A pyramid source encoding can be partitioned into two concatenated mappings; the first from source word to reproduction codeword within a codebook, and the second from the reproduction codeword to a binary string. The first mapping typically allows distortion and is referred to herein as quantization or compression. The second mapping is lossless, and is referred to as enumeration. Enumerative source coding was introduced by Cover [8] for the lexicographic ordering of n-tuples. The ordering developed in this paper is different due to the pyramid formulation and the trellis structure. The pyramid codes use trellis coded quantization [9]. The contribution of the paper is to describe efficient pyramid enumeration encoding and decoding algorithms. Fixed-to-fixed length and fixed-to-variable length pyramid trellis codes are easily constructed using the enumeration algorithms.

Trellis coded quantization (TCQ) [9] is an efficient, low-complexity source coding technique that, when used with entropy coding [10],[11], can provide near-optimum rate-distortion performance for a broad class of memoryless sources. The uniform pyramid trellis codes described here use a (possibly scaled or translated) subset of the integer lattice $\mathbf{Z}$ as the codebook, partitioned into 2^{m+1} subsets, with integer $m \geq 1$. The subsets are assigned to the 2^m branches leaving each state in an N-state trellis defined by a rate-$m/(m+1)$ convolutional encoder. The entropy-constrained TCQ results in [10] indicate that most of the available granular gain is achieved with $m = 1$.

Consider the integer lattice $\mathbf{Z}$, partitioned into 4 subsets, D_j, $j = 0,1,2,3$, as shown in Figure 1. The lattice point z is in D_j if and only if $z \bmod 4 = j$. A time-invariant labeling assigns one subset to every branch leaving each trellis state in an N-state trellis. For each time step, i, a positive integer weight, say w_i, is assigned to the trellis transition. If $\mathbf{y} = [y_1, y_2, \cdots, y_L]^T$ is the sequence of codewords corresponding to an L-step trellis path, then the weighted ℓ_1 norm of the path symbols (the length-L path weight) is given by

$$\|\mathbf{y}\|_{1,\mathbf{w}} = \sum_{i=1}^{L} w_i |y_i|.$$

For any fixed initial and final trellis states, say s and t, denote the number of paths that begin in state s, end in state t, and have weight k as $N(s,t,L,\mathbf{w},k)$, $k = 0,1,\cdots$. Computation of the weight enumeration of the code is the basis for the enumeration encoding and decoding.

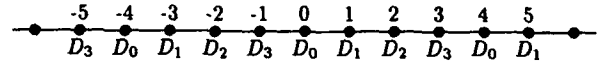


Fig. 1. Integer lattice codebook and partition into subsets.

A pyramid trellis encoding first maps an input sequence into the sequence of reproduction letters corresponding to the minimum distortion path through the trellis. Fixed- and variable-length enumeration codes are considered for mapping the sequence of reproduction letters to binary strings. In the first, consecutive L-tuples of the trellis reproduction sequence are mapped to consecutive fixed-length strings of (integral) RL bits. The trellis search is constrained to allow only sequences of encoded symbols of pre-selected weights. That is, if the length-L trellis path begins in state s and ends in state t, then all code sequences $\mathbf{y}$ must have one of M possible weights, say k_i, $i = 1,\cdots,M$, such that

$$\sum_{i=1}^{M} N(s,t,L,\mathbf{w},k_i) \leq 2^{RL}.$$

If $k_i = i-1$, $i = 1,\cdots,M$, then the trellis codewords have been shaped in sequence space so that each L-tuple lies within a pyramid. By partitioning the integers $0,\cdots,2^{RL}-1$ into M consecutive sets, each of size $N(s,t,L,\mathbf{w},k_i)$, it is seen that the essence of the enumeration is simply to map each of the $N(s,t,L,\mathbf{w},k_i)$ sequences of weight k_i to the integers $0,\cdots,N(s,t,L,\mathbf{w},k_i)-1$.

In the variable-length encoding, the trellis is searched in the usual way to find the minimum distortion path. Then, if the length-L path has weight k, the codeword $\mathbf{y}$ is assigned a binary string, say $b(\mathbf{y})$, of length $\lceil \log_2 N(s,t,L,\mathbf{w},k) \rceil$ bits. A prefix-free code is used to encode k as $c(k)$, and the binary string representing $\mathbf{y}$ is $(c(\|\mathbf{y}\|_{1,\mathbf{w}}), b(\mathbf{y}))$.

† This work was supported, in part, by the National Science Foundation under Grants NCR-8821764 and MIP-9247526.

REFERENCES

[1] T. R. Fischer, "A Pyramid Vector Quantizer," *IEEE Trans. Inform. Th.*, Vol IT-32, pp. 568–583, July 1986.
[2] T. R. Fischer, "Geometric source coding and vector quantization," *IEEE Trans. Inform. Th.*, vol. IT-35, pp. 137-145, Jan. 1989.
[3] D.-G. Jeong and J. D. Gibson, "Uniform and piecewise unifrom lattice vector quantization for memoryless Gaussian and Laplacian sources," *IEEE Trans. Inform. Th.*, to appear.
[4] H.-C. Tseng and T. R. Fischer, "Transform and Hybrid Transform/DPCM Coding of Images Using Pyramid Vector Quantization," *IEEE Trans. Commun.*, Jan., 1987.
[5] D.-G. Jeong and J. D. Gibson, "Image coding with uniform and piecewise-uniform vector quantizers," submitted to *IEEE Trans. Image Processing*, May 1992.
[6] M. Barlaud, P. Sole, M. Antonini, P. Mathieu, and T. Gaidon, "Pyramidal lattice vector quantization for multiscale image coding," submitted to *IEEE Trans. Image Processing*, May 1992.
[7] M. E. Blain and T. R. Fischer, "A comparison of vector quantization techniques in transform and subband coding of imagery," Elsevier *Signal Processing: Image Communication*, vol. 3, pp. 91-105, 1991.
[8] T. M. Cover, "Enumerative Source Encoding," *IEEE Trans. Inform. Th.*, Vol IT-19, pp. 73-77, Jan. 1973.
[9] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. COM-38, pp. 82-93, Jan. 1990.
[10] T. R. Fischer and M. Wang, "Entropy-constrained trellis-coded quantization," *IEEE Trans. Inform. Th.*, vol. IT-38, pp. 415-426, Mar. 1992.
[11] M. W. Marcellin, "On entropy-constrained trellis-coded quantization," *IEEE Trans. Commun.*, to appear.

Information Rates of Pre/Post Filtered Dithered Quantizers

Ram Zamir and Meir Feder

Dept. of Electrical Engineering - Systems, Tel-Aviv University
Tel-Aviv, 69978, ISRAEL

Abstract

In this work the improvement in rate-distortion performance of an entropy coded dithered uniform (or lattice) quantizer, incorporating appropriate pre/post filters, is shown and analyzed. The proposed scheme attains good coding performance, under MSE criterion, for any source distribution, although its design depends only on the second order statistics of the source.

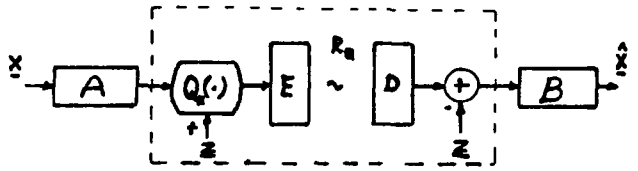


Figure 1: The Pre/Post Filtered Coded Dithered Quantization Scheme

We examine the enhancement in performance achieved by incorporating pre/post linear filters into the universal coding scheme, composed of a dithered quantizer and a lossless (entropy) coder. We assume that the second order statistics of the source are known, and a Mean Square Error (MSE) criterion is used. Considering Figure 1, the source is denoted $\underline{X} \in \mathcal{R}^n$, its reconstructed value is $\hat{\underline{X}}$, the (coded) dithered quantizer is described by the dashed area in the figure, and the matrices A and B represent the pre and post filters, respectively. $Q_k(\cdot)$ is a uniform scalar or k-dimensional lattice quantizer, characterized by G_k - the normalized second moment of its lattice. The pseudo-random dither $\underline{Z}$ is distributed uniformly over the basic Voronoi cell of the lattice (e.g., over $(-\Delta/2, \Delta/2)$ when a scalar uniform quantizer with a step size Δ is used) and is available to both encoder and decoder. The quantizer output is encoded, conditioned on the dither values, by a possibly universal, lossless (entropy) coder. The coding rate of the scheme is thus assumed to be the quantizer entropy, conditioned on the dither, i.e., $R_Q = \frac{1}{n} H(Q_k(A\underline{X} + \underline{Z})|\underline{Z})$ (see [1]).

As shown in [1], the coded dithered quantizer is equivalent to an additive noise channel in rate-distortion sense. The quantization noise is independent of the source and it is distributed as $-\underline{Z}$, which implies that the overall MSE distortion of the scheme is $D = \frac{1}{n} E\|B(A\underline{X} - \underline{Z}) - \underline{X}\|^2$. The coding rate of the scheme is the mutual-information between the input and the output of the equivalent additive noise channel, which in our case can be written as $R_Q = \frac{1}{n} I(A\underline{X}; A\underline{X} - \underline{Z})$.

Intuitively speaking, coding performance are enhanced by incorporating pre/post filtering, since one may allow at first higher distortion in coding - and thus save rate - relying on noise power reduction at the reconstruction by the post filter. Specifically, in designing the quantization scheme we try to simulate the structure of the optimal *"forward channel"*, achieving the rate-distortion function of a Gaussian source (see e.g. [2]). Now, unlike the "forward channel" realization, which combines filtering and additive *Gaussian* noise, the additive quantization noise in the scheme is usually not Gaussian. Nevertheless, we suggest to use the optimal filters of the Gaussian case, and here we examine the resulting performance for an arbitrary source and the actual quantization noise.

The redundancy of the scheme over the rate-distortion function of the source is defined as

$$\rho = R_Q(D) - R(D) \quad . \tag{1}$$

We derive the following bounds for this redundancy:

1. For any source, which is $\mathcal{D}$ bits away from Gaussianity,

$$\rho \leq \mathcal{D} + \frac{1}{2} \log 2\pi e G_k \tag{2}$$

where $\mathcal{D} = \frac{1}{n} \int f_{\underline{X}} \log \frac{f_{\underline{X}}}{f_{\underline{X}^*}}$ is the divergence between $\underline{X}$ and $\underline{X}^*$, $f_{\underline{X}}$ is the source density and $f_{\underline{X}^*}$ is a Gaussian density with the same mean and covariance as $f_{\underline{X}}$. Note that for a Gaussian source $\mathcal{D} = 0$, and if we further assume that $G_k \to \frac{1}{2\pi e}$ for lattice quantizer with large dimension, then the scheme achieves the rate-distortion function of the (Gaussian) source.

2. For any source with a density,

$$\lim_{D \to 0} \rho = \frac{1}{2} \log 2\pi e G_k \tag{3}$$

as in dithered quantization without pre/post filtering.

3. For any source with a covariance matrix R_x,

$$\rho \leq C \,|_{S_{in}=D} \tag{4}$$

where C is the power constraint capacity (at input level S_{in} equals to the allowed distortion D) of the equivalent additive channel $\underline{Y} = A\underline{X} - \underline{Z}$, and A is the appropriate pre-filter (for R_x and D). Note that this bound implies low redundancy at high distortion, and in particular it follows that $\rho \to 0$ as D goes to the source average power (since in that case $A \to 0$). In general $C \leq \frac{1}{2} \log 4\pi e G_k$ which is the upper bound for the redundancy of dithered quantization without filters (see [1]).

The combination of (2), (3) and (4) leads to useful bounds on the performance of pre/post filtered dithered quantization in the general case. In figure 2 (A), the information rate curve versus MSE, of a Gaussian source encoded by a pre/post filtered scalar dithered quantizer, is illustrated. The graph is compared to Shannon's rate-distortion function (R(D)) and to the performance of a dithered quantizer without filtering (B). Note that in this case the pre/post filtering saves ~ 0.75 bits at high distortion.

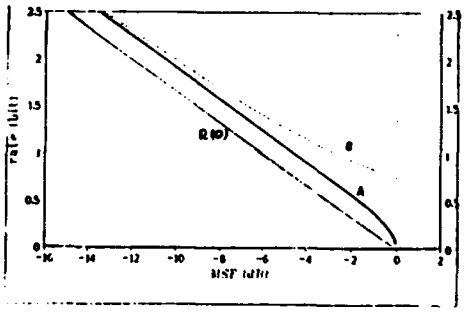


Figure 2: Information rate Curves versus MSE of a Gaussian Source: quantization with and without filtering, and R(D).

References

[1] R. Zamir and M. Feder. On universal quantization by randomized uniform lattice quantizer. *IEEE Trans. Information Theory*, pages 428–436, March 1992.

[2] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

A Frequency Domain Approach to the Optimization of Scalar Quantizers

Marcelo S. Alencar *†
Universidade Federal da Paraiba
Departamento de Engenharia Elétrica
Av. Aprígio Veloso, 882
Campina Grande PB Brasil

Abstract

This article presents a procedure for optimising the scalar quantiser, based on the power spectrum density of the quantisation noise. The input signal is assumed stationary in the wide sense, but no restriction is made concerning its probability density function.

Summary

The usual optimisation techniques for the scalar quantiser are centered on properties of the probability density function (pdf) of the input signal [1] [2]. In fact, there seems to be a tendency of the proposed schemes to obtain an output pdf that more closely resembles the uniform type [3]. Usually, the information on the noise power is sufficient to approach a given problem. Sometimes, as in the case of matched filters design, the shape of the noise spectrum plays a more important role.

The spectrum of the quantisation noise was shown in a recent paper to be quite independent of the spectrum of the applied signal and remarkably related to the probability density function of the signal derivative [4]. A small quantisation step, as well as a uniform quantisation scheme were considered in that proposed model. A generalisation of that model is proposed in this article, to account for the nonuniform case.

Quantisation noise can be thought as the result of the application of the signal $x(t)$ to a circuit with characteristic $f(x)$. The function $f(x)$ is periodic, with period d, as shown below

$$f(x) = x - m.d$$
$$(m - \frac{1}{2})d < x < (m + \frac{1}{2})d, m = 0, \pm 1, \pm 2, \cdots \quad (1)$$

The autocorrelation function of the quantisation noise can be evaluated, as described in [5], and its power spectrum density can be obtained by using the Wiener-Khintchine theorem [6],

$$S_N(w) = \frac{d^2}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} p_{X'}(\frac{wd}{2\pi n}) \quad (2)$$

where $p'_X(.)$ is the probability density function of the derivative of the input signal.

Equation 2 demonstrates that the power spectral density of the quantisation noise is related to the probability density function of the derivative of the input signal. The convergence of the noise spectrum to Equation 2, as the stepsize decreases, is a result of a previous work [7]. The noise spectrum reflects an infinite sum of contributions, each one with the shape of the probability density function, but with decreasing intensity and increasing bandwidth.

For the nonuniform case, one can assume that the signal is transformed by a nonlinear function $g(\cdot)$ prior to the quantisation process. This gives

$$S_N(w) = \frac{d^2}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} p_{g'(x)}(\frac{wd}{2\pi n}) \quad (3)$$

here, $g'(x)$ is the derivative of the compression function.

Careful selection of $g(x)$ can minimise the following expression and maximise the signal to quantisation noise ratio. The compression function must be chosen in order to displace the peak of the quantisation noise spectrum far outside the signal bandwidth.

$$P'_N = \int_{-w_M}^{w_M} S_N(w)dw = \frac{d^2}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{-w_M}^{w_M} p_{g'(x)}(\frac{wd}{2\pi n})dw \quad (4)$$

where P'_N represents the noise power that falls inside the signal bandwidth. This quantity can be made quite small, compared to the total noise $P_N d^2/2$. A procedure for solving Equation 4 involves the linearisation of the function $g(x)$.

References

[1] Stuart P. Lloyd. "Least Squares Quantisation in PCM". *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982.

[2] Joel Max. "Quantising for Minimum Distortion". *IEEE Transactions on Information Theory*, 6(1):7–12, March 1960.

[3] N. S. Jayant and Peter Noll. *Digital Coding of Waveforms*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984.

[4] Marcelo S. Alencar. "Power Spectrum Density of Quantisation Noise for Uniform Quantisers". In *Proceedings of the IASTED International Symposium on Computers, Electronics, Communication and Control*, pages 274–275, Calgary, Canada, April 1991.

[5] Marcelo S. Alencar. "A Model for Evaluating the Quantisation Noise Power Spectral Density". In *Anais do Simpósio Brasileiro de Telecomunicações*, pages 10.4.1–10.4.3, São Paulo, Brasil, Setembro 1991.

[6] Marcelo S. Alencar. "Measurement of the Probability Density Function of Communication Signals". In *Proceedings of the IEEE Instrumentation and Measurement Technology Conference - IMTC'89*, pages 513–516, Washington, D. C., April 1989.

[7] Marcelo S. Alencar and Benedito G. Aguiar Neto. "Estimation of the Probability Density Function by Spectral Analysis: A Comparative Study". In *Proceedings of the Treizième Colloque sur le Traitement du Signal et des Images - GRETSI*,, pages 377–380, Juan-Les-Pins, France, September 1991.

*This research was supported by a grant from the Canadian Institute for Telecommunications Research under the NCE program of the Government of Canada.

†The author is currently with the Department of Electrical and Computer Engineering, University of Waterloo, Canada.

Performance Comparisons of TCQ with Difference Distortion Measures and Short Coding Delays

Min Wang
Rockwell International Corp.
4311 Jamboree Road
Newport Beach, CA 92658-8902

Thomas R. Fischer
School of Electrical Engineering and Computer Science
Washington State University
Pullman, Washington 99164-2752

Abstract

The design of trellis coded quantization (TCQ) to minimize the mean-squared error (MSE) has been generalized to minimize $E\{|x-y|^v\}$ for positive integer v. Simulation results for memoryless uniform and Gaussian sources with $v = 1$ and 4 show that TCQ outperforms scalar quantization (SQ) in the same way as for $v = 2$.

Entropy-constrained trellis coded quantization (ECTCQ) has also been studied for difference distortion measure $\rho(x,y) = |x-y|^v$. Two ECTCQ realizations have been considered. One is to design the TCQ codebook subject to the output entropy of the source encoder. The other is to design the TCQ codebook independent of the output entropy while the followed entropy code is designed based on the probabilities of the (locally) optimized TCQ codewords for the source sequence. The latter is suboptimal but requires less computations. Simulations show that the performance of the TCQ system is generally improved by combining with an entropy encoder. ECTCQ outperforms entropy-constrained scalar quantization (ECSQ) in all cases considered. The performance difference between the two ECTCQ realizations at low output entropy increases as v gets large, but vanishes as the output entropy increases.

TCQ system with short coding delays and squared-error criterion has been studied. Simulations show that, with the same amount of coding delay, the performance of such designed TCQ is comparable to that of vector quantizers (VQ) for the memoryless uniform and Gaussian sources, while slightly inferior to VQ for the memoryless Laplacian source. However, TCQ requires much less computations than VQ, expecially for large vector dimensions or encoding rates.

Summary

Trellis coded quantization (TCQ) [1] is a low-complexity form of trellis coding [2] in which the trellis branches are labeled with reproduction subsets instead of individual reproduction levels. The idea of designing TCQ to minimize the mean-squared error (MSE) [1] is generalized to minimize the average distortion between the input and the output of the quantizer, given a distortion measure $\rho(x,y) = |x-y|^v$ for positive integer v. Let the encoding rate be $R \geq 0$. The TCQ codebook contains $N = 2^{R+R''}$ codewords partitioned into $K = 2^{R'+R''}$ subsets (according to the rules in [1]), each subset of $L = 2^{R-R'}$ codewords. $0 \leq R' \leq R$ and $0 \leq R''$. The N_s-state encoding trellis is defined by a rate-$R'/(R'+R'')$ convolutional encoder with $2^{R'}$ branches entering/leaving each state. Let $\mathcal{X} = \{x_j : j = 1,2,...,\|\mathcal{X}\|\}$ and $\mathcal{Y} = \{y_i\}_{i=1}^{N}$ represent the source (training) sequence and the TCQ codebook, respectively. Let $B_i = \{x_j : x_j \in \mathcal{X} \text{ is encoded as } y_i\}$. Then the TCQ codebook $\mathcal{Y}$ should be designed to satisfy the following conditions

$$\sum_{x_j \in B_i} \mathrm{sgn}(x_j - y_i)(x_j - y_i)^{v-1} = 0, \quad \text{for odd positive integer } v, \quad (1)$$

or

$$\sum_{x_j \in B_i} (x_j - y_i)^{v-1} = 0, \quad \text{for even positive integer } v, \quad (2)$$

for $i = 1,2,...,N$.

The performance of the TCQ system with distortion measure $\rho(x,y) = |x-y|^v, v = 1$ and 4 is evaluated by simulation, for zero mean and unit variance memoryless uniform and Gaussian sources. The encoding trellises used are 4-, 8- and 256-state rate-1/2 Ungerboeck amplitude modulation trellises [3]. The simulation results show that i) TCQ always outperforms the scalar quantizer (SQ); ii) as the number of trellis states increases, the TCQ performance also increases; iii) for the Gaussian source, the gap between the TCQ performance and the performance promised by the Shannon Lower Bound [4] is significant, and increases as R increases; and iv) for the Gaussian source, the improvement of TCQ over SQ increases as v increases. The first three results are the same as those for $v = 2$ in [1].

The performance of TCQ for non-uniformly distributed sources is improved by combining entropy encoding. Such a scheme is called entropy-constrained trellis coded quantization (ECTCQ) [5]. We study ECTCQ with distortion measure $\rho(x,y) = |x-y|^v$. Two kinds of ECTCQ systems are considered. The first realization is the natural ECTCQ (as for $v = 2$ described in [5] and [6]). That is, the TCQ and entropy encoder are jointly designed to minimize the functional [8]

$$J = E\{\rho(x,y)\} + \lambda E\{l(y), \quad (3)$$

where λ is a Lagrange multiplier and $l(y)$ is the number of bits used by the entropy code to represent y. The second realization is based on designing the TCQ alone, in the same way as in the fixed-rate TCQ system, and then designing the entropy encoder based on the probabilities of the (locally) optimized TCQ codewords. The latter approach requires less computations than the former, in the sense of either design or implementation. The updating equation for the TCQ codebook of either ECTCQ realization is given by (1) or (2) depending on the value of v. The entropy encoder is realized as the state-entropy encoder [6], which assigns a single entropy coder for each union codebook of subsets that appear as labels for the branches leaving each trellis state.

Simulation results with $v = 1,2$, and 4 for zero mean and unit variance memoryless Gaussian and Laplacian sources show that the performance improvement of the jointly designed system over the corresponding independently designed system at low output entropy depends on the distortion criterion. The larger v, the larger the improvement. However, such improvement vanishes as the output entropy of the encoder increases. Overall, entropy-constrained techniques outperform the corresponding fixed-rate schemes.

The TCQ with short coding delays is studied. Simulation results with $v = 2$ show that such designed TCQ has comparable performance as vector quantization (VQ) (designed with the clustering algorithm [7]) for the memoryless uniform and Gaussian sources, but is slightly inferior to VQ for the memoryless Laplacian source.

References

1. M.W. Marcellin and T.R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. COM-38, pp.82-93, Jan. 1990.

2. L.C. Stewart, R.M. Gray, and Y. Linde, "The design of trellis waveform coders," *IEEE Trans. Commun.*, vol. COM-30, pp. 702-710, April 1982.

3. G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Th.*, vol. IT-28, pp. 55-67, Jan. 1982.

4. T. Berger, *Rate Distortion Theory*, Prentice-Hall, 1971.

5. T.R. Fischer and M. Wang, "Entropy-constrained trellis coded quantization," *IEEE Trans. Inform. Th.*, vol. IT-38, pp. 416-425, March 1992.

6. M.W. Marcellin, "On entropy-constrained trellis coded quantization," *IEEE Trans. Commun.*, to appear.

7. Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Inform. Th.*, vol. IT-31, pp.106-109, Jan. 1985.

8. P.A. Chou, T. Lookbaugh, and R.M. Gray, "Entropy-Constrained vector quantization," *IEEE Trans. ASSP*, vol. 37, pp. 31-42, Jan. 1989.

ASYMPTOTIC QUANTIZATION FOR NOISY CHANNELS

Steven W. McLaughlin
Electrical Engineering Dept.
Rochester Institute of Technology
Rochester, NY 14623

David L. Neuhoff
EECS Department
University of Michigan
Ann Arbor, MI 48109

Abstract

We consider the problem of asymptotic quantization in conjunction with a noisy binary symmetric channel. For a *noiseless* channel, Bennett's integral is a formula for the distortion of a scalar quantizer given in terms of the source density, the number of quantization points (assumed to be large), and the distribution of quantization points, or *point density*. In this paper we extend Bennett's integral to the case where the quantizer is used in conjunction with a noisy binary symmetric channel, assuming that channel codewords are assigned randomly. We also derive an expression for the optimum noisy channel point density.

Summary

One of Shannon's fundamental results is that source and channel coding can be treated separately without loss of performance. This usually leads to the separate design of source and channel coders (e.g. a source code can be designed assuming a noiseless channel). Shannon's result, however, is one that requires arbitrarily complex source and channel coders, which is not reasonable in practice. The practical channel code cannot guarantee zero error probability and, consequently, the performance of a source code designed for a noiseless channel will degrade when used in conjunction with a noisy channel. Thus one must analyze and design the source code with the noisy channel in mind.

The purpose of this paper is first to develop an expression for the distortion of a quantizer used in conjunction with a noisy binary symmetric channel, and then to find the optimum distribution of quantization points, or *point density*, when the number of quantization points is large and the channel codewords are assigned randomly. In previous work [1] we derived lower bounds to distortion in terms of the point density using a "greedy" codeword assignment. This paper gives more accurate estimates of the distortion caused by the noisy channel. Recently, Zeger and Manzella [2] have derived a similar, but not quite identical expression for distortion, without optimizing the distribution of quantization points.

A noisy channel quantizer is described by a set of N ($= 2^L$) quantization points $C = \{y_i\}_{i=1}^N$, a partition $S = \{S_i\}_{i=1}^N$ of the real line, and a codeword assignment $A = \{c_i\}_{i=1}^N$, where $c_i \in \{0,1\}^L$ is the $L = \log_2 N$ bit codeword assigned to quantization point y_i. Given a source sample u, the encoder determines in which cell S_i the sample u lies and produces an L-bit binary sequence c_i which is transmitted across a binary symmetric channel with crossover probability $q<.5$. The channel output is an L-bit binary sequence c_j. The decoder receives c_j and outputs the quantization point y_j. The mean squared error that results can be written

$$D = D_S + D_C \tag{1}$$

where

$$D_S = \sum_{i=1}^{N} \int_{S_i} (u-y_i^*)^2 \, p(u)\, du \tag{2}$$

$$D_C = \sum_{i=1}^{N} P(U \in S_i) \sum_{j=1}^{N} q_L(c_j/c_i)\,(y_i^*-y_j)^2 \tag{3}$$

where $q_L(c_j/c_i)$ is the probability that c_j is received given that c_i is transmitted, $p(u)$ is the probability density of the source, and $y_i^* = E[U|U \in S_i]$ is the probabilistic centroid of S_i. The first term in (1) (source distortion D_S) is the distortion that results assuming a noiseless channel and codebook consisting of the y_i^*'s. The second term (channel distortion D_C) is the distortion due to channel errors.

Bennett's integral is an asymptotic formula (i.e. for large N) for D_S that depends on the distribution of quantization points, or point density $\lambda(u)$, the source density $p(u)$ and the number of points N. For a scalar quantizer with N large,

$$D_S = \frac{1}{N^2} \int \frac{1}{\lambda(u)^2}\, p(u)\, du \tag{4}$$

This expression can be used to find the optimum (in minimum mean squared error) point density for a noiseless channel, which is found to be

$$\lambda(u) = c\, p^{1/3}(u) \tag{5}$$

where c is a constant independent of N such that $\int \lambda(u)\, du = 1$.

The principal result of this paper is the following expression for the channel distortion D_C in terms of the point density $\lambda(u)$, the number of points N, and the channel crossover probability q, when the codeword assignment is random:

$$D_C = (1-(1-q)^L)\left(\int u^2\, \lambda(u)\, du + \sigma^2 - 2 D_S \right) \tag{6}$$

where σ^2 is the variance of the source and D_S is the source distortion. For a given source density $p(u)$, size N, and channel crossover probability q, the total source plus channel distortion is

$$D = \frac{1}{N^2} \int \frac{1}{\lambda(u)^2}\, p(u)\, du + (1-(1-q)^L)\left(\int u^2\, \lambda(u)\, du + \sigma^2 - 2 D_S \right)$$

One may minimize this expression with respect to the point density $\lambda(u)$ subject to the constraint that $\lambda(u)$ integrates to 1. This is an isoperimetric problem of the calculus of variations which yields the noisy channel point density

$$\lambda^*(u) = \frac{p^{1/3}(u)}{((1-Q)\, u^2 + \rho)^{1/3}}\, c^{1/3} \tag{7}$$

where $c = (4Q-1)/12N^2$, $Q = (1-q)^L$ and ρ is a constant such that $\lambda(u)$ integrates to 1.

To see that the optimum point density in (7) performs better than the point density in (5), consider a uniform source on [-.5,.5]. From (7), the optimum noisy channel point density is

$$\lambda_U^*(u) = \frac{1}{((1-Q)\, u^2 + \rho)^{1/3}}\, c^{1/3} \tag{8}$$

where c, Q and ρ were defined previously. Figure 1 compares the signal-to-quantization noise performance as a function of q for an $N=32$ "Channel-optimized" quantizer with the optimal point density in (8) and a "Non-channel optimized" uniform scalar quantizer (the optimum noiseless channel scalar quantizer) as suggested in [2]. The channel optimized point density performs about 3dB better for sufficiently large channel error probabilities.

References

[1] S.W. McLaughlin and D.L. Neuhoff, "Asymptotic Bounds in Source-Channel Coding," *Proceedings of the 1991 International Symposium on Information Theory*, Budapest, Hungary, July 1991.

[2] K. Zeger and V. Manzella, "Asymptotically Optimal Noisy Channel Quantization via Random Coding," presented at Joint DIMACS/IEEE Workshop on Coding and Quantization, Rutgers, Oct. 1992.

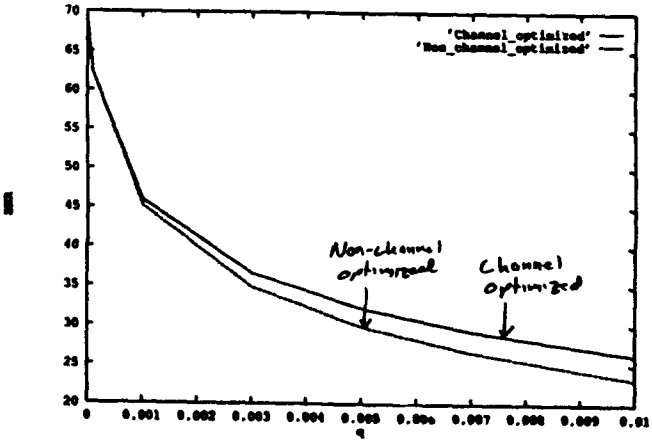


Figure 1: Signal-to-noise ratio for scalar quantizers with uniform and channel-optimized point density, in terms of crossover probability q.

The Design of Finite-State Machines for Quantization Using Simulated Annealing

Ercan Engin Kuruoğlu
Electrical and Electronics Engineering Department
Bilkent University
Ankara, 06533, Turkey

Ender Ayanoğlu
AT&T Bell Laboratories
101 Crawfords Corner Road 4F-507
Holmdel, NJ 07733-3030, USA

ABSTRACT

In this paper, the combinatorial optimization algorithm known as *simulated annealing* is used for the optimization of the trellis structure or the next-state map of the decoder finite-state machine in trellis waveform coding. The generalized Lloyd algorithm which finds the optimum codebook is incorporated into simulated annealing. Comparison of simulation results with previous work in the literature shows that this combined method yields coding systems with good performance.

1 Introduction

A high-performance waveform coding technique is known as trellis, lookahead, or delayed decision source or waveform coding [1]. Trellis waveform coding uses a finite-state machine as the decoder. This machine is defined by an output map, corresponding to the codebook, and a next-state map, corresponding to the trellis structure, both of which being functions of the channel symbol and the current state. The extension of the next-state map or the state transition diagram in time is known as a trellis structure, a weighted directed graph consisting of identical stages. Each stage corresponds to a time instant. The encoder is matched to the decoder, it examines the trellis and finds the channel sequence that leads to minimum distortion, which is the sum of the distortion values between the input and reproduction symbols. This can be accomplished by a trellis search algorithm, such as the Viterbi Algorithm (VA). The encoder in a trellis waveform coding system is simply a trellis search algorithm matched to the decoder finite-state machine. Therefore, the design problem reduces to the design of the decoder finite-state machine. This problem has been addressed by several authors in the literature, see, e.g., [1], [2]. The design of the finite-state machine for a quantizer, using a trellis search, or in the context of finite-state vector quantization, without any search, has also been addressed in the literature, see, e.g., [3]. In this work, we optimize both the codewords and the finite-state machine structure of a scalar trellis waveform coder that uses the Viterbi algorithm, using a near-optimum approach. For the optimization of the decoder finite-state machine, we make the observation that since the decoder is equivalent to the trellis structure, for a given set of codewords, and a given input sequence, it is clear that finding the optimum decoder is equivalent to finding the trellis structure that will generate a channel sequence with minimum distortion at the decoder output. This is a combinatorial optimization problem and can be solved by known optimization methods. In this paper, we propose the *simulated annealing* algorithm [4] for this purpose.

2 The Design Method

In this work, the state space is chosen to be all the possible state transitions in a single stage of the trellis. We are interested in trellis waveform coders with rate 1 bit/sample. This imposes a constraint on the encoder structure: from each node, there are two outgoing branches which correspond to values of 0 and 1 for the binary channel code. We also constrain the number of input branches going into each node: there are two incoming branches. This constraint is imposed in order to obtain a more symmetric structure so that the search space is minimized and the possibility of pathological trellis structures is certainly eliminated. The move set has been chosen to be just the flipping of two branches, so that the output of a move is again in the state space. The cost function is simply the minimum metric calculated by VA. The initial value of the control parameter is calculated as suggested by Johnson et al. [5]. Geometric improvement is used as the cooling schedule. The length of Metropolis loops are determined experimentally. As the source, a first order Gauss-Markov source with autocorrelation coefficient 0.9 is used. This source is chosen since it is a common model of real data and it is widely used in comparing data compression systems. For the design of codewords, we used the generalized Lloyd algorithm (GLA) [1].

In this work, GLA and SA are run together. For a given codebook, the trellis structure is optimized using SA, and for this structure, the codebook is modified using GLA. The process is stopped when the system reaches an equilibrium, with respect to the SA criteria.

3 Results

Trellis waveform coding systems of different constraint lengths were trained using a first order Gauss-Markov source, and were coded using SA and GLA. For constraint lengths of $K = 2$–8, signal-to-quantization-noise ratios (SQNR) were computed. Then the system was tested using another first order Gauss-Markov source. In Table 1, the computed SQNR [dB] values are given (SA+GLA) together with the results of Stewart et al. (GLA) [1], and of Ayanoglu and Gray (PS) [6]. Results obtained using SA are better than those of [1]. This is expected since in [1] the trellis structure was fixed, not optimized. The results obtained via the predictive system [6] are better than the combined system, especially for low constraint lengths. Again, this is expected since the predictive system has a higher system complexity. However, our results are sufficiently close to those of [6] for intermediate constraint lengths, so that the nonpredictive system once again becomes attractive. Alternatively, SA can be incorporated into the predictive system design with possibly better performance.

References

[1] L. C. Stewart, R. M. Gray, Y. Linde, "The Design of Trellis Waveform Coders," *IEEE Trans. Comm.*, Vol. COM-30, pp. 702–711, April 1982.

[2] G. H. Freeman, J. W. Mark, I. F. Blake, "Trellis Source Codes Designed by Conjugate Gradient Optimization," *IEEE Trans. Comm.*, Vol. COM-36, pp. 1–12, January 1988.

[3] J. Foster, R. M. Gray, M. O. Dunham, "Finite-State Vector Quantization for Waveform Coding," *IEEE Trans. Info. Theo.*, Vol. IT-31, pp. 348–359, May 1985.

[4] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, Vol. 220, pp. 671–680, May 1983.

[5] D. S. Johnson, C. R. Aragon, L. A. McGeoch, C. Shevon, 'Optimization by Simulated Annealing: an Experimental Evaluation, Parts I and II,' *Operations Research*, Vol. 37, pp. 865–892, December 1989, and Vol. 39, pp. 378–406, June 1991.

[6] E. Ayanoglu, R. M. Gray, "The Design of Predictive Trellis Waveform Coders Using the Generalized Lloyd Algorithm," *IEEE Trans. Comm.*, Vol. COM-34, pp. 1073–1081, November 1986.

	SA+GLA		GLA		PS	
K	train	test	train	test	train	test
2	6.92	6.86	6.92	6.86	11.08	10.73
3	9.81	9.45	8.77	8.59	11.53	11.18
4	11.24	11.13	10.13	9.87	11.84	11.47
5	11.90	11.77	11.05	10.67	12.18	11 .83
6	12.00	11.90	11.56	11.09	12.38	11.96
7	12.29	11.98	11.8,	11.70	12.52	12.52
8	12.32	11.97	12.13	11.91	12.64	12.58

Table 1: SQNR [dB] values, SA+GLA: Simulated Annealing and Generalized Lloyd Algorithm, GLA: Generalized Lloyd Algorithm only, PS: Predictive System. K: Constraint Length.

Conference Author Index

A

Aazhang, B. 43, 206, 379
Abdel-Ghaffar, K. A. S. 125
Abrahams, J. 216
Adachi, F. 345
Agrell, E. 394
Ahlswede, R. 396
Alabbadi, M. 199
Alajaji, F. 426
Al-Bassam, S. 9
Albuquerque, A. A. 198
Alencar, F. M. R. 227
Alencar, M. S. 190, 440
Ali, I. 325
Alon, N. 433
Al-Rumaih, R. M. 257
Amit, Y. 185
Amrani, O. 61
Anderson, J. 268
Anderson, J. B. 19, 203, 271, 417
Ansari, A. 424
Antweiler, M. 411
Araki, K. 34
Ariel, M. 25
Arikan, E. 152
Arimoto, S. 218, 395
Ashley, J. J. 4
Atlas, L. E. 173
Aulin, T. 272
Axtell, M. L. 350
Ayanoglu, E. 219, 317, 443
Aygölü, U. 290

B

Balamesh, A. S. 435
Balram, N. 281
Baram, Y. 430
Baras, J. S. 340
Barg, A. 129
Barron, A. R. 51, 54
Baum, C. W. 109
Be'ery, Y. 29, 61
Bégin, G. 386
Belfiore, J. C. 342
Bell, M. R. 428
Belongie, M. 238
Belzile, J. 108, 267
Bender, P. E. 6
Benedetto, S. 413
Benyamin-Seeyar, A. 99
Berger, T. 151, 223
Betz, J. W. 16, 18
Bhargava, V. K. 97, 301
Biglieri, E. 287, 360
Biniashvili, A. 239
Bitzer, D. L. 270
Blake, I. F. 78, 96
Blakley, B. 229
Blakley, G. R. 229
Blaum, M. 125, 126, 294, 295
Bloemen, A. H. A. 300
Blostein, S. D. 89
Blum, R. S. 12
Bobrowski, R. 163
Boncelet, C. G., Jr. 118
Bose, B. 7, 9
Boullé, K. 342
Bours, P. A. H. 127
Bovik, A. C. 427
Boztas, S. 410
Brady, D. 48, 50
Brandt-Pearce, M. 379
Breen, M. A. 350
Bross, S. 78
Brown, C. P. 405
Brualdi, R. A. 366
Bruck, J. 126, 294, 295, 433
Burnashev, M. V. 15
Burr, A. G. 67, 284

C

Calderbank, A. R. 137, 141, 154, 183, 251
Cambanis, S. 315, 328
Camion, P. 146
Campbell, L. L. 310
Campello de Souza, R. M. 227
Capocelli, R. M. 7
Carlet, C. 305
Castelli, V. 355
Cenkl, M. 405
Cercas, F. A. B. 198
Chabanne, H. 398
Chan, A. H. 229
Chan, F. 414
Chan, W. K. 211
Chan, W.-Y. 335
Chandran, S. R. 318
Chang, C. S. 161
Chang, C.-S. 215
Chang, S. C. 79
Chao, C.-C. 346
CharnKeitKong, P. 138
Chayat, N. 259
Chen, B. 92
Chen, C.-C. 27
Chen, J. 282
Chen, P.-N. 11
Chen, X. 302
Chen, Z. 223
Cheng, R. S. 209, 260
Cherubini, G. 241, 253
Cheung, K.-M. 381
Chiu, M.-C. 346
Chou, P. A. 53
Cimadevilla, M. O. 94
Cioffi, J. M. 47
Clark, J. J. 331
Clarke, B. 54
Clarke, W. A. 299
Cochran, D. 330, 331
Coffey, J. T. 158, 303
Cohen, G. D. 150, 370
Cohen, J. E. 1
Cohn, D. 176
Collins, O. 20
Conte, E. 87
Cooper, A. B., III 82
Costello, D. J., Jr. 139, 144, 415
Courteau, B. 146
Cover, T. M. 311, 355
Csibi, S. 319
Csiszár, I. 73

D

Dabak, A. 207
da Costa e Silva, M. A. O. 63
Dale, M. 376
Dallal, Y. E. 168
da Rocha, V. C., Jr. 80
Dettmar, U. 382
Dholakia, A. 270
Di Bisceglie, M. 87, 90
Di Porto, A. 37
Divsalar, D. 381
Dixit, C. 224
Dodunekov, S. M. 306
Dolinar, S. 381
Domaszewicz, J. 437
Dorsch, B. G. 399
Drakul, S. L. 287
Drane, C. R. 157
Dumer, I. I. 31

E

Effros, M. 53
Ehrhard, D. 250
Elia, M. 360
Ephremides, A. 324
Ericson, T. 296
Etzion, T. 197

F

Fang, G. 39
Fang, Y. 389
Faragó, A. 431
Farrell, P. G. 32
Farvardin, N. 169, 392, 425
Feder, M. 52, 72, 74, 420, 439
Feng, G. L. 95, 304
Ferland, G. 383
Ferreira, H. C. 162, 299
Fessler, J. A. 131
Filip, P. 391
Fine, T. L. 351, 432
Finesso, L. 186
Fischer, T. R. 438, 441
Fishburn, P. C. 137, 141
Fitz, M. P. 164
Fitzpatrick, P. 98
Fonseka, J. P. 349
Forest, S. 108
Forney, G. D., Jr. 177
Franaszek, P. A. 3
Francos, J. M. 93
Frankl, P. 154
Freeman, G. H. 119
Fuja, T. 102, 122, 426
Fujiwara, E. 40, 244
Fujiwara, H. 26
Fujiwara, T. 68
Fukumasa, H. 404

G

Gabidulin, E. M. 412
Gagliardi, R. 376
Gagnon, F. 344
Games, R. A. 405
Gao, Y. 382
Garello, R. 413
Gehrman, C. 230
German, D. 133
Gersho, A. 170, 335
Gibson, J. D. 94
Giraud, X. 342
Gorman, J. D. 136
Gozzo, F. 203
Graham, R. L. 154
Greenberg, G. 197
Grenander, U. 185
Gu, J. 122
Gubner, J. A. 309
Guida, F. 37
Gulliver, T. A. 301
Günther, C. G. 70
Guo, N. 338
Györfi, L. 51, 55, 321

H

Haccoun, D. 108, 267, 414
Hagen, R. 171
Hajek, B. 320
Halpenny, L. 225
Hamada, M. 244
Hammons, R. 196
Han, T. S. 71, 153
Han, Y. S. 27
Hardin, R. H. 60
Harris, C. F. 279
Hartmann, C. R. P. 27

Hasan, M. A. 97
Hashimoto, T. 101, 385
Hassan, A. A. 111, 377
Hassner, M. 358
Hata, M. 297
Hauge, E. R. 361
Haussler, D. 54
Hedelin, P. 171
Heegard, C. 238, 283, 341, 401
Hekstra, A. P. 21
Helberg, A. S. J. 299
Helleseth, T. 361
Henkel, W. 285
Hero, A. O. 131, 187, 191
Hershey, J. E. 377
Herzberg, H. 62
Higgie, G. R. 336
Hirasawa, S. 397
Hirschfeld, J. W. P. 246
Hole, K. J. 242
Holubowicz, W. 163
Honary, B. 23
Honig, M. L. 49, 372
Honkala, I. S. 39, 197
Horowitz, J. 133
Hou, X.-d. 193
How, S. K. 416
Hu, I. 316
Huang, S.-C. 276
Huang, Y.-F. 276
Huber, K. 359
Hughes, B. 82, 210, 323
Hussain, Y. 392

I

Imai, H. 64, 138, 404
Immink, K. A. S. 2
Irie, H. 347
Itoh, S. 282

J

Ji, C. 434
Jia, M. 99
Jiang, S. 40
Jinushi, H. 36, 292
Johannesson, R. 268, 288
Johansson, T. 231
Johnson, D. H. 17, 207

K

Kabatianskii, G. A. 255
Kadota, T. T. 13
Kailath, T. 188
Kaleh, G. K. 201
Kallel, S. 100, 273
Kamabe, H. 8
Kamiya, N. 307
Kantorovitz, M. R. 393
Kao, Y.-H. 340
Kaplan, G. 105, 266
Kasami, T. 68
Kassam, S. A. 12
Kawabata, T. 112
Kennedy, G. T. 365
Kerpez, K. J. 261
Ketseoglou, T. J. 322
Khachatrian, L. H. 295
Khayrallah, A. S. 308
Kieffer, J. C. 277
Kiely, A. B. 158
Kim, D. 205
Kløve, T. 147
Knagenhjelm, P. 339
Kobayashi, H. 220
Kobayashi, K. 113, 406
Koch, M. 285
Kofman, Y. 160
Koga, H. 395
Kohno, R. 404
Kolodziejski, K. R. 16
Komo, J. J. 362
Koplowitz, J. 278
Koshelev, V. N. 212
Koski, T. 436
Kot, A. D. 291
Kötter, R. 33
Krauss, T. P. 88
Krouk, E. A. 255
Krzyzak, A. 353
Kschischang, F. R. 195
Kubo, J. 116
Kulkarni, S. 217
Kumar, P. V. 196, 298
Kurtas, E. 208
Kuruoglu, E. E. 443
Kuznetsov, A. V. 128

L

Lachaud, G. 247
Ladner, R. 173, 176
Lai, C. H. 100
Lam, W.-M. 217
Laţidoth, A. 143, 263, 266
Laroia, R. 169
Larsson, P. 38
Lazic, D. 264
Lee, C.-C. 169
Lee, J. S. 110
Leland, R. P. 314
Le-Ngoc, T. 99
Léo, A. M. P. 227
Letaief, K. B. 274
Leung, C. 291
Levenshtein, V. 245, 296
Levitin, L. B. 76
Levy, Y. 139
Li, K. 273
Li, W.-C. W. 154
Li, Y.-X. 236
Liesenfeld, B. 399
Likhanov, N. B. 320
Lin, M.-C. 419
Lin, S. 68, 184, 289, 318, 343
Lindell, G. 418
Linder, T. 65, 390
Linne von Berg, D. C. 380
Lipman, M. J. 216
Litsyn, S. 370
Liu, C.-C. 186
Liu, Y. 89
Livingston, J. N. 286
Loeliger, H.-A. 180, 182
Longo, M. 10, 90
Lops, M. 10, 87
Lorenzelli, F. 332
Lu, C.-C. 200
Lugosi, G. 356, 431
Lunn, T. J. 67
Lyons, D. F. 333

M

Ma, S.-C. 419
Madhow, U. 49, 372
Madhusudhana, H. S. 400
Mandayam, N. B. 43
Mandell, M. I. 252
Mao, R. 349
Maragos, P. 427
Marcellin, M. W. 5
Marcus, B. H. 4
Markarian, G. 23
Markman, I. 271
Marton, K. 189
Masry, E. 315
Massey, J. L. 229, 373
Massey, P. 181
Mathys, P. 181, 232, 257
Matsumoto, T. 345
Matsushima, T. K. 348
Mattson, H. F., Jr. 371
McEliece, R. J. 142, 252
McKellips, A. L. 310
McLane, P. J. 161
McLaughlin, S. W. 442
Meeuwissen, H. B. 300
Melas, C. M. 126
Merhav, N. 72, 265, 266, 352, 420
Middleton, D. 86
Miller, D. 172
Miller, L. E. 110
Miller, M. I. 134, 185
Mills, D. G. 144
Milstein, L. B. 202
Mittelholzer, T. 182
Mittenthal, L. 233
Miura, S. 307
Modestino, J. W. 107, 279, 422
Modiano, E. 324
Mondin, M. 413
Montolivo, E. 37
Montorsi, G. 413
Montpetit, A. 146
Morelos-Zaragoza, R. H. 184
Moreno, C. J. 145
Moreno, O. 145, 298, 358, 405
Mori, S. 116, 167
Morii, M. 34
Morita, H. 113, 406
Moulin, P. 135
Moura, J. M. F. 281
Muhammad, K. 274
Müller, F. 91, 280
Murad, A. H. 102

N

Nagata, O. 397
Nakagawa, K. 222
Narasimhan, A. 93
Narayan, P. 186
Nasiri-Kenari, M. 240
Nelson, G. 277
Nelson, L. B. 44
Neuhoff, D. L. 333, 435, 442
Nill, C. 22
Nilsson, J. E. M. 306
Nilsson, M. 364
Nishijima, T. 397
Nobel, A. B. 334
Noneaker, D. L. 106
Norton, G. H. 398
Noviskey, M. J. 350

O

Oda, H. 327
Offer, E. 19
Oguz, N. C. 317
Ogiwara, H. 347
Ohtsuki, T. 167
Oka, I. 26
Ölçer, S. 241, 243
Olshen, R. A. 226, 334
Onyszchuk, I. 142, 381
Orcutt, E. K. 5
O'Reilly, J. J. 124
Orlitsky, A. 155, 213
Orsak, G. C. 14, 206
Osthoff, H. 268, 269

P

Palazzo, R., Jr. 63
Páli, I. 55
Pan, J. 438
Panayirci, E. 290, 357
Papamarcou, A. 11
Paris, B.-P. 14
Parks, T. W. 88
Parsavand, D. 374
Paterson, K. G. 408
Pawlak, M. 356
Pei, P. 405
Peng, X.-H. 32
Penzhorn, W. T. 103
Pereira, J. M. N. 166
Perry, P. 123
Persson, J. 288
Phamdo, N. 425

Pless, V. 365, 366
Polemi, D. 358
Pollara, F. 381
Poltyrev, G. 62, 84, 159
Poor, H. V. 15, 44, 423
Popken, L. 165
Popplewell, A. 124
Poscetti, G. M. 37
Pottie, G. J. 251
Proakis, J. G. 16
Psaltis, D. 354, 434
Pursley, M. B. 106, 109

Q

Quatieri, T. F. 427

R

Rabinovich, A. 141
Raghavan, S. A. 28, 202
Rajan, B. S. 400
Rajpal, S. 289
Rao, P. S. 17
Rao, T. R. N. 304
Rapajic, P. B. 45
Rasmussen, L. K. 384
Redinbo, R. 387
Reed, I. S. 302
Reid, W. J., III 362
Ren, Q. 220
Rhee, D. J. 289
Rimoldi, B. 81, 85
Riskin, E. A. 173, 176
Rissanen, J. 52
Riza, N. A. 377
Roche, J. R. 214
Rose, K. 75, 172
Ross, T. D. 350
Rossin, E. J. 283
Roth, R. M. 4, 35
Ruf, M. J. 391
Rupf, M. 373
Ruprecht, J. 375
Rushanan, J. J. 405
Rushforth, C. K. 240
Ruszinkó, M. 367
Ryabko, B. Y. 57

S

Sadowsky, J. S. 313
Safavi-Naini, R. 234
Said, A. 417
Saints, K. 401
Sakaniwa, K. 36, 292
Sakata, S. 363
Salehi, M. 208
Samarasooriya, V. N. S. 421
Sasaki, G. 224
Sasase, I. 116, 167
Sato, Y. 327
Schaewe, T. J. 134
Schalkwijk, J. P. M. 300
Schapiro, B. 76
Schlegel, C. 24, 65
Schneider, W. R. 70
Schulz, T. J. 132
Sendrier, N. 402
Šenk, V. 264
Seshadri, N. 183
Shah, A. A. 326
Shamai, S. 105, 160, 168, 259, 262, 266
Shamoon, T. 341
Shen, S. 429
Shenoy, R. G. 88
Shepp, L. 130, 154
Sheppard, J. A. 284
Shi, J.-J. 83
Shibuya, T. 36
Shields, P. C. 115, 189
Shtarkov, Y. M. 56, 59
Shwedyk, E. 275
Siddiqi, M. U. 258, 400, 407
Siegel, P. H. 35
Simonis, J. 148
Sloane, N. J. A. 60
Smeets, B. 156
Smyth, C. J. 225
Snapp, R. R. 354
Snyder, D. L. 120
Snyders, J. 25, 84
Solé, P. 174, 368
Solomon, G. 192
Sorger, U. K. 30, 382
Stark, W. E. 111, 253
Steinberg, Y. 423
Steiner, M. 204, 403
Stevenson, T. J. 256
Stiller, C. 91, 280
Stokes, P. 368
Struik, R. 369
Stutzer, M. J. 312
Su, Y. 328
Sun, F.-W. 66
Sundberg, C.-E. W. 22
Sung, W. 303
Suzuki, H. 218
Suzuki, J. 58
Swanson, L. 381
Swarts, F. 162
Swaszek, P. F. 175

T

Takada, M. 34
Takata, T. 68
Takumi, I. 297
Tallini, L. 7
Tanaka, H. 348
Tassiulas, L. 221
Tempel, D. J. 275
Thelen, B. J. 136
Thomas, J. A. 3, 215
Tjalkens, T. 59, 121
Tombak, L. 234
Tomlinson, M. 198
Tong, L. 188
Trott, M. D. 177, 178
Truong, T. K. 302
Tsfasman, M. A. 249
Tsybakov, B. S. 320
Tu, C. 293, 409
Tufts, D. W. 326
Turmon, M. J. 432
Tzeng, K. K. 95

U

Udaya, P. 258, 407
Udén, J. 418
Ungerboeck, G. 243
Urbanke, R. 85
Uyematsu, T. 41

V

Vaishampayan, V. 104, 437
Vajda, I. 321
Valdez, C. 26
van der Meulen, E. C. 51, 55
van der Vleuten, R. J. 388
van Tilborg, H. C. A. 66, 126
van Trung, T. 228
Varanasi, M. K. 42, 46, 374
Vardy, A. 29, 61, 370
Varshney, P. K. 421
Vastola, K. S. 325
Vasudevan, S. 46
Venkatesh, S. S. 316, 354
Verdú, S. 71, 153, 209, 262
Veugen, T. 235
Vinck, A. J. H. 128, 162, 245, 299
Viswanathan, R. 424
Viterbi, A. J. 254
Viterbi, A. M. 254
Vladut, S. G. 248
Vouk, M. A. 270
Vucetic, B. S. 45

W

Wan, Z.-x. 179
Wang, F.-Q. 415
Wang, M. 441
Wang, R.-Y. 173
Wang, Y.-Y. 200
Watanabe, Y. 83
Weber, J. H. 125, 388
Wei, C. 330
Weinberger, M. J. 52
Wicker, S. B. 199, 384
Wigderson, A. 155
Willems, F. M. J. 59
Williamson, C. J. 358
Wilson, S. G. 380
Wilson, S. K. 47
Winick, K. A. 237
Wittke, P. H. 310
Wolf, J. K. 6, 202
Wolfmann, J. 149
Woods, J. W. 93
Wu, J. 343
Wu, J.-L. 114
Wu, X. 389
Wu, Y. W. 79
Wyner, A. D. 117

X

Xia, X.-G. 293, 329
Xu, G. 188
Xu, L. 353

Y

Yaghoobian, T. 96
Yamaguchi, K. 64, 138
Yamamoto, H. 69
Yamazato, T. 116
Yang, E.-h. 337
Yang, G.-c. 378
Yang, S.-H. 237
Yang, S.-M. 104
Yao, K. 205, 332
Yates, K. W. 256
Ye, Z. 429
Yeung, R. W. 77
Yoshida, K. 292
Ytrehus, O. 140, 242
Yu, C.-L. 114
Yuan, J.-L. 351
Yuille, A. 353

Z

Zamir, R. 74, 439
Zeger, K. 65, 390, 393
Zehavi, E. 160, 239
Zémor, G. 150, 370
Zhang, J. 92
Zhang, Z. 293, 298, 329, 409
Zigangirov, K. 269, 288
Ziv, J. 117, 352
Zvonar, Z. 48
Zyablov, V. V. 23