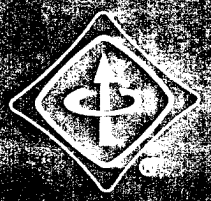
Proceedings

1995 IEEE

INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY

Whistler Conference Center
Whistler, British Columbia, Canada

September 17-22, 1995



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

Proceedings
1995 IEEE
International Symposium on
Information Theory

DTIC QUALITY INSPECTED 4

Whistler Conference Centre
Whistler, British Columbia, Canada
17-22 September, 1995

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

19960606 130

Proceedings 1995 IEEE International Symposium on Information Theory

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits 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, 222 Rosewood Drive, Danvers, MA 01923. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For other copying, 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 © 1995 by The Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Numbers: 95CH35738
95CB35738

Library of Congress Number: 72-179437

ISBN - Softbound: 0-7803-2453-6
Casebound: 0-7803-2454-4
Microfiche: 0-7803-2455-2

Additional copies of this Conference Record are available from

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

Chairmen

Vijay Bhargava Michael Pursley

Finance:	Norman Secord
Registration:	Aaron Gulliver
Publications:	Anwar Hasan
Publicity:	Amir Khandani
Local Arrangements:	Paul Ho (Chairman)
	Jim Ritcey
	Qiang Wang
International Advisory Committee:	Han Vinck (Chairman)

Acknowledgements

Grateful acknowledgement is given to the following organizations for their financial support of the 1995 IEEE ISIT:

Natural Sciences and Engineering Research Council of Canada
The National Science Foundation, USA
The U.S. Army Research Office
The U.S. Office of Naval Research

Program Committee

Ian F. Blake (Chairman)
John Anderson
Andrew Barron
Norman Beaulieu
Toby Berger
Richard Blahut
Rob Calderbank
Jim Cavers
Tony Ephremides
David Forney
David Haccoun
Brian Hughes
Hideki Imai
Saleem Kassam
Frank Kschischang
Vijay Kumar
Robert McEliece
Peter McLane
Larry Milstein
Vince Poor
Elvino Sousa
Stafford Tavares
Sergio Verdu
Victor Wei
Steve Wilson

Table of Contents

The Shannon Lecture
Performance and complexity--G. David Forney Jr. 1

Plenary Session Lectures
Symbolic dynamics and coding applications--Marcus, B. 2
Inequalities for source coding: Some are more equal than others--Ziv, J. 3
Quantum information theory--Brassard, G. 4
Wavelets: An overview, with recent applications--Daubechies, I. 5

Session MOAM1 Shannon Theory
Generalized projections for non-negative functions--Csiszár, I. 6
Capacity of channels with uncoded-message side-information--Shamai (Shitz), S., and Verdú, S. 7
Zero-error list capacities of discrete memoryless channels--Telatar, I.E. 8
Information efficiency in investment--Cover, T.M., and Erkip, E. 9
Sensitivity of Gaussian channel capacity and rate-distortion function to nonGaussian contamination--Pinsker, M.S., Prelov, V.V., and Verdú, S. 10
Determining the independence of random variables--Massey, J.L. 11
Information-theoretic bounds in authentication theory--Maurer, U.M. 12
The empirical distribution of good codes--Shamai (Shitz), S., and Verdú, S. 13

Session MOAM2 Universal Lossless Source Coding
Asymptotic behavior of the Lempel-Ziv parsing scheme and digital trees--Jacquet, P., and Szpankowski, W. 14
Empirical context allocation for multiple dictionary data compression--Franaszek, P., and Thomas, J. 15
Universal coding for arbitrarily varying sources--Feder, M., and Merhav, N. 16
A two-stage universal coding procedure using sufficient statistics--Matsushima, T., and Hirasawa, S. 17
Adaptive limitation of the dictionary size in LZW data compression--Ouaissa, K., Abdat, M., and Plume, P. 18
Universal coding of integers and unbounded search trees--Ahlswede, R., Han, T.S., and Kobayashi, K. 19
On the context tree maximizing algorithm--Volf, P.A.J., and Willems, F.M.J. 20
A fixed-to-variable variation of the Ziv-Lempel coding--Iwata, K., Uyematsu, T., and Okamoto, E. 21
Proposal of partially decodable Ziv-Lempel code--Iwata, K., Uyematsu, T., and Okamoto, E. 22

Session MOAM3 Code Division Multiple Access
Coded multicarrier code division multiple access--Rowitch, D.N., and Milstein, L.B. 23
The performance of voice and data communications in a mobile cellular CDMA system--Gass, J.H., Jr., Noneaker, D.L., and Pursley, M.B. 24
Successive cancellation in fading multipath CDMA channels--Varanasi, M.K. 25
Adaptive interference suppression for DS-CDMA with impulsive noise--Mandayam, N.B. 26
Error-and-erasure decoding of convolutional coded DS/SSMA communications in AWGN and Rayleigh fading channels--Kwon, J.M., and Kim, S.W. 27
Diversity performance of $\pi/4$-DQPSK for the reverse link in a DS-CDMA cellular system--Miller, L.E., and Lee, J.S. 28

On the expected value of the average interference parameter of code-sequences in CDMA systems--Schotten, H.D. 29
Synchronous frequency-hopped CDMA using wavelets--Daneshgaran, F., and Mondin, M. 30
Near-orthogonal coding for spread spectrum and error correction--Halford, K.W., Rozenbaum, Y., and Brandt-Pearce, M. 31

Session MOAM4 Turbo Codes

Unveiling turbo codes: some results on parallel concatenated codes--Benedetto, S., and Montorsi, G. 32
Unit-memory Hamming turbo codes--Cheng, J.-F., and McEliece, R.J. 33
Distance spectrum of the turbo-codes--Podemski, R., Holubowicz, W., Berrou, C., and Glavieux, A. 34
Low-rate turbo codes for deep-space communications--Divsalar, D., and Pollara, F. 35
'Turbo' coding for deep space applications--Andersen, J.D. 36
Interleaver design for three dimensional turbo codes--Barbulescu, A.S., and Pietrobon, S.S. 37
Weight distributions of turbo-codes--Svirid, Y.V. 38
Threshold decoding of turbo-codes--Svirid, Y.V., and Riedel, S. 39

Session MOAM5 Broadband Networks and Protocols

An efficient reservation connection control protocol for gigabit networks--Varvarigos, E.A., and Sharma, V. 40
Guaranteeing spatial coherence in real-time multicasting--Pokam, M.R., and Michel, G. 41
Peakedness of stochastic models for high-speed network traffic--Mark, B.L., Jagerman, D.L., and Ramamurthy, G. 42
Fault detection in communication protocols using signatures--Noubir, G., Vijayananda, K., and Raja, P. 43
Sporadic information sources--Loher, U. 44
An analysis approach for cell loss rate of shared buffer ATM switching--Zhao, Y.-B., Yu, J.-P., and Liu, Z.-J. 45
A scheme to adopt dynamic selection of error-correcting codes in hybrid ARQ protocol--Zhong, L., Xuemai, G., Qing, G., and Shilou, J. 46

Session MOAM6 Signal Processing and Coding

Importance sampling for TCM scheme on additive non-Gaussian noise channel--Sakai, T., and Ogiwara, H. 47
BRM sequence generators based on the field GF(2^n) for DSP implementations--Lee, S.-J., Goh, S.-C., Kim, K.-J., and Lee, D.-K. 48
Shift register synthesis for multiplicative inversion over GF(2^m)--Hasan, M.A. 49
On the probability of undetected error and the computational complexity to detect an error for iterated codes--Nishijima, T., and Hirasawa, S. 50
Wavefront decoding of trellis codes--Larsson, T. 51
Potential-decoding, error correction beyond the half minimum distance for linear block codes--Löhnert, R. 52
Information set decoding complexity for linear codes in bursty channels with side information--Sung, W., and Coffey, J.T. 53
When is hard decision decoding enough?--Swaszek, P.F. 54
First order approximation of the ordered binary symmetric channel--Fossorier, M.P.C., and Lin, S. 55
An asymptotic evaluation on the number of computation steps required for the nearest point search over a binary tree--Suzuki, H., and Arimoto, S. 56
New estimation of the probability of undetected error--Blinovsky, V. 57
Some remarks on efficient inversion in finite fields--Paar, C. 58

Session MOAM7 Coded Modulation

Multilevel coding with the 8-PSK signal set--Persson, J. 59

Soft-decision decoding for trellis coding and phase-difference modulation--Pursley, M.B., and Shea, J.M. 60

Coding and decoding of punctured QAM trellis codes--Chan, F., and Haccoun, D. 61

Power efficient rate design for multilevel codes with finite blocklength--Huber, J., and Wachsmann, U. 62

Trellis coding of Gaussian filtered MSK--Tyczka, P., and Holubowicz, W. 63

Bit error rate reduction of TCM systems using linear scramblers--Gray, P.K., and Rasmussen, L.K. 64

Multidimensional signaling for bandlimited channels--Daneshgaran, F., and Mondin, M. 65

Effect of encoder phase on sequential decoding of linear coded modulation--Balachandran, K., and Anderson, J.B. 66

Reduced complexity algorithms in multistage decoding of multilevel codes--Bobrowski, R., and Holubowicz, W. 67

A novel general approach to the optimal synthesis of trellis-codes for arbitrary noisy discrete memoryless channels--Baccarelli, E., Cusani, R., and Piazzo, L. 68

Session MOPM1 Information Theory and Applications

Identification via compressed data--Ahlswede, R., Yang, E.-H., and Zhang, Z. 69

Asymptotics of Fisher information under weak perturbation--Prelov, V.V., and van der Meulen, E.C. 70

A matrix form of the Brunn-Minkowski inequality--Zamir, R., and Feder, M. 71

The influence of the memory for a special permutation channel--Tamm, U. 72

The relation of description rate and investment growth rate--Erkip, E., and Cover, T.M. 73

Multi-way alternating minimization--Yeung, R.W., and Berger, T. 74

Generating non-Markov random sources with high Shannon entropy--D'yachkov, A.G., and Sidelnikov, V.M. 75

On the equivalence of some different definitions of capacity of the multiple-access collision channel with multiplicity feedback--Ruszinkó, M. 76

Matrix approach to the problem of matrix partitioning--Stasevich, S.I., and Koshelev, V.N. 77

Using the ideas of the information theory to study communication systems of social animals--Reznikova, Z., and Ryabko, B.Y. 78

Session MOPM2 Universal Lossy Source Coding

Fixed-slope universal algorithms for lossy source coding via lossless codeword length functions--Yang, E.-H., Zhang, Z., and Berger, T. 79

A lossy data compression based on an approximate pattern matching--Luczak, T., and Szpankowski, W. 80

The gold-washing algorithm (II): Optimality for ϕ-mixing sources--Zhang, Z., and Yang, E.-H. 81

Universal estimation of the optimal probability distributions for data compression of discrete memoryless sources with fidelity criterion--Koga, H., and Arimoto, S. 82

A universal data-base for data compression--Muramatsu, J,, and Kanaya, F. 83

Universal compression algorithms based on approximate string matching--Sadeh, I. 84

Session MOPM3 CDMA Sequence Families

Certain exponential sums over Galois rings and related constructions of families of sequences--Lahtonen, J. 85

Generalization of No sequences--No, J.-S. 86

Codes for optical transmission at different rates--Moreno, O., and Marić, S.V. 87

An upper bound for extended Kloosterman sums over Galois rings--Shanbhag, A.G., Kumar, P.V., and Helleseth, T. 88

Optimization of the ambiguity function of binary sequences and their mismatched filters for use in CTDMA systems--Schotten, H.D., and Ruprecht, J. 89

Optimal sequence sets meeting Welch's lower bound--Mow, W.H. 90

New signal design method by coded addition of sequences--Suehiro, N. 91

An upper bound for the aperiodic correlation of weighted-degree CDMA sequences--Shanbhag, A.G., Kumar, P.V., and Helleseth, T. 92

Construction of signal sets with constrained amplitude spectrum with upper bounds on cross-correlation--Chandran, G., and Jaffe, J.S. 93

Session MOPM4 Algebraic Geometry Codes

On Gröbner bases of the error-locator ideal of Hermitian codes--Chen, X., Reed, I.S., and Helleseth, T. 94

A new approach to determine a lower bound of generalized Hamming weights using an improved Bezout theorem--Feng, G.L., and Rao, T.R.N. 95

Fast erasure-and-error decoding of any one-point AG codes up to the Feng-Rao bound--Sakata, S. 96

The (64,32,27) Hermitian code and its application in fading channels--Chen, X., and Reed, I.S. 97

New construction of codes from algebraic curves--Shen, B.-Z., and Tzeng, K.K. 98

A fast parallel decoding algorithm for general one-point AG codes with a systolic array architecture--Kurihara, M., and Sakata, S. 99

Effective construction of self-dual geometric Goppa codes--Haché, G. 100

On codes containing Hermitian codes--Blahut, R.E. 101

Multilevel codes based on matrix completion--Dabiri, D., and Blake, I.F. 102

Session MOPM5 Modeling Analysis and Stability in Networks

Rate distortion functions and effective bandwidth of queueing processes--Chang, C.S., and Thomas, J.A. 103

Large bursts don't cause instability--Hajek, B. 104

The extension of optimality of threshold policies in queueing systems with two heterogeneous constant-rate servers--Traganitis, A., and Ephremides, A. 105

Elimination of bistability in spread-spectrum multiple-access networks--Murali, R., and Hughes, B.L. 106

Transient analysis of media access protocols by diffusion approximation--Ren, Q., and Kobayashi, H. 107

Multi-media multi-rate CDMA communications--Chang, Y.-W., Yao, S., and Geraniotis, E. 108

Stability analysis of quota allocation access protocols in ring networks with spatial reuse--Georgiadis, L., Szpankowski, W., and Tassiulas, L. 109

Capacity of digital cellular CDMA system with adaptive receiver--Oppermann, I., Vucetic, B.S., and Rapajic, P.B. 110

Session MOPM6 Signal Processing
Quadratic-inverse spectrum estimates for non-stationary processes--Thomson, D.J. 111
Parameter estimation and order determination in the low-rank linear statistical model--Tufts, D.W., Vaccaro, R.J., and Shah, A.A. 112
Blind identification in non-Gaussian impulsive environments--Ma, X., and Nikias, C.L. 113
Algorithms for blind identification of digital communication channels--Buisán Gómez del Moral, J., and Biglieri, E. 114
A class of iterative signal restoration algorithms--Noonan, J.P., Natarajan, P., and Achour, B. 115
Stochastic processes and linear combinations of periodic clock changes--Aakvaag, N.D., Duverdier, A., and Lacaze, B. 116
Near optimum filtering of quantized signals--Alencar, M.S., and Scharcanski, J. 117
Likelihood ratio partitions for distributed signal detection in correlated Gaussian noise--Chen, P.-N., and Papamarcou, A. 118
Rational moment mapping--Ferreira, H.C. 119
Real-time tracking of nonstationary signals using the Jacobi SVD algorithm--Lorenzelli, F., and Yao, K. 120
Improved LMS estimation via structural detection--Homer, J., Mareels, I., Bitmead, R., Wahlberg, B., and Gustafsson, F. 121

Session MOPM7 Trellis Structures annd Trellis Decoding
The trellis structure of maximal fixed-cost codes--Kschischang, F.R. 122
On trellis complexity of block codes: optimal sectionalizations--Lafourcade-Jumenbo, A., and Vardy, A. 123
On trellis complexity of block codes: lower bounds--Lafourcade-Jumenbo, A., and Vardy, A. 124
Trellis complexity versus the coding gain of lattices--Tarokh, V., and Blake, I.F. 125
Rotationally invariant, punctured trellis coding--Rossin, E.J., Rowe, D.J., and Heegard, C. 126
Trellises with parallel structure for block codes with constraint on maximum state space dimension--Moorthy, H.T., Lin, S., and Uehara, G. 127
Reconfigurable trellis decoding of linear block codes--Kot, A.D., and Leung, C. 128
On the twisted squaring construction, symmetric-reversible designs and trellis diagrams of block codes--Berger, Y., and Be'ery, Y. 129
Codes which satisfy the two-way chain condition and their state complexities--Encheva, S.B. 130
The trellis complexity of convolutional codes--McEliece, R.J., and Lin, W. 131

Session TUAM1 Channel Capacity
If binary codes existed that exceed Gilbert-Varshamov bound they could not reach the cutoff rate of BSC--Beth, T., and Lazic, D.E. 132
A simple proof that time-invariant convolutional codes attain capacity--Shulman, N., and Feder, M. 133
A construction of codes with exponential error bounds on arbitrary discrete memoryless channels--Uyematsu, T., and Okamoto, E. 134
Restricted two-way channel: bounds for achievable rates region for given error probability exponents--Haroutunian, E.A., Haroutunian, M.E., and Avetissian, A.E. 135
Lattice codes can achieve capacity on the AWGN channel--Urbanke, R., and Rimoldi, B. 136

Achieving symmetric capacity of a L-out-of-K Gaussian channel using single-user codes and successive decoding schemes--Cheng, R.S. 137
Capacity of a memoryless quantum communication channel--Fujiwara, A., and Nagaoka, H. 138
The effect of a randomly time-varying channel on mutual information--Medard, M., and Gallager, R.G. 139
Source coding and transmission of signals over time-varying channels with side information--Khansari, M., and Vetterli, M. 140

Session TUAM2 Coding for Recording Channels

One and two dimensional parallel partial response for parallel readout optical memories--Olson, B.H., and Esener, S.C. 141
Maximum likelihood decoding of block codes on (1-D) partial response channels--Markarian, G., and Honary, B. 142
Constrained block codes for class-IV PRML channels--Abdel-Ghaffar, K.A.S., and Weber, J.H. 143
On efficient high-order spectral-null codes--Tallini, L., Al-Bassam, S., and Bose, B. 144
Viterbi decoding considering insertion/deletion errors--Mori, T., and Imai, H. 145
Further results on cosets of convolutional codes with short maximum zero-run lengths--Hole, K.J., and Ytrehus, Ø. 146
A new coding technique: integer multiple mark modulation (IMMM)--Menyennett, C., and Ferreira, H.C. 147
A class of optimal fixed-byte error protection codes-(S+Fb)EC codes--Ritthongpitak, T., Fujiwara, E., and Kitakami, M. 148
A forbidden rate region for generalized cross constellations--Gelblum, E.A., and Calderbank, A.R. 149

Session TUAM3 Fading Channels I

Efficient multiuser communication in the presence of fading--Wornell, G.W. 150
Information theoretic limits on communication over multipath fading channels--Buz, R. 151
Computational cutoff rate of BDPSK signaling over correlated Rayleigh fading channels--Dam, W.C., Taylor, D.P., and Luo, Z.-Q. 152
On the construction of MPSK block codes for fading channels--Portugheis, J., and de Alencar, C.D. 153
Multilevel block coded 8-PSK modulation using unequal error protection codes for Rayleigh fading channels--Morelos-Zaragoza, R.H., Kasami, T., and Lin, S. 154
A change-detection approach to monitoring fading channel bandwidth--Blostein, S.D., and Liu, Y. 155
On the correlation and scattering functions of mobile uncorrelated scattering channels--Sadowsky, J.S., and Kafedziski, V.G. 156
High diversity lattices for fading channels--Boutros, J., and Viterbo, E. 157
Real-number DFT codes on a fading channel--Shiu, J., and Wu, J.-L. 158

Session TUAM4 Convolutional Codes

Good k/(k+1) time-varying convolutional encoders from time-invariant convolutional codes--Yamaguchi, K., and Imai, H. 159
Cascaded convolutional codes--Perez, L.C., and Costello, D.J., Jr. 160
An algorithm for identifying rate (n-1)/n catastrophic punctured convolutional encoders--Sun, F.-W., and Vinck, A.J.H. 161
Generalized Hamming weights of convolutional codes--Rosenthal, J., and Von York, E. 162

Upper bounds on the probability of the correct path loss for list decoding of fixed convolutional codes--Johannesson, R., and Zigangirov, K.S. 163
Minimal, minimal-basic, and locally invertible convolutional encoders--Dholakia, A., Bitzer, D.L., Koorapaty, H., and Vouk, M.A. 164
First order representations for convolutional encoders--Rosenthal, J., and Von York, E. 165
Some remarks on convolutional codes--Wan, Z.-X. 166
Improved union bound for Viterbi decoder of convolutional codes--Burnashev, M.V. 167

Session TUAM5 Neural Networks and Learning
Assessing generalization of feedforward neural networks--Turmon, M.J., and Fine, T.L. 168
Optimal stopping and effective machine complexity in learning--Wang, C., Venkatesh, S.S., and Judd, J.S. 169
On batch learning in a binary weight setting--Fang, S.C., and Venkatesh, S.S. 170
Pattern recognition via match between coded patterns and feature vectors--Lee, L.L., and Sosa, B.R.M. 171
Training recurrent networks using Hessian information--Coelho, P.H.G. 172
An artificial neural net Viterbi decoder--Wang, X.-A., and Wicker, S.B. 173
Combining neural network classification with fuzzy vector quantization and hidden Markov models for robust isolated word speech recognition--Xydeas, C.S., and Cong, L. 174
An EM-based algorithm for recurrent neural networks--Ma, S., and Ji, C. 175

Session TUAM6 Estimation
Sufficient conditions for norm convergence of the EM algorithm--Hero, A., and Fessler, J. 176
Deterministic EM algorithms with penalties--O'Sullivan, J.A., and Snyder, D.L. 177
Hidden Markov models estimation via the most informative stopping times for Viterbi algorithm--Kogan, J.A. 178
Model parameter estimation for 2D noncausal Gauss-Markov random fields--Cusani, R., Baccarelli, E., and Galli, S. 179
Achievable regions in the bias-variance plane for parametric estimation problems--Hero, A., and Usman, M. 180

Session TUAM7 Codes for the Gaussian Channel
New spherical designs in three and four dimensions--Hardin, R.H., and Sloane, N.J.A. 181
Computing the Voronoi cell of a lattice: The diamond-cutting algorithm--Viterbo, E., and Biglieri, E. 182
A new sphere packing in 20-dimensional Euclidean space--Vardy, A. 183
Asymptotically optimal spherical codes--Hamkins, J., and Zeger, K. 184
Applications of TCM with σ-tree constellations over the AWGN channel--Zaidan, M.Y., Barnes, C.F., and Wicker, S.B. 185
Generalized minimum distance decoding of Reed-Muller codes and Barnes-Wall lattices--Wang, C., Shen, B., and Tzeng, K.K. 186
Constellation shaping for the Gaussian channel--Heegard, C. 187
Random exploration of the three regular polytopes--Blachman, N.M. 188
Codes for the Lee metric and lattices for the l_1-distance--Siala, M., and Kaleh, G.K. 189

Session TUPM1 Rate Distortion Theory
On the redundancy of lossy source coding--Zhang, Z., Yang, E.-H., and Wei, V.K. 190
Mutual information and mean square error--Ihara, S. 191
Critical distortion of Potts model--Ye, Z., and Berger, T. 192
On the role of mismatch in rate distortion theory--Lapidoth, A. 193
On rate-distortion bounds of sub-Gaussian random vectors--Müller, F. 194
Rate distortion efficiency of subband coding with crossband prediction--Wong, P.W. 195
Operational rate distortion theory--Sadeh, I. 196
Distortion measures for variable rate coding--McClellan, S.A., and Gibson, J.D. 197
Multiple-access channels with correlated sources - coding subject to a fidelity criterion--Salehi, M. 198

Session TUPM2 Coding for Constrained Channels
Coding for channels with cost constraints--Khayrallah, A.S., and Neuhoff, D.L. 199
On the capacity of M-ary run-length-limited codes--McLaughlin, S.W., Luo, J., and Xie, Q. 200
Joint multilevel RLL and error correction coding--Siala, M., and Kaleh, G.K. 201
(4, 20) Runlength limited modulation code for high density storage system--Kim, M.-G., and Lee, J.H. 202
On properties of binary maxentropic DC-free runlength-limited sequences--Braun, V. 203
Coding for low complexity detection of multiple insertion/deletion errors--Clarke, W.A., and Ferreira, H.C. 204
Single-error-correcting codes for magnetic recording--Levitin, L.B., and Vainstein, F.S. 205
Single-track Gray codes--Hiltgen, A.P., Paterson, K.G., and Brandestini, M. 206
Optimizing the encoder/decoder structures in a discrete communication system--Khandani, A.K. 207

Session TUPM3 Fading Channels II
The effect of space diversity on coded modulation for the fading channel--Ventura-Traveset, J., Caire, G., and Biglieri, E. 208
Multilevel concatenated coded modulation schemes for the shadowed mobile satellite communication channel--Rhee, D.J., and Lin, S. 209
A hidden Markov model (HMM)-based MAP receiver for Nakagami fading channels--Kong, H., and Shwedyk, E. 210
Polynomial representation of burst error statistics--Kittel, L. 211
Coherent detection for transmission over severely time and frequency dispersive multipath fading channels--Bejjani, E., Belfiore, J.-C., and Leclair, P. 212
Optimised multistage coded modulation design for Rayleigh fading channels--Burr, A.G. 213
Distributed reception of fading signals in noise--Blum, R.S. 214
Performance of trellis coded direct-sequence spread-spectrum with noncoherent reception in a fading environment--Cheng, V.W., and Stark, W.E. 215
Reliable communication over the Rayleigh fading channel with I-Q TCM--Al-Semari, S.A., and Fuja, T.E. 216
A highly adaptive high-speed wireless transceiver--Pottie, G.J. 217

Session TUPM4 Decoding of Convolutional Codes
Adaptive forward error control shemes in channels with side information at the transmitter--Larrea-Arrieta, J., and Tait, D.J. 218

Novel scarce-state-transition syndrome-former error-trellis decoding of (n, n-1) convolutional codes--Lee, L.H.C., Tait, D.J., Farrell, P.G., and Leung, P.S.C. 219
Construction of trellis codes at high spectral efficiencies for use with sequential decoding--Wang, F.-Q., and Costello, D.J., Jr. 220
Real number convolutional code correction and reliability calculations in fault-tolerant systems--Redinbo, R. 221
Bidirectional Viterbi decoding algorithm with repeat request and estimation of unreliable region--Tajima, M. 222
Reduced complexity algebraic type Viterbi decoding of q-ary convolutional codes--Zigangirov, K.S. 223
On the general threshold decoding rule and related codes--Peng, X.-H., and Farrell, P.G. 224
Optimum distance profile trellis encoders for sequential decoding--Ljungberg, P. 225
Error burst detection with high-rate convolutional codes--Said, A. 226
A table-based reduced complexity sequential decoding algorithm--Koorapaty, H., Bitzer, D.L., Dholakia, A., and Vouk, M.A. 227

Session TUPM5 MDL and Learning

Characterization of the Bayes estimator and the MDL estimator for exponential families--Takeuchi, J.-I. 228
Concept learning using complexity regularization--Lugosi, G., and Zeger, K. 229
Minimax redundancy through accumulated estimation error--Yu, B. 230
An algorithm for designing a pattern classifier by using MDL criterion--Tsuchiya, H., Itoh, S., and Hashimoto, T. 231
An extension on learning Bayesian belief networks based on MDL principle--Suzuki, J. 232
Optimal universal learning and prediction of probabilistic concepts--Feder, M., Freund, Y., and Mansour, Y. 233

Session TUPM6 Combinatorics and Coding

Covering radius 1985-1994--Cohen, G.D., Litsyn, S.N., Lobstein, A., and Mattson, H.F., Jr. 234
Greedy generation of non-binary codes--Monroe, L., and Pless, V. 235
Lee distance Gray codes--Bose, B., and Broeg, B. 236
Diffuse difference triangle sets--Kløve, T., and Svirid, Y.V. 237
Two classes of binary optimum constant-weight codes--Fu, F.-W., and Shen, S.-Y. 238
Tensor codes for the rank metric--Roth, R.M. 239
On the asymptotic properties of a class of linearly expanded maximum distance separable codes--Ray-Chaudhuri, S. 240
Generalized partial spreads, geometric forms of bent functions--Carlet, C. 241
Constructing covering codes via noising--Charon, I., Hudry, O., and Lobstein, A. 242

Session TUPM7 Two Dimensional Channel Coding

On the diamond code construction--Baggen, C.P.M.J., and Tolhuizen, L.M.G.M. 243
A five-head, three-track, magnetic recording channel--Soljanin, E., and Georghiades, C.N. 244
Multi-track MSN codes for magnetic recording channels--Kurtas, E., Proakis, J.G., and Salehi, M. 245
MDS array codes with independent parity symbols--Blaum, M., Bruck, J., and Vardy, A. 246

On the capacity rates of two-dimensional runlength limited codes--Ye, Z., and Zhang, Z. 247
Physical limits on the storage capacity of magnetic recording media--Porter, D.G., and O'Sullivan, J.A. 248
Entropy estimates for simple random fields--Forchhammer, S., and Justesen, J. 249
Wavelets and lattice spaces--Moon, T.K. 250

Session WEAM1 Nonparametric Estimation and Classification
Nonparametric regression estimation for arbitrary random processes--Posner, S.E., and Kulkarni, S.R. 251
Model selection criteria and the orthogonal series method for function estimation--Moulin, P. 252
On nonparametric curve estimation with compressed data--Pawlak, M., and Stadtmüller, U. 253
Complexity regularization using data-dependent penalties--Lugosi, G., and Nobel, A. 254
On the existence of strongly consistent rules for estimation and classification--Kulkarni, S.R., and Zeitouni, O. 255
k nearest neighbors in search of a metric--Snapp, R.R., and Venkatesh, S.S. 256
An information-theoretic framework for optimization with application to supervised learning--Miller, D., Rao, A., Rose, K., and Gersho, A. 257
Nonparametric classification using radial basis function nets and empirical risk minimization--Krzyzak, A., Linder, T., and Lugosi, G. 258
Universal, nonlinear, mean-square prediction of Markov processes--Modha, D.S., and Masry, E. 259

Session WEAM2 New Directions in Data Compression
The quadratic Gaussian CEO problem--Viswanathan, H., and Berger, T. 260
Gaussian multiterminal source coding--Oohama, Y. 261
Multilevel diversity coding with symmetrical connectivity--Roche, J.R., Yeung, R.W., and Hau, K.P. 262
An error exponent for lossy source coding with side information at the decoder--Jayaraman, S., and Berger, T. 263
Error exponents for successive refinement by partitioning--Kanlis, A., and Narayan, P. 264
Remote coding of correlated sources with high resolution--Zamir, R., and Berger, T. 265
A sliding window Lempel-Ziv algorithm for differential layer encoding in progressive transmission--Subrahmanya, P., and Berger, T. 266
On the compression dimension of data strings and data sets--Kieffer, J., and Nelson, G. 267

Session WEAM3 Synchronization and Interference in Communication Systems
Optimal linear receivers for synchronizing pseudo random sequences--Dabak, A.G. 268
Maximum likelihood synchronization and frequency measurements--Collins, O.M., and Zhuang, Z. 269
Multilevel coding to combat quantization of the sum of the transmitted signal, a noise and a known interference--Herzberg, H., and Saltzberg, B.R. 270
Probability distribution of differential phase perturbed by tone interference and its application--Zeng, M., and Wang, Q. 271
Cyclotomic cosets and steady state solutions to a dynamic jamming game--Mallik, R.K. , and Scholtz, R.A. 272

Duration of a search for a fixed pattern in random data: the distribution function--Bajić, D., Drajić, D., and Katić, O. 273

Session WEAM4 Cyclic Codes

New codes with the same weight distributions as the Goethals codes and the Delsarte-Goethals codes--Helleseth, T., Kumar, P.V., and Shanbhag, A.G. 274

A cyclic [6,3,4] group code and the hexacode over GF(4)--Ran, M., and Snyders, J. 275

Decoding binary expansions of low-rate Reed-Solomon codes far beyond the BCH bound--Retter, C.T. 276

Multisequence generation and decoding of cyclic codes over Z_q--Interlando, J.C., and Palazzo, R., Jr. 277

On the maximality of BCH codes--Levy-dit-Vehel, F., and Litsyn, S. 278

Weights of long primitive binary BCH-codes are not binomially distributed--Lazic, D.E., Kalouti, H., and Beth, T. 279

On decoding doubly extended Reed-Solomon codes--Jensen, J.M. 280

The generalized Hamming weight of some BCH codes and related codes--Helleseth, T., and Winjum, E. 281

Cyclic codes and quadratic residue codes over Z_4--Pless, V., and Qian, Z. 282

Improved estimates for the minimum distance of weighted degree Z_4 trace codes--Helleseth, T., Kumar, P.V., Moreno, O., and Shanbhag, A.G. 283

Session WEAM5 Modeling and Analysis of Communication Systems

Representation of nonlinear OQPSK-type modulated waveforms as a sum of linear OQPSK-components--Gusmão, A., and Gonçalves, V. 284

Properties of guided scrambling encoders and their coded sequences--Fair, I.J., Bhargava, V.K., and Wang, Q. 285

Channel codes that exploit the residual correlation in CELP-encoded speech--Alajaji, F., Phamdo, N., and Fuja, T. 286

Modal analysis of linear nonbinary block codes used on stochastic finite state channels--Zepernick, H.-J. 287

Decoding procedure capacities for the Gilbert-Elliott channel--Bratt, G., Johannesson, R., and Zigangirov, K.S. 288

Real-time channel estimation using fuzzy logic--Arani, F., Smietana, R., and Honary, B. 289

Diversity waveform sets for high-resolution delay-doppler imaging--Guey, J.-C., and Bell, M.R. 290

Reduced complexity symbol-by-symbol demodulation--Fitz, M.P., and Gelfand, S.B. 291

The representation of multicomponent chirp signals using frequency-shear distribution--Yao, S., and He, Z.Y. 292

Session WEAM6 Signal Detection

A paradigm for distributed detection under communication constraints--Yu, C.-T., and Varshney, P.K. 293

Decentralized quickest change detection--Veeravalli, V. 294

Performance loss computation in distributed detection--Amirmehrabi, H., and Viswanathan, R. 295

HOS-based noise models for signal-detection optimization in non-Gaussian environments--Tesei, A., and Regazzoni, C.S. 296

Optimum detection of Gaussian signals in non-Gaussian noise--Buzzi, S., Conte, E., and Lops, M. 297

Asymptotically robust detection using statistical moments--Kolodziejski, K.R., and Betz, J.W. 298
Conditional testing in two-input detectors with single input conditioning--Seyfe, B., and Kahrizi, M. 299
Signal detection in continuous-time white Gaussian channel--Ihara, S., and Sakuma, Y. 300
A recursive formulation for quadratic detection on Rayleigh fading channels--Castoldi, P., and Raheli, R. 301
Robust detection of impulse signals in random impulse interferences--Grishin, Y.P., Sokolov, A.I., and Yurchenko, Y.S. 302

WEAM7 - Room EL4 Group Codes
Rotating group codes for the ISI channel--Massey, P., and Mathys, P. 303
On the trellis of convolutional codes over groups--Loeliger, H.-A. 304
Minimality tests for rational encoders over rings--Mittelholzer, T. 305
Permutation decoding of group codes--Biglieri, E. 306
On the algebraic fundamentals of convolutional encoders over groups--Arpasi, J.P., and Palazzo, R., Jr. 307
Useful groups for trellis codes--Sarvis, J.P., and Trott, M.D. 308
Codes from iterated maps--Andersson, H., and Loeliger, H.-A. 309
On binary-to-q-ary codes over groups--Wilhelmsson, L. 310
Abelian group codes, duality and MacWilliams identities--Ericson, T., and Zinoviev, V. 311

Session THAM1 Multiuser Systems
Detection of spread-spectrum signals for linear multi-user receivers--Mitra, U., and Poor, H.V. 312
MMSE interference suppression for joint acquisition and demodulation in CDMA systems--Madhow, U. 313
Orthogonally anchored blind interference suppression using the Sato cost criterion--Honig, M.L. 314
Optimal soft multi-user decoding for vector quantization in a synchronous CDMA system--Skoglund, M., and Ottosson, T. 315
Linear multiuser detectors for synchronous code-division multiple-access systems with continuous phase modulation--Papasakellariou, A., and Aazhang, B. 316
A near ideal whitening filter for M-algorithm detection in an asynchronous time-varying CDMA system--Wei, L., and Rasmussen, L.K. 317
A new projection receiver for coded synchronous multi-user CDMA systems--Schlegel, C., and Xiang, Z. 318
Adaptive multilevel coding associated with CCI cancellation for CDMA--Saifuddin, A., and Kohno, R. 319
Performance bounds for decorrelator detectors in a QS-CDMA system--Iltis, R.A. 320
A non-orthogonal synchronous DS-CDMA case, where successive cancellation and maximum-likelihood multiuser detectors are equivalent--Kempf, P. 321

Session THAM2 Entropy and Noiseless Source Coding
An inequality on guessing and its application to sequential decoding--Arikan, E. 322
Weighted coding methods for binary piecewise memoryless sources--Willems, F., and Casadei, F. 323
A D-ary Huffman code for a class of sources with countably infinite alphabets--Kato, A., Han, T.S., and Nagaoka, H. 324

Data expansion with Huffman codes--Cheng, J.-F., Dolinar, S., Effros, M., and McEliece, R. 325
Minimizing the maximum codeword cost--Abrahams, J. 326
Entropy reduction, ordering in sequence spaces, and semigroups of non-negative matrices--Hollmann, H.D.L., and Vanroose, P. 327
Variable-to-fixed length codes and the conservation of entropy--Savari, S.A. 328
An inequality on entropy--McEliece, R.J., and Yu, Z. 329
On entropies, divergences, and mean values--Basseville, M., and Cardoso, J.-F. 330

Session THAM3 Dispersive Channels and Equalization I
When are the MLSD respectively the matched filter receiver optimal with respect to the BER--Ödling, P., Eriksson, H.B., Koski, T., and Börjesson, P.O. 331
A genie-aided detector with a probabilistic description of the side information--Eriksson, H.B., Ödling, P., Koski, T., and Börjesson, P.O. 332
A new family of decision delay-constrained MAP decoders for data transmission over noisy channels with ISI and soft-decision demodulation--Baccarelli, E., Cusani, R., and Di Blasio, G. 333
MMSE-Optimal feedback and its applications--Tarköy, F. 334
A comparison of a single-carrier system using a DFE and a coded OFDM system in a broadcast Rayleigh-fading channel--Wilson, S.K., and Cioffi, J.M. 335
A decision feedback filter--Gelfand, S.B., and Fitz, M.P., 336
Combined decision feedback equalization and coding for high SNR channels--Yellin, D., Vardy, A., and Amrani, O. 337
Maximum likelihood sequence estimation for non-Gaussian band-limited channels--Cordier, M., and Geraniotis, E. 338
Incoherent diversity detection of fading signals in correlated non-Gaussian noise--Izzo, L., and Tanda, M. 339
Statistics of error recovery times of decision feedback equalizers--Choy, W.W., and Beaulieu, N.C. 340

Session THAM4 Structure of Block Codes
A lower bound on the undetected error probability of block codes--Abdel-Ghaffar, K.A.S. 341
The worst-case probability of undetected error for linear codes on the local binomial channel--Kløve, T. 342
Good error detection codes satisfy the expurgated bound--Hashimoto, T. 343
On the binomial approximation to the distance distribution of codes--Krasikov, I., and Litsyn, S. 344
Extensions of linear codes--Hill, R., and Lizak, P. 345
Tabu search in coding theory--Nurmela, K.J., and Östergard, P.R.J. 346
Reducing the complexity of trellises for block codes--Aguado-Bayón, L.E., and Farrell, P.G. 347
Canonical representation of quasi-cyclic codes--Esmaelli, M., Gulliver, T.A., and Secord, N.P. 348
Newton's identities for minimum codewords of a family of alternant codes--Augot, D. 349

Session THAM5 Cryptographic Techniques
Information theoretical lower bounds for unconditionally secure group authentication--Gehrmann, C. 350
Spectral properties and information leakage of multi-output Boolean functions--Youssef, A.M., and Tavares, S.E. 351
Cryptographic redundancy and mixing functions--Collins, O. 352
Design of the extended-DES cryptography--Oh, H.-S., and Han, S.-J. 353

Constructions of asymmetric authentication systems--Johansson, T. 354
Two simple schemes for access control--Chan, M.Y., and Yeung, R.W. 355
Attacks on Tanaka's non-interactive key sharing scheme--Park, S., Kim, Y., Lee, S., and Kim, K. 356
Simply implemented identity-based non-interactive key sharing--Tanaka, H. 357
The degeneration and linear structures of multi-valued logical functions--Feng, D.-G., Liu, B., and Xiao, G.-Z. 358
A fast identification scheme--Véron, P. 359

Session THAM6 Image Processing

Transmission of two-tone images over noisy channels with memory--Burlina, P., Alajaji, F., and Chellappa, R. 360
Multilevel resolution of digital binary images--Koplowitz, J., and DeLeone, J. 361
A new efficient coding method of a still image using three-dimensional DCT--Kondo, H., Agus, S., and Komori, S. 362
Nonlinear filters in joint source channel coding of images--Khayrallah, A.S. 363
The polynomial phase difference operator for parametric modeling of 2-D nonhomogeneous signals--Friedlander, B., and Francos, J.M. 364
Adaptive image restoration using discrete polynomial transforms--Neyt, X., and Acheroy, M. 365
Extreme elements and granulometries in the estimation problem--Villalobos, I.R.T. 366
A noise tolerant algorithm for the object recognition of warning system--Kang, D.-S. 367
Variable step search algorithm for motion estimation--Cai, Z.Q., and Tran, V.N. 368
A simple, general and mathematically tractable sense of depth--Saadat, A., and Fahimi, H. 369

Session THAM7 Quantization

On the cost of finite block length in quantizing unbounded memoryless sources--Linder, T., and Zeger, K. 370
Universal quantization of parametric sources has redundancy k/2 logn/n--Chou, P.A., Effros, M., and Gray, R.M. 371
On the encoding complexity of scalar quantizers--Hui, D., and Neuhoff, D.L. 372
Reduced-complexity waveform coding via alphabet partitioning--Said, A., and Pearlman, W.A. 373
Some results on quantization of a narrowband process--Bist, A. 374
A statistical analysis of adaptive quantization based on causal past--Yu, B. 375
Probability quantization for multiplication-free binary arithmetic coding--Cheung, K.-M. 376
Affine index assignments for binary lattice quantization with channel noise--Méhes, A., and Zeger, K. 377
Optimal quantization for finite state channels--Duman, T.M., and Salehi, M. 378
Performance of the adaptive quantizer--Pilipchouk, N.I. 379

Session THPM1 Multiuser Communications

Combined multipath and spatial resolution for multiuser detection: potentials and problems--Huang, H.C., Schwartz, S.C., and Verdú, S. 380
Multi-user communication with multiple symbol rates--Honig, M.L., and Roy, S. 381
Multisensor multiuser receivers for time-dispersive multipath fading channels--Stojanovic, M., and Zvonar, Z. 382
An iterative multiuser receiver: the consensus detector--Grant, A.J., and Schlegel, C. 383
Blind multiuser deconvolution in fading and dispersive channels--Fonollosa, J.R., Fonollosa, J.A.R., Zvonar, Z., and Vidal, J. 384

On the least possible decoding error probability for truly asynchronous single sequence hopping--Csibi, S. 385
On the performance of partial-response DS/SS systems in a specular multipath environment--Wang, Y.-P., and Stark, W.E. 386
Distributed access control in wireline and wireless systems--Hansen, C.J., and Pottie, G.J. 387

Session THPM2 Lossless Source Coding
A Bayes coding algorithm for FSM sources--Matsushima, T., and Hirasawa, S. 388
A CTW scheme for some FSM models--Suzuki, J. 389
On tree sources, finite state machines, and time reversal--Seroussi, G., and Weinberger, M. 390
Approximation of Bayes code for Markov sources--Takeuchi, J.-I., and Kawabata, T. 391
Markov random field models for natural language --Mark, K.E., Miller, M.I., and Grenander, U. 392
A multialphabet arithmetic coding with weighted history model--Hsieh, M.-H., and Wei. C.-H. 393
Bit-wise arithmetic coding for data compression--Kiely, A.B. 394
Fast enumerative source coding--Ryabko, B. 395

Session THPM3 Dispersive Channels and Equalization II
A new spectral shaping scheme without subcarriers--Yang, Y., and Welch, L.R. 396
On optimizing multicarrier transmission--Willink, T.J., and Wittke, P.H. 397
Multitone modulation and demodulation for channels with SNR uncertainty--Baum, C.W., and Conner, K.F. 398
Achievable rates for Tomlinson-Harashima precoding--Wesel, R.D., and Cioffi, J.M. 399
Phase-shifted linear partial-response modulation--Said, A., and Anderson, J.B. 400
Self-training adaptive equalization for multilevel partial-response transmission systems--Cherubini, G., Ölçer, S., and Ungerboeck, G. 401
On the existence and uniqueness of joint channel and data estimates--Chugg, K.M., and Polydoros, A. 402
Data detection of coded PSK in the presence of unknown carrier phase--Nassar, C.R., and Soleymani, M.R. 403
Delayed decision feedback equalization--Varanasi, M.K. 404
Tree search algorithms for self-adaptive maximum-likelihood sequence estimation--Paris, B.-P., and Shah, A.R. 405
Modified/quadrature partial response-trellis coded modulation (M/QPR-TCM) systems--Uçan, O.N. 406

Session THPM4 Decoding Block Codes
Soft decoding employing algebraic decoding beyond e_{BCH}--Nilsson, J.E.M. 407
The algebraic decoding of the Z_4-linear Goethals code--Helleseth, T., and Kumar, P.V. 408
The Welch-Berlekamp and Berlekamp-Massey algorithms--Blackburn, S.R. 409
The optimal erasing strategy for concatenated codes--Hsuan, Y., Coffey, J.T., and Collins, O.M. 410
Error and erasure decoding of binary cyclic code up to actual minimum distance--Lee, H., Tzeng, K.K., and Chen, C.J. 411
Enhanced decoding of interleaved error correcting codes--Blaum, M., and van Tilborg, H.C.A. 412
A decoding algorithm for linear codes over Z_4--Goel, M., and Rajan, B.S. 413
Decoding linear block codes using optimization techniques--Shih, C.-C., Wulff, C.R., Hartmann, C.R.P., and Mohan, C.K. 414

On maximum likelihood soft decision syndrome decoding--Fossorier, M.P.C., Lin, S., and Snyders, J. 415
Fast error magnitude evaluations for Reed-Solomon codes--Komo, J.J., and Joiner, L.L. 416
On neural decoding for some cyclic codes--Yu, J.-P., Zhao, Y.-B., and Wang, X.-M. 417

Session THPM5 Random Processes
Poisson approximation for excursions of adaptive algorithms--Zerai, A.A., and Bucklew, J.A. 418
Additive random sampling and exact reconstruction--Lacaze, B., and Duverdier, A. 419
Distortion measures via parametric filtering--Li, T.-H., and Gibson, J.D. 420
A two-step Markov point process--Hayat, M.M., and Gubner, J.A. 421
Noisy attractors of Markov maps--Salazar-Anaya, G., and Urías, J. 422
Revisiting the Huber-Strassen minimax theorem for capacities--Schwarte, H., and Sadowsky, J.S. 423
Hypothesis testing for arbitrarily varying source with exponential-type constraint--Fu, F.-W., and Shen, S.-Y. 424
A simple formula for the rate of maxima in the envelope of normal processes having unsymmetrical spectra--Abdi, A., and Nader-Esfahani, S. 425

Session THPM6 Wavelets
Measuring time-frequency information and complexity using the Rényi entropies--Baraniuk, R.G., Flandrin, P., and Michel, O. 426
Random fields and their wavelet transforms and representation: covariance and spectral properties--Masry, E. 427
Finite field wavelet transforms and multilevel error protection--Sarkar, S., and Poor, H.V. 428
Galois theory and wavelet transforms--Klappenecker, A., and Beth, T. 429
The discrete-time biorthogonal wavelet transform--Sola, M.Á., and Sallent, S. 430
Compression of square integrable functions: Fourier vs wavelets--Krichevskii, R.E., and Potapov, V.N. 431

Session THPM7 Trellises, Vectors and Quantization
Generalized vector quantization: Jointly optimal quantization and estimation--Rao, A., Miller, D., Rose, K., and Gersho, A. 432
Universal trellis coded quantization--Kasner, J.H., and Marcellin, M.W. 433
Maximum mutual information vector quantization--Wilcox, L.D., and Niles, L.T. 434
Robust quantization for transmission over noisy channels--Chen, Q., and Fischer, T.R. 435
Soft decoding for vector quantization in combination with block channel coding--Skoglund, M., and Hedelin, P. 436
A fixed-rate trellis source code for memoryless sources--Yang, L., and Fischer, T.R. 437
Why vector quantizers outperform scalar quantizers on stationary memoryless sources--Neuhoff, D.L. 438
Vector quantization of spherically invariant random processes--Müller, F., and Geischläger, F. 439
Optimal quantization for distributed estimation via a multiple access channel--Duman, T.M., and Salehi, M. 440
Vector quantization of images with the ECMPNN algorithm--Comaniciu, D. 441

Session FRAM1 Multiuser Information Theory
A new universal random coding bound for the multiple-access channel--Liu, Y.-S., and Hughes, B.L. 442
On the factor-of-two bound for Gaussian multiple access channels with feedback--Ordentlich, E. 443

On the capacity for the T-user M-frequency noiseless multiple access channel without intensity information--Vinck, A.J.H., Keuning, J., and Kim, S.W. 444
An extension of the achievable rate region of Schalkwijk's 1983 coding strategy for the binary multiplying channel--Meeuwissen, H.B., Schalkwijk, J.P.M., and Bloemen, A.H.A. 445
Interference cancellation in groups--Hanly, S.V., and Whiting, P.A. 446
Multiterminal estimation theory with binary symmetric source--Shimokawa, H., and Amari, S.-I. 447
Single user coding for the discrete memoryless multiple access channel--Grant, A., Rimoldi, B., Urbanke, R., and Whiting, P. 448
Rate-distortion theory for a triangular communication system--Yamamoto, H. 449
Extended Shannon's inequality and the asymptotic capacity of T-user binary adder channel--Mow, W.H. 450

Session FRAM2 Interactive Communications and Computation
Coding for computing--Orlitsky, A., and Roche, J.R. 451
Coding for interactive communication--Schulman, L.J. 452
Coding for distributed computation--Rajagopalan, S., and Schulman, L.J. 453
Random access from compressed datasets with perfect value hashing--Miller, J.W. 454
Information retrieval from databases--Leung, N.K.N., Coffey, J.T., and Sechrest, S. 455
Information theory and noisy computation--Evans, W.S., and Schulman, L.J. 456
Multiple repetition feedback coding for discrete memoryless channels--Veugen, T. 457

Session FRAM3 Sequences for Synchronization
Extremal polyphase sequences--Golomb, S.W. 458
A unified construction of perfect polyphase sequences--Mow, W.H. 459
Asymptotic autocorrelation of Golomb sequences--Gabidulin, E.M., Fan, P.Z., and Darnell, M. 460
Perfect sequences derived from M-sequences--Darnell, M., Fan, P.Z., and Jin, F. 461
Balanced quadriphase sequences with near-ideal autocorrelations--Fan, P.Z., and Darnell, M. 462
Quasi-linear synchronization codes--van Wijngaarden, A.J. 463
Extended sonar sequences--Moreno, O., Golomb, S.W., and Corrada, C.J. 464
Extended prefix synchronization codes--van Wijngaarden, A.J., and Morita, H. 465
Maximal prefix synchronized codes by means of enumerative coding--Morita, H., van Wijngaarden, A.J., and Vinck, A.J.H. 466
Partial classification of sequences with perfect auto-correlation and bent functions--Gabidulin, E.M. 467

Session FRAM4 Iterative Decoding Techniques
Codes and iterative decoding on general graphs--Wiberg, N., Loeliger, H.-A., and Kötter, R. 468
Concatenated coding system with iterated sequential inner decoding--Jensen, O.R., and Paaske, E. 469
The least stringent sufficient condition on the optimality of suboptimally decoded codewords--Kasami, T., Koumoto, T., Takata, T., Fujiwara, T., and Lin, S. 470
Implementation and performance of a serial MAP decoder for use in an iterative turbo decoder--Pietrobon, S.S. 471
On the convergence of the iterated decoding algorithm--Caire, G., Taricco, G., and Biglieri, E. 472
An iterative decoding scheme for serially concatenated convolutional codes--Siala, M., Papproth, E., Taieb, K.H., and Kaleh, G.K. 473

Soft-decision decoding of binary linear block codes based on an iterative search algorithm--Moorthy, H.T., Lin, S., and Kasami, T. 474
A soft output decoding algorithm for concatenated systems--Wang, X.-A., and Wicker, S.B. 475

Session FRAM5 Performance of Optical Communication Systems
On the impact of laser's relaxation oscillation on quadratically detected heterodyned lightwave signals--Dallal, Y.E., Jacobsen, G., and Shamai (Shitz), S. 476
On a new detection scheme of optical PPM signal--Yamazaki, K. 477
Tight lower bounds on capacity and cutoff rate of DOPPM in optical direct-detection channel--Ohtsuki, T., Sasase, I., and Mori, S. 478
Bit-error-rate of an optical two-users communications technique--Hassan, A.A., Molnar, K.J., and Imai, H. 479
BER of optical communication system using fiber source--Nguyen, L., Aazhang, B., and Young, J.F. 480
The channel capacity of hybrid fiber/coax (HFC) networks--Kerpez, K.J. 481
Error probability evaluation of optical systems disturbed by phase noise and additive noise--Einarsson, G., Strandberg, J., and Monroy, I.T. 482

Session FRAM6 Cryptographic Structures
Applications of coding and design theory to constructing the maximum resilient systems of functions--Levenshtein, V.I. 483
McEliece public key cryptosystems using algebraic-geometric codes--Janwa, H., and Moreno, O. 484
Binary trinomials divisible by a fixed primitive polynomial--Games, R.A., Key, E.L., and Rushanan, J.J. 485
New digital multisignature scheme in electronic contract systems--Kang, C.G. 486
The binary symmetric broadcast channel with confidential messages, with tampering--van Dijk, M. 487
Ideal perfect threshold schemes and MDS codes--Blakley, G.R., and Kabatianski, G.A. 488
Product codes and private-key encryption--Campello de Souza, J., and Campello de Souza, R.M. 489
On interval linear complexity of binary sequences--Balakirsky, V.B. 490

Session FRAM7 Special Code Constructions
The permutation group of affine-invariant codes--Berger, T.P., and Charpin, P. 491
Mixed-rate multiuser codes for the T-user binary adder channel--Cooper III, A.B., and Hughes, B.L. 492
New quaternary linear codes of dimension 5--Gulliver, T.A. 493
Hecke modules as linear block codes and block m-PSK modulation codes--Zimmermann, K.-H. 494
Reed-Solomon group codes--Zain, A.A., and Rajan, B.S. 495
A new construction of nonlinear unequal error protection codes--Chiu, M.-C., and Chao, C.-C. 496
Linear code construction for the 2-user binary adder channel--Cabral, H.A., and da Rocha, V.C., Jr. 497
Extending Reed-Solomon codes to modules--Desmedt, Y. 498
Quasi-cyclic Goppa codes--Bezzateev, S.V., and Shekhunova, N.A. 499
New ternary linear codes--Boukliev, I.G. 500

Performance and Complexity

G. David Forney, Jr.
Motorola, Inc.
20 Cabot Boulevard, Mansfield MA 02048 USA

Shannon showed that it was possible to achieve arbitrarily low error rates at any data rate less than channel capacity. By the early Sixties, it had been realized that the real problem was how to achieve reasonable error rates with acceptable decoding complexity at data rates anywhere near capacity. The author's research has been primarily motivated by this problem [1]-[22]. This lecture will offer an account of some of his adventures in this pursuit, and some preliminary conclusions.

References

[1] G. D. Forney, Jr., "On decoding BCH codes," IEEE Trans. Inform. Theory, vol. IT-11, pp. 549-557, 1965.

[2] G. D. Forney, Jr., "Generalized minimum distance decoding," IEEE Trans. Inform. Theory, vol. IT-12, pp. 125-131, 1966.

[3] G. D. Forney, Jr., Concatenated Codes, MIT Press, 1966.

[4] G. D. Forney, Jr., "Review of random tree codes," Appendix A of Final Report on Contract NAS2-3637, NASA CR73176, NASA Ames Res. Ctr., Calif., 1967.

[5] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," IEEE Trans. Inform. Theory, vol. IT-14, pp. 206-220, 1968.

[6] G. D. Forney, Jr., "Convolutional codes I: Algebraic structure," IEEE Trans. Inform. Theory, vol. IT-16, pp. 720-738, 1970.

[7] G. D. Forney, Jr., "Burst-correcting codes for the classic bursty channel," IEEE Trans. Commun. Technol., vol. COM-19, pp. 772-781, 1971.

[8] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," IEEE Trans. Inform. Theory, vol. IT-18, pp. 363-378, 1972.

[9] G. D. Forney, Jr., "The Viterbi algorithm," Proc. IEEE, vol. 61, pp. 268-278, 1973.

[10] G. D. Forney, Jr., "Structural analysis of convolutional codes via dual codes," IEEE Trans. Inform. Theory, vol. IT-19, pp. 512-518, 1973.

[11] G. D. Forney, Jr., "Convolutional codes II. Maximum-likelihood decoding" and "Convolutional codes III. Sequential decoding," Inform. Control, vol. 25, pp. 222-297, 1974.

[12] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff and S. U. Qureshi, "Efficient modulation for band-limited channels," IEEE J. Select. Areas Commun., vol. SAC-2, pp. 632-647, 1984.

[13] G. D. Forney, Jr., "Coset codes Part I: Introduction and geometrical classification" and "Coset codes Part II: Binary lattices and related codes," IEEE Trans. Inform. Theory, vol. IT-34, pp. 1123-1187, 1988.

[14] 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.

[15] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations Part I: Introduction, figures of merit, and generalized cross constellations," IEEE J. Select. Areas Commun., vol. SAC-7, pp. 877-892, 1989.

[16] G. D. Forney, Jr., "Multidimensional constellations Part II: Voronoi constellations," IEEE J. Select. Areas Commun., vol. SAC-7, pp. 941-958, 1989.

[17] G. D. Forney, Jr. and A. R. Calderbank, "Coset codes for partial response channels; or, coset codes with spectral nulls," IEEE Trans. Inform. Theory, vol. IT-35, pp. 925-943, 1989.

[18] G. D. Forney, Jr., "Geometrically uniform codes," IEEE Trans. Inform. Theory, vol. IT-36, pp. 1241-1260, 1991.

[19] G. D. Forney, Jr., "Trellis shaping," IEEE Trans. Inform. Theory, vol. 38, pp. 281-300, 1992.

[20] M. V. Eyuboglu and G. D. Forney, Jr., "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," IEEE Trans. Inform. Theory, vol. 38, pp. 301-314, 1992.

[21] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams, and canonical encoders," IEEE Trans. Inform. Theory, vol. 39, pp. 1491-1513, 1993.

[22] G. D. Forney, Jr., "Dimension-length profiles and trellis complexity of linear block codes" and "Density-length profiles and trellis complexity of lattices," IEEE Trans. Inform. Theory, vol. 40, pp. 1741-1772, 1994.

Symbolic Dynamics and Coding Applications

Brian Marcus
IBM Almaden Research Center, K53-802
650 Harry Rd.
San Jose, CA, 95120, USA
e-mail: marcus@almaden.ibm.com

The purpose of this talk is twofold: to give an elementary and concrete introduction to symbolic dynamics and to discuss two applications to coding problems.

We will begin with a brief discussion of the origins of symbolic dynamics going back to the work of Hadamard in 1898. The rough idea is that symbolic dynamics provides a model for the orbits of a classical dynamical system via a space of sequences. Next we will introduce the basic concepts of symbolic dynamics, emphasizing sliding block codes. We will survey some of the fundamental problems, solved and unsolved, in the subject. Then we will see how work on these problems has led to coding applications in two different settings:

1. The state splitting algorithm for constructing encoders/decoders adapted to input-constrained channels such as magnetic and optical recording channels.

2. An analysis of a class of spaces with homogeneity properties that naturally generalize convolutional codes, group codes [3], geometrically uniform codes [2], and orbit systems [6].

For introductory reading on symbolic dynamics and its applications, see the monograph [1], the textbook [4], and the article [6](§IV). For a tutorial on the state splitting algorithm see [5].

THE SPACE BELOW HAS BEEN RESERVED FOR YOUR DOODLING PLEASURE:

References

[1] M.-P. Béal, Codage Symbolique, Masson, 1993.

[2] G. D. Forney, Jr., "Geometrically uniform codes," IEEE Transactions on Information Theory, IT-37, 1991, 1241–1260.

[3] G. D. Forney, Jr. and M. Trott, "The dynamics of linear codes over groups: state spaces, trellis diagrams and canonical encoders," IEEE Transactions on Information Theory, IT-39, 1993, 1491–1513.

[4] D. Lind and B. Marcus,
An Introduction to Symbolic Dynamics and Coding,
Cambridge University Press, 1995.

[5] B. H. Marcus, P. H. Siegel and J. K. Wolf, "Finite-state modulation codes for data storage," IEEE Journal on Selected Areas in Communications, 10, 1992, 5–37.

[6] E. Rossin, N.T. Sindhushayana, and C. Heegard, "Trellis group codes for the Gaussian channel," IEEE Transactions on Information Theory, to appear (1995).

Inequalities for source coding: Some are more equal than others

Jacob Ziv[1]
Department of Electrical Engineering
Technion, Haifa 32000, Israel

Abstract — **An important class of universal encoders is the one where the encoder is fed with two inputs:**

a) The incoming string of data to be compressed.

b) A "training sequence" that consists of the last N data symbols that have been processed (i.e. a Sliding Window algorithm).

We consider Fixed-to-Variable universal encoders that noiselessly compress blocks of some fixed length and derive universal bounds on the rate of approach of the compression to the l-th order (per letter) entropy $H(X_1^\ell)$ or to the smaller conditional entropy $H(X_1^{\ell-k}|X_{-k+1}^0)$ as a function of ℓ and of the length N of the training sequence X_{-N+1}^0.

We describe non-asymptotic uniform bounds on the performance of data-compression algorithms in cases where the length N of the training sequence ("history") that is available to the encoder is not large enough so as to yield the ultimate compression, namely the entropy of the source. Two characteristic ultimate goals are considered: The l-th order entropy $H(X_1^\ell)$, and the associated conditional entropy $H(X_1^{\ell-k}|X_{-k+1}^0)$. The bounds are based on classical information-theoretic convexity arguments. Nevertheless, it is demonstrated that convexity arguments that work for one case are totally useless for the other and vice versa. Furthermore, these classical convexity arguments, when properly used, lead to efficient universal data compression algorithms for each of the two cases. For the sake of simplicity we confine our attention to binary stationary ergodic sources.

The first case to be considered is the one where we would like to find an upper bound on the length of a training sequence needed in order to guarantee that any source in the given class will yield a compression close to its l-th order entropy H_ℓ, and to derive a uniform bound on the rate of approach to this entropy as a function of ℓ and N.

"Intuition" tells us to use the "plug-in" method: namely, given a training sequence of length N, find the relative frequency $Q(X_1)^\ell$ of all l-vectors in it . Find the appropriate Huffman code and use it to encode the incoming l-blocks. The expected compression will be $-E\log Q(X_1^\ell)$. Clearly, by convexity, $-E\log Q(X_1^\ell) \geq \ell H(X_1^\ell)$ and eventually converges to it. Alas, the convergence is not uniform!

Let the training sequence be denoted by X_{-N+1}^0 and let: $N(X_{-N+1}^0|X_1^\ell)$=smallest $i \geq 1$ such that $X_{-i}^{\ell-i} = X_1^\ell$. If no such i can be found, $N(X_{-N+1}^0|X_1^\ell) = N$. It then follows from Kac's Lemma [1] that there exists a universal algorithm (a variant of the LZ algorithm) with a length function $L(X_1^\ell|X_{-N+1}^0)$ which is roughly equal to $logN(X_{-N+1}^0|X_1^\ell)$ when $N(X_{-N+1}^0|X_1^\ell) \neq N$ or to ℓ otherwise, such that $EL(X_1^\ell|X_{-N+1}^\ell) \leq \ell[H(X_1^\ell)+O(\log\ell/\ell)+\delta+2^{-\delta\ell}]$ where δ is some arbitrarily small positive number. This uniform bound holds if $N \geq 2^{(B+\delta)\ell}$ where B satisfies: $P[X_1^\ell : X_1^\ell \leq 2^{-B\ell}] \leq \delta$.

But why be satisfied with achieving $H(X_1^\ell)$ and not try to aim at some smaller conditional entropy where the conditioning is on some suffix of the training sequence X_{-N+1}^0?

Our second goal is to achieve a universal compression that is close to $H(X_1^{\ell-k}|X_{-k+1}^0)$ where $1 \leq k \leq \ell-1$. It is now assumed that a certain mixing condition is satisfied [2]. By Kac's lemma [1] and by convexity, $(\ell-k)H(X_1^{\ell-k}|X_{-k+1}^0) \geq E\log N(X_{-N+1}^0|X_{-k+1}^{\ell-k}) - kH(X_{-k+1}^0) \geq E\log N(X_{-N+1}^0|X_{-k+1}^{\ell-k}) + E\log\frac{n(X_{-N+1}^{-k}|X_{-k+1}^0)-O(1)}{N}$
$= E\log\frac{N(X_{-N+1}^0|X_{-k+1}^{\ell-k})[n(X_{-N+1}^{-k}|X_{-k+1}^0)-O(1)]}{N}$ where $n(X_{-N+1}^{-k}|X_{-k+1}^0)$ is the number of occurrences of an index i; i=k,k+1,...N such that $X_{-k+1-i}^{-i} = X_{-k+1}^0$ (i.e. a "plug-in" method!). Clearly, since X_{-k+1}^0 is a suffix of the training sequence it is available to both the encoder and the decoder prior to the processing of $X_1^{\ell-k}$.

Thus, the existence of a simple universal encoding algorithm that can uniformly approximate the lower bound on the conditional entropy that is derived above follows immediately.

A conditional version of the Kac's Lemma leads to yet another algorithm (a conditional LZ variant) that applies to *all* finite alphabet ergodic sources. [3]. It is demonstrated in [3] that in a sense, this algorithm is efficient in that no other universal data compression algorithm can do better, when the length of the training sequence is bounded by N (for large N).

[1] This work was supported by the Technion Fund for the Promotion of Research

References

[1] A. D. Wyner and J. Ziv, "The Sliding-Window Lempel-Ziv Algorithm is Asymptotically Optimal" (Invited paper), Proceeding of the IEEE, Vol. 82 , June 1994, pp. 872–877.

[2] A. D. Wyner and J. Ziv,"Classification with Finite-Memory", submitted to the IEEE Transactions on Information Theory.

[3] Y. Hershkovits and J. Ziv, "On Sliding-Window Universal Data Compression with Limited-Memory", submitted to the IEEE Transactions on Information Theory.

Quantum Information Theory

Gilles Brassard[1]

Département IRO, Université de Montréal
C. P. 6128, Succ. Centre-Ville, Montréal (Québec), Canada H3C 3J7

Abstract — **Quantum information theory is at the confluent of computer science and quantum mechanics. We survey some of the most striking recent developments in the field.**

I. Introduction: the Qubit

Classical and quantum information are very different. Classical information can be read, copied, and transcribed into any medium; it can be transmitted and broadcast, but it cannot travel faster than light. Quantum information cannot be read or copied without disturbance, but in some instances it appears to propagate instantaneously or even backward in time. Together the two kinds of information can perform feats that neither could achieve alone. For more details, references, and appropriate credit to the many researchers who made this work possible, please refer to my full paper in *Current Trends in Computer Science*, Jan van Leeuwen (Editor), Lecture Notes in Computer Science, Volume 1000 (special anniversary volume), Springer–Verlag, 1995.

Quantum information theory has the potential to bring about a spectacular revolution in computer science. Even though current-day computers use quantum-mechanical effects in their operation, for example through the use of transistors, they are still very much classical computing devices. A supercomputer is not fundamentally different from a purely mechanical computer that could be built around simple relays: their operation can be described purely in terms of classical physics and they can simulate one another in a straightforward manner, given sufficient storage. By contrast, computers could in principle be built to profit from genuine quantum phenomena that have no classical analogue, sometimes providing exponential speed-up compared to classical computers. Quantum information is also at the core of other phenomena that would be impossible to achieve in a purely classical world, such as unconditionally secure distribution of secret cryptographic material.

At the heart of it all is the quantum bit, or *qubit*. In classical information theory, a bit can take either value 0 or value 1. According to quantum information theory, a qubit can be in linear *superposition* of the two classical states, with complex coefficients. It is best visualized as a point on the surface of a unit sphere whose North and South poles correspond to the classical values. (This is not at all the same as taking a value *between* 0 and 1 as in classical analogue computing.) In general, qubits cannot be measured reliably: not more than one classical bit of information can be extracted from any given qubit and the more information you obtain about it, the more you disturb it irreversibly. As an example of how quantum information differs from classical information, it is possible in some situations to extract more than twice as much information from two *identical* qubits than from either one alone.

[1] Research supported in part by NSERC and FCAR.
Written while visiting the University of Wollongong, Australia.
Email: brassard@iro.umontreal.ca

II. Quantum Cryptography

The impossibility to measure quantum information reliably is at the core of quantum cryptography. When information is encoded with non-orthogonal quantum states, any attempt from an eavesdropper to access it necessarily entails a probability of spoiling it irreversibly, which can be detected by the legitimate users. This phenomenon can be exploited to implement a key distribution system that is provably secure even against an eavesdropper with unlimited computing power. Several prototypes have been built, including one that is fully operational over 30 kilometres of ordinary optical fibre. Further experiments are currently under way across the lake of Geneva. Quantum techniques may also assist in the achievement of subtler cryptographic goals, such as protecting private information while it is being used to reach public decisions.

III. Quantum Computing

Independent qubits are sufficient to produce nontrivial cryptographic phenomena, but they are not very interesting for computational purposes. For this, we must consider quantum *registers* composed of n qubits. Such registers can be in an arbitrary superposition of all 2^n classical states. In principle, a quantum computer can be programmed so that exponentially many computation paths are taken simultaneously in a single piece of hardware, a phenomenon known as *quantum parallelism*. What makes this so powerful—and mysterious—is the exploitation of constructive and destructive interference, which allows for the reinforcement of the probability of obtaining desired results while at the same time the probability of spurious results is reduced or even annihilated. The most famous example of quantum computation allows in principle for the quick factorization of large integers on a quantum computer, which has dramatic cryptographic significance.

IV. Quantum Teleportation

Even though quantum information cannot be measured in general, it can be teleported from one place to another. It is possible for two spatially separated qubits to be *entangled*, in the sense that each of them behaves randomly when measured, but they always give opposite results to the same measurement. Let Alice and Bob share such a pair. If Alice makes her mystery particle interact in the proper way with her share of the pair, Bob's share will instantaneously become a replica of the mystery particle up to rotation; at the same time Alice's mystery particle loses its information but she learns which rotation Bob must perform on his replica to match the original. Imperfect stores of nonlocal qubit pairs can be *purified* by local transformations and exchange of classical information.

Acknowledgements

I have benefited from discussions with too many people to attempt to list them all here, but one of them stands out: none of this would have been possible without the ubiquitous presence of Charles H. Bennett, our 12th–century troubadour.

Wavelets: An overview, with recent applications

Ingrid Daubechies

Department of Mathematics, Princeton University
Princeton, New Jersey, 08544 USA

Wavelets have emerged in the last decade as a synthesis from many disciplines, ranging from pure mathematics (where forerunners were used to study singular integral operators) to electrical engineering (quadrature mirror filters), borrowing in passing from quantum physics, from geophysics and from computer aided design.

The first part of the talk will present an overview of the ideas in wavelet theory, and show how it fits into the different disciplines in which it is rooted. The second part of the talk will discuss some recent applications, such as, in particular, a nonlinear "squeezing' of the wavelet transform, inspired by auditory models, with applications to speech processing; and a discussion of nonlinear approximation and why wavelets are so succesful in nonlinear approximation.

Generalized Projections for Non-negative Functions

I. Csiszár[1]

Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

Abstract — **The problem of minimizing a functional over a convex set of non-negative functions is considered, when the functional to be minimized is an f-entropy, or f-divergence resp. Bregman distance from a given function.**

I. Motivation

The motivation for this paper is the problem of inferring a function $p(x)$ on a set X when the only available information is $p \in E$, where E is a known convex set of functions on X. Possibly a prior guess q is also available, namely $p(x) = q(x)$ would be inferred were $q \in E$. A familiar method is to take that $p \in E$ which minimizes a certain functional.

II. Maximum-entropy type methods

For inferring non-negative functions, it is usual to minimize one of the functionals

$$J_f(p) = \int f(p)d\mu, \quad D_f(p,q) = \int qf(\frac{p}{q})d\mu, \tag{1}$$

$$B_f(p,q) = \int [f(p) - f(q) - f'(q)(p-q)]d\mu, \tag{2}$$

called f-entropy, f-divergence and Bregman distance, respectively. Here f is a strictly convex differentiable function on R_+ and μ is a σ-finite measure on X. $B_f(p,q)$ is a distance in the sense that it is non-negative and equals 0 iff $p = q \quad [\mu]$. $D_f(p,q)$ is also a distance if $f(1) = f'(1) = 0$.

The choice $f_1(t) = t\log t - t + 1$ gives the method of maximum entropy or ME ($J_{f_1}(p)$ for a probability density p is negative Shannon entropy, and $D_{f_1} = B_{f_1}$ is Kullback-Leibler I-divergence). Other familiar choices are $f_0(t) = -\log t + t + 1$, leading to Burg's method and to minimizing reversed I-divergence, and $f_\alpha(t) = [t^\alpha - \alpha t + \alpha - 1]\,\mathrm{sign}\,(\alpha - 1)$, $\alpha > 0$. There are strong arguments, both probabilistic and axiomatic, that support ME, cf. [1], [3]. For axiomatic justifications of alternative methods with some other f cf. [1], [4]. A probabilistic justification of these methods can be given by an extension of ME [2] in the case when f can be represented as the convex conjugate of the log of the moment generating function of a non-negative valued random variable. Among the functions f_α above, those with $0 \le \alpha \le 1$ have this property.

III. Main results

Theorem 1: Let E be a convex set of non-negative functions such that the infimum for $p \in E$ of $J_f(p)$, $D_f(p,q)$ or $B_f(p,q)$ is finite. Then each sequence $\{p_n\} \subset E$ approaching this infimum converges to a function p^* in the sense of convergence in measure on every set with finite μ-measure, providing in the case of D_f that either $q > 0 \quad [\mu]$ or

$$\lim_{t\to\infty} f'(t) = \infty\,. \tag{3}$$

Moreover, the difference of $J_f(p)$ resp. $B_f(p,q)$ from its infimum is lower bounded by $B_f(p,p^*)$, for every $p \in E$.

Notice that here p^* does not necessarily belong to E. The minimum of the considered functional over E is attained iff $p^* \in E$. If $p^* \notin E$, it is considered a generalized solution of the minimization problem or (in the case of D_f or B_f) a generalized projection of q onto E.

Theorem 2: The statement of Theorem 1 can be strengthened to convergence in $L_1(\mu)$ norm

(a) for J_f, if μ is a finite measure and (3) holds,

(b) for D_f, if $q \in L_1(\mu)$ and (3) holds,

(c) for B_f, if μ is a finite measure, $q \in L_1(\mu)$, and

$$\inf_{v\ge 1}(f'(Kv) - f'(v)) > 0 \quad \text{for some } K > 1\,. \tag{4}$$

Corollary: Under the conditions in Theorem 2, the $L_1(\mu)$ closedness of E is a sufficient condition for $p^* \in E$, i.e., for the existence of a (unique) solution of the minimization problem.

Remark: (4) is a stronger hypothesis than (3), but for the functions f_α either holds iff $\alpha \ge 1$. When (3) is not satisfied, no good sufficient conditions are available for $p^* \in E$.

In most applications, the feasible set E is defined by linear constraints,

$$E = \{p\colon \int a_\gamma(x)p(x)\mu(dx) = b_\gamma,\ \gamma \in \Gamma\}\,. \tag{5}$$

Then, by the above Corollary, under the hypotheses of Theorem 2 the boundedness of the functions a_γ is a sufficient condition for $p^* \in E$. For the functionals (1), a somewhat weaker sufficient condition is given in

Theorem 3: Under the hypotheses of Theorem 2 (a) or (b), the finiteness of $\int f^*(\lambda|a_\gamma|)d\mu$ or $\int f^*(\lambda|a_\gamma|)q\,d\mu$ for every $\lambda > 0$ and $\gamma \in \Gamma$ is sufficient for $p^* \in E$. Here f^* denotes the convex conjugate of f.

[1]This work was supported by the Hungarian National Foundation for Scientific Research, Grant 1906

References

[1] I. Csiszár, "Why least squares and maximum entropy?" *Ann. Statist.* vol. 19, pp. 2031-2066, 1991.

[2] D. Dacunha-Castelle, F. Gamboa, "Maximum d'entropie et problème des moments," Ann. Inst. H. Poincaré, vol. 4, pp. 567-596, 1990.

[3] E. T. Jaynes, Papers on Probability, Statistics, and Statistical Physics. R. D. Rosenkrantz, ed., Reidel, Dordrecht, 1981.

[4] L. Jones, G. Byrne, "General entropy criteria for inverse problems," *IEEE Trans. Inform. Th.* vol. 36. pp. 23-30, 1990.

Capacity of Channels with Uncoded-Message Side-Information

Shlomo Shamai (Shitz)[1] and Sergio Verdú[2]

[1]Department of Electrical Engineering, Technion-Israel Inst. of Technology, Haifa 32000, Israel sshlomo@ee.technion.ac.il
[2]Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A. Verdu@Princeton.edu

Abstract — **Parallel independent channels where no encoding is allowed for one of the channels are studied. The Slepian-Wolf theorem on source coding of correlated sources is used to show that any information source whose entropy rate is below the sum of the capacity of the coded channel and the input/output mutual information of the uncoded channel is transmissible with arbitrary reliability. The converse is also shown. Thus, coding of the side information channel is unnecessary when its mutual information is maximized by the source distribution. An information-theoretic interpretation of Parallel-Concatenated channel codes and, in particular, Turbo codes is put forth.**

I. Model

Consider the model depicted in the Figure 1 where two independent channels operate in parallel. If the inputs to both channels were allowed to be encoded, then Shannon's coding theorem tells us that the source is reliably transmissible provided its entropy rate is below the sum C_1+C_2 of the channel capacities; conversely, if the source entropy rate exceeds $C_1 + C_2$ then reliable transmission is not possible. The new twist in the model in Figure 1 is that the information going through channel 2 is not encoded. The following practical scenarios which fit into this model are studied in this paper: an existing uncoded communication link is to be upgraded with the addition of a coded channel in order to provide reliable transmission; the receiver obtains a noisy version of the raw data in addition to the coded channel output; a single channel time-multiplexed into several independent subchannels.

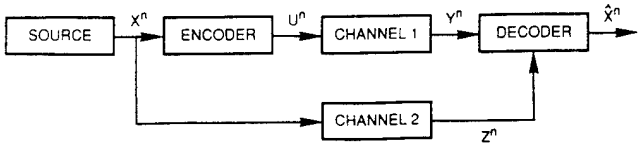


Fig. 1: Channel with Uncoded Side Information

II. Coding Theorem

Our main result states that *the source can be transmitted reliably provided that its conditional entropy rate given the output of the uncoded channel, $H(\mathbf{X}|\mathbf{Z})$, is below the capacity C_1 of channel 1, and, conversely, it cannot be transmitted reliably if the conditional entropy rate exceeds C_1.*

This result suggests that we view the information rate of the source as split into two nonoverlapping components, $H(\mathbf{X}) = H(\mathbf{X}|\mathbf{Z}) + I(\mathbf{X};\mathbf{Z})$. Even though the information quantified by the second term is transmitted uncoded, the source is reproducible with arbitrary reliability at the output. If, furthermore, the source is matched to the uncoded channel in the sense that it maximizes its input/output mutual information, then it is possible to transmit information at rate $C_1 + C_2$ even though no coding is provided for the information going through one of the channels. This implies that the sum of the capacities of K independent parallel binary symmetric channels can be achieved even if only one of them is encoded. This observation is most striking when the encoded BSC has very poor crossover probability.

Our coding theorem is proved under very mild conditions on the channels and the source. The source and the output of the uncoded channel are assumed to be jointly ergodic/stationary and the coded channel is assumed to be such that its capacity is equal to the limit of maximal mutual informations.

To prove the converse part of the result we show that even if the encoder were to observe the output of the uncoded channel, it would not be possible to send information at a faster rate. The proof of the achievability part is by construction of an encoder where the source coding and channel coding operations are performed separately. The source encoder does not operate at the full entropy rate of the source. Rather it is a *Slepian-Wolf encoder* [1] operating at rate $H(\mathbf{X}|\mathbf{Z})$. In the special case of binary-input memoryless channels, optimal encoding is possible by restricting attention to linear codes.

III. Parallel-Concatenated Codes

Parallel-Concatenated codes, and in particular Turbo codes [2], exhibit favorable complexity/performance tradeoffs. They can be cast within the model of this paper by considering a single-channel time-multiplexed into several independent subchannels. For example, one subchannel transmits the uncoded raw data (the Turbo codes are systematic), and two parallel channels are driven by *partial* encoders which can be viewed as joint source-channel encoders driven by a redundant source. A practically appealing way to ensure that the information encoded by the partial encoders is nonoverlapping is by prepending a sufficiently long interleaver at the input of one of the encoders. This setup is more attractive than simply multiplexing the source because of the complexity reductions of combined source/channel coding with high compression ratios. Good component codes in Parallel-Concatenated schemes are able to trade to some extent the traditional role of reducing the uncertainty of the source given the channel outputs for the easier goal of preserving mutual information.

Acknowledgements

This work was supported in part by a grant from the U.S.-Israel Binational Science Foundation

References

[1] D. Slepian and J. K. Wolf, Noiseless Coding of correlated information sources, *IEEE Trans. Information Theory*, IT-19, pp. 471-480, July 1973

[2] C. Berrou, A. Glavieux and P. Thitimajshima, Near Shannon Limit Error Correcting Coding and Decoding: Turbo-Codes, *Proc. International Conference on Communications*, Geneva, Switzerland, pp. 1064-1070, May 23-26, 1993

Zero-Error List Capacities of Discrete Memoryless Channels

İ. Emre Telatar
Room 2C-174, AT&T Bell Laboratories, Murray Hill, NJ 07974, USA
telatar@research.att.com

Abstract — **We define zero-error list capacities for discrete memoryless channels. We find lower bounds to, and a characterization of, these capacities. As is usual for such zero-error problems in information theory, the characterization is not generally a single-letter one. Nonetheless, we exhibit a class of channels for which a single letter characterization exists. We also show how the computational cutoff rate relates to the capacities we have defined.**

I. Introduction

It is sometimes desirable that the decoder of a communication system declare not just one, but several estimates of the transmitted data. For example, the encoder and the decoder may be the inner code of a more complex transmission system, the structure of the outer code can then be used to choose among the estimates the inner code provides. A decoder that may produce more than one estimate is called a *list decoder.*

Suppose we are given a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and transition probabilities $\{P(y|x), y \in \mathcal{Y}, x \in \mathcal{X}\}$.

Let $\mathcal{C}$ be a block code of length n for P. A *zero-error list decoder* for $\mathcal{C}$ is a decoder that assigns to every output $y \in \mathcal{Y}^n$ the set of codewords $\mathcal{L}(y,\mathcal{C}) \subset \mathcal{C}$ that could have produced that output with positive probability: $\mathcal{L}(y,\mathcal{C}) = \{c \in \mathcal{C} : P^n(y|c) > 0\}$. Let $L(y,\mathcal{C}) = |\mathcal{L}(y,\mathcal{C})|$ be the size of the list. The uniform distribution on $\mathcal{C}$ induces a distribution on L, and we will be interested in the moments of L.

For any $\rho > 0$ and P define the *zero-error ρ^{th}-moment list capacity* $C_{0\ell}(\rho, P)$ as the largest rate R such that for all $\epsilon > 0$ there exists a code of rate at least R for which the ρ^{th} moment of the list size is at most $1 + \epsilon$.

II. Summary of Results

To state our results, we introduce

$$F_0(\rho, P) = \max_{Q} \min_{\substack{V, W: V, W \ll P \\ WQ = VQ}} \rho I(Q,W) + D(V\|P|Q),$$

where Q ranges over the distributions on the input alphabet of P, $D(V\|P|Q)$ is the conditional informational divergence and $I(Q,W)$ is the mutual information. In the minimization V and W range over the set of channels with the same input and output alphabets as P, the notation $VQ = WQ$ means that the distribution on the output alphabet of the channels V and W are the same when their inputs have distribution Q, and $V \ll P$ means that $V(j|k) = 0$ whenever $P(j|k) = 0$.

Theorem 1 *For all $\rho > 0$, $C_{0\ell}(\rho, P) \geq F_0(\rho, P)/\rho$. Moreover, if we compute the lower bound for P^n, normalize, and pass to the limit, $C_{0\ell}(\rho, P) = \lim_{n\to\infty} n^{-1} F_0(\rho, P^n)/\rho$.*

The case of $\rho = 1$ is of particular interest; the corresponding capacity $C_{0\ell}(1, P)$ is called the *zero-error average list size capacity.* The substitution $\rho = 1$ in Theorem 1 recovers the results of [1].

Another special case is obtained by letting ρ become vanishingly small. The constraint on the ρ^{th} moment of the list size is then equivalent to demanding that $\Pr[L > 1]$ gets arbitrarily small. Taking the limit as $\rho \to 0$ in Theorem 1 we recover the previously known lower bound for *zero-undetected-error capacity* C_{0u} [1, 2, 3].

As a further special case, consider the limit $C_{0\ell}(\infty, P) \triangleq \lim_{\rho\to\infty} C_{0\ell}(\rho, P)$.

Theorem 2 $C_{0\ell}(\infty, P) = \min_{W: W \ll P} C(W)$, *where* $C(W) = \max_Q I(Q,W)$ *is the ordinary capacity of a discrete memoryless channel W.*

We have thus seen that $C_{0\ell}(\infty, P)$ has a single letter characterization. A more surprising result is that for a special class of DMCs one can obtain a single letter expression for $C_{0\ell}(\rho, P)$ for any $\rho > 0$:

Theorem 3 *Given a DMC P with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, construct the bipartite graph $G(P)$ with vertices $\mathcal{X} \cup \mathcal{Y}$ and edges $\{(x,y) : x \in \mathcal{X}, y \in \mathcal{Y}, P(y|x) > 0\}$. If $G(P)$ is acyclic then $C_{0\ell}(\rho, P) = E_0(\rho, P)/\rho$, where $E_0(\rho, P) = \max_Q -\ln \sum_y \left[\sum_x Q(x)P(y|x)^{1/(1+\rho)}\right]^{1+\rho}$.*

The quantity $E_0(\rho)/\rho$ is the largest rate for which the ρ^{th} moment of the number of computations per symbol remains bounded in sequential decoding [4, 5]. Theorem 3 is similar to the result of [6] where it is shown that for the same class of channels the zero-undetected-error capacity C_{0u} is identical to the ordinary capacity C.

References

[1] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list and detection zero-error capacities for low noise and a relation to identification," in *Proc. of 1994 IEEE Int. Symp. on Information Theory*, June 27 – July 1 1994.

[2] İ. E. Telatar and R. G. Gallager, "Zero error decision feedback capacity of discrete memoryless channels," in *Proc. of 1990 Bilkent Int. Conference on New Trends in Communication, Control, and Signal Processing*, (Bilkent University, Ankara, Turkey), July 1990.

[3] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," in *Proc. of 1994 IEEE Int. Symp. on Information Theory*, June 27 – July 1 1994.

[4] I. M. Jacobs and E. R. Berlekamp, "A lower bound to the distribution of computation for sequential decoding," *IEEE Trans. Information Theory*, vol. IT-13, pp. 167–174, April 1967.

[5] E. Arıkan, "An upper bound on the cutoff rate of sequential decoding," *IEEE Trans. Information Theory*, vol. IT-34, pp. 55–63, January 1988.

[6] M. S. Pinsker and A. Y. Sheverdyaev, "Transmission capacity with zero error and erasure," *Problems in Information Transmission (Problemy Peredachi Informatsii)*, vol. 6, no. 1, pp. 20–24, 1970.

Information efficiency in investment

Thomas M. Cover and Elza Erkip[1]
Stanford University, Information Systems Lab, Stanford, CA 94305-4055
cover@isl.stanford.edu, elza@isl.stanford.edu

Abstract — **We answer the question, what should we say about V when we want to gamble on X, and what is it worth? If $V = X$, we show that every bit of description at rate R is worth a bit of increase $\Delta(R)$ in the doubling rate. Thus the efficiency $\Delta(R)/R$ is equal to 1. For general V, we provide a single letter characterization for $\Delta(R)$. When applied specifically to jointly normal (V, X) with correlation ρ, we find the initial efficiency $\Delta'(0)$ is ρ^2. If V and X are Bernoulli random variables connected by a binary symmetric channel with parameter p, the initial efficiency is $(1-2p)^2$.**

We finally show how much increase in doubling rate is possible when the sender can provide R bits of information about V and side information S is available only to the investor.

Summary

Suppose we are interested in gambling on the outcome of a random variable X. The gambling consists of betting a proportion of wealth $b(x)$ on the outcome x. We would like to maximize the doubling rate, which is the growth rate of wealth when the gambler uses a fixed betting strategy on independent realizations of X. It is well known that Kelly gambling, which is to bet in proportion to the probability mass function of X, is optimal.

Now suppose we allow a description of X at rate R bits per symbol. Let $\Delta(R)$ be the maximum increase in the doubling rate of wealth for transmission rate of R. We prove that $\Delta(R) = R$. Any bit of information one sends about X is worth a bit of increase in the doubling rate.

We next consider the effectiveness of sending information when side information S is available to the investor but not to the encoder. The gambler combines this side information with the partial description of X to form the bet.

Theorem 1 *If X is described at rate R, and side information S is available to the gambler, then,*

$$\Delta(R) = R.$$

We ask what information should be given about a correlated random variable V if we want to help the investor gamble on X. This problem shows some similarities to source coding with side information [4, 1]. The encoder sends R bits about V and the investor uses this information to gamble on X. Here maximal efficiency is not generally possible.

Theorem 2 *When the encoder observes V correlated with X,*

$$\Delta(R) = \max_{p(\tilde{v}|v,x):\ I(\tilde{V};V)\le R,\ \tilde{V}\to V\to X} I(\tilde{V};X).$$

We establish certain properties of $\Delta(R)$ using entropy maximization results from Witsenhausen and Wyner [3].

Next, we find the increase in the doubling rate when the encoder sends information at rate R about a correlated random variable V with side information S present only at the investor. The investor uses these R bits together with the side information S to invest in the outcome of X.

Theorem 3 *When the encoder observes V, and side information S is available at the investor,*

$$\Delta(R) = \max_{p(\tilde{v}|v,x,s):\ I(V;\tilde{V}|S)\le R,\ \tilde{V}\to V\to(X,S)} I(\tilde{V};X|S)$$

Finally, we investigate the efficiency of descriptions based on correlated variables. If X and V are both Bernoulli$(\frac{1}{2})$ and are associated by a binary symmetric channel with crossover probability p, it can be shown that $\Delta(R)$ has a derivative of $(1-2p)^2$ at $R = 0$. Thus, even the most effective description of V relative to the investment in X pays off at the rate of only $(1-2p)^2$ bits of doubling per bit of description.

Now suppose that V and X are jointly Gaussian with correlation ρ. In this case the initial efficiency, $\Delta'(0)$, is equal to ρ^2.

The functional form of $\Delta(R)$ for binary and Gaussian random variables will be developed in [2]. Also, the relationship between the derivative of $\Delta(R)$ at $R = 0$ and the Renyi maximal correlation of V and X will be investigated.

References

[1] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Transactions on Information Theory*, vol. 21, pp. 629-637, 1975.

[2] E. Erkip and T. M. Cover, "The relation of investment rate and description growth rate," *Proceedings of 1995 IEEE International Symposium on Information Theory.*

[3] H. S. Witsenhausen and A. D. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Transactions on Information Theory*, vol. 21, pp. 493-501, 1975.

[4] A. A. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 21, pp. 294-300, 1975.

[1] This work was supported by NSF Grant NCR-9205663, ARPA Contract J-FBI-94-218 and JSEP Contract DAAH04-94-G-0058.

Sensitivity of Gaussian Channel Capacity and Rate-Distortion Function to nonGaussian Contamination

Mark S. Pinsker[1], Vyacheslav V. Prelov[1] and Sergio Verdú[2]

[1]Institute for Problems of Information Transmission, Ermolovoy str 19, Moscow 101447, Russia Prelov@ippi.ac.msk.su
[2]Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A. Verdu@Princeton.edu

Abstract — **In some applications, channel noise is the sum of a Gaussian noise and a relatively weak non-Gaussian contaminating noise. Although the capacity of such channels cannot be evaluated in general, we analyze the decrease in capacity, or sensitivity of the channel capacity to the weak contaminating noise. We show that for a very large class of contaminating noise processes, explicit expressions for the sensitivity of a discrete-time channel capacity do exist. Sensitivity is shown to depend on the contaminating process distribution only through its autocorrelation function and so it coincides with the sensitivity with respect to a Gaussian contaminating noise with the same autocorrelation function. A key result is a formula for the derivative of the water-filling capacity with respect to the contaminating noise power.**

Parallel results for the sensitivity of rate-distortion function relative to a mean-square-error criterion of almost Gaussian processes are obtained.

I. Sensitivity of Channel Capacity

Consider a discrete-time stationary channel:

$$Y_j = X_j + N_j + \theta Z_j \tag{1}$$

We assume that the random sequences $X = \{X_j\}$, $N = \{N_j\}$ and $Z = \{Z_j\}$ are second-order and mutually independent. The nominal noise N is Gaussian, $\mathbf{E}N_j = \mathbf{E}Z_j = 0$, $\mathbf{E}N_j^2 = \sigma^2$, $\mathbf{E}Z_j^2 = 1$. Denote by $C_P(\theta)$ the capacity of channel (1) under the assumption that the input power is constrained to some fixed constant P. The sensitivity of channel capacity with respect to the contaminating noise power is defined as

$$\mathbf{S}_P = \lim_{\theta \to 0} \frac{C_P(0) - C_P(\theta)}{\theta^2} \tag{2}$$

II. Gaussian Contamination

If the contaminating process $\{Z_i\}$ is Gaussian, then the capacity of (1) admits the well-known water-filling solution

$$C(\theta) = \frac{1}{2} \int_{-1/2}^{1/2} \ln \left(1 + \frac{[K_\theta - N_0(f) - \theta^2 Z(f)]^+}{N_0(f) + \theta^2 Z(f)} \right) df \,, \tag{3}$$

where $N_0(f)$ and $Z(f)$ are the power spectral densities of the nominal and contaminating noises, respectively, and the water level K_θ is adjusted so that the integral of the optimum input power spectral density $S_\theta(f)$ is equal to P, where $S_\theta(f)$ is the numerator in (3).

We show in this paper that the sensitivity of the water-filling channel capacity formula admits the following simple expression:

$$\mathbf{S}_P = \frac{1}{2K_0} \int_{-1/2}^{1/2} Z(f) \frac{S_0(f)}{N_0(f)} df \,, \tag{4}$$

where K_0 is the nominal water level. It follows that the sensitivity is maximized by a contaminating random process which concentrates its power at those frequencies where the nominal noise spectral density is minimum. Note that the worst-case sensitivity is minimized over the nominal noise spectral density by white noise, in which case the sensitivity is equal to

$$\mathbf{S} = \frac{P}{2\sigma^2} \, \frac{1}{P + \sigma^2} \,, \tag{5}$$

regardless of the power spectral density of the contaminating process.

III. NonGaussian Contamination

Since Gaussian noise minimizes capacity for a given power spectral density, the expression in (4) is an upper bound to sensitivity for nonGaussian contamination. Despite the lack of an expression for $C(\theta)$ in the nonGaussian case, this paper shows that

- The sensitivity is equal to (5) if the nominal Gaussian noise is white and the contaminating noise is regular (cf. [2]).
- The sensitivity is equal to (4) if both the nominal and contaminating noises are regular and if the ratio of spectral densities of contaminating to nominal noises: $Z(f)/N_0(f)$ is bounded on $\left[0, \frac{1}{2}\right]$.
- The sensitivity is equal to 0 if the nominal noise is regular and the contaminating noise is entropy-singular.

IV. Rate-Distortion Function

Consider the random process $N + \theta Z$ and denote by $R_D(\theta)$ its rate-distortion function relative to the mean-square-error criterion. We have shown (under the same conditions as above) that if $D \leq \sigma^2$, then the sensitivity of the rate-distortion function is

$$\lim_{\theta \to 0} \frac{R_D(\theta) - R_D(0)}{\theta^2} = \int_0^{1/2} \frac{Z(f)}{\max\{\lambda_0, N_0(f)\}} \, df \tag{6}$$

where λ_0 is defined by the equation

$$2 \int_0^{1/2} \min\{\lambda_0, N_0(f)\} \, df = D. \tag{7}$$

References

[1] Yu. A. Rozanov, *Stationary Random Processes*. Moscow: Fizmatgiz, 1963.

[2] I. A. Ibragimov and Yu. A. Rozanov, *Gaussian Random Processes*. Moscow: Nauka, 1970.

[3] M. S. Pinsker, V. V. Prelov, S. Verdú. "Sensitivity of Channel Capacity", to appear in *IEEE Thans. Inform. Theory*.

[4] M. S. Pinsker, V. V. Prelov, S. Verdú. "Asymptotics of the ϵ-Entropy of Stationary Almost Gaussian Processes." Proc. of IEEE Workshop, Moscow, 1994.

Determining the Independence of Random Variables

James L. Massey
Signal & Info. Proc. Lab., Swiss Federal Inst. Tech., CH-8092 Zurich, Switzerland

Abstract — **A graphical calculus is presented for determining the independence and conditional independence of random variables in a specified probabilistic setting. The calculus is developed first for the case of random variables that form a Markov chain. The calculus is then extended to the "general causal case" where the random variables are obtained from a sequence of random experiments in which each experiment can be carried out in full when the results of specified previous experiments are made available to it.**

I. Introduction

Because mutual information is essentially a measure of probabilistic dependence, information theory can be used to devise a convenient calculus for reasoning about probabilistic dependence. For example, because $I(X;Y) \geq 0$ with equality if and only if the random variables X and Y are independent, it follows that the determination of whether X and Y are independent is equivalent to determining whether $I(X;Y)$ vanishes. Moreover, the vanishing of $I(X;Y)$ can alternatively and conveniently be taken as the definition of (probabilistic) independence. Similarly, the vanishing of the conditional mutual information $I(X;Y \mid Z)$ can be taken as the definition of the independence of X and Y when conditioned on knowledge of Z.

Conditional independence will be seen to play an important role in the study of probabilistic dependence. Independence and conditional independence are in general unrelated properties of random variables in the sense that X and Y can be independent but dependent when conditioned on Z and, conversely, X and Y can be dependent but independent when conditioned on Z.

II. Markov Chains

A Markov chain can alternatively and conveniently be defined as a sequence $X_1, X_2, \ldots X_n$ of random variables such that, for all i strictly between 1 and n, $[X_1, X_2, \ldots X_{i-1}]$ and $[X_{i+1}, X_{1+2}, \ldots X_n]$ are independent when conditioned on X_i. An immediate consequence of the symmetry of mutual information, i.e., of the fact that $I(X;Y \mid Z) = I(Y;X \mid Z)$, is that the reversed sequence $X_n, X_{n-1}, \ldots X_1$ is also a Markov chain, which is a well-known fact but one that is awkward to prove from the usual definition of a Markov chain. Another immediate consequence of this alternative definition of a Markov chain is that any subsequence of a Markov chain $X_1, X_2, \ldots X_n$ is also a Markov chain, which again is a well known fact that is awkward to prove from the usual definition.

The following result is as useful in formulating a calculus of dependence as it is trivial to prove.

Proposition 1 *(Independence Inheritance)*
If $I(WX;Z \mid Y) = 0$, *then also* $I(X;Z \mid Y) = 0$ *and* $I(X;Z \mid WY) = 0$.

In other words, if some (possibly conditional) mutual information is zero, then any random variable *not in the conditioning* can be discarded or moved into the conditioning with the mutual information remaining zero.

The above proposition is the basis for the following calculus of independence for Markov chains: The random variables $X_1, X_2, \ldots X_n$ in the Markov chain are used to label in the natural order the nodes of a simple (undirected) linear graph with n nodes. Then any (possibly conditional) mutual information involving only the random variables $X_1, X_2, \ldots X_n$ is zero if, for every pair of random variables with one to the left and one to the right of the semicolon in the mutual information expression, there is a random variable in the conditioning whose node in the graph lies between the nodes for these two random variables. Moreover, this is the strongest statement that can be made in general about the (conditional) independence of the random variables in a Markov chain in the sense that there are chains for which the given mutual information is non-zero when this condition is not fulfilled. It is thus natural from the graphical viewpoint to think of conditioning as "blocking" dependence between the random variables in a Markov chain.

III. General Causal Systems

The graphical calculus of independence developed for Markov chains can be extended to apply to any random variables that can be described as the results of a sequence of random experiments in which the results of only previous experiments affect the results of following experiments, i. e., the random variables in the sequence have a well defined defined *causal dependence*. The distinction between causal dependence, which is directed, and probabilistic dependence, which is undirected, is crucial to the formulation of this extended graphical calculus. In contrast to the Markov chain case, conditioning can in general create probabilistic dependence between random variables that would be independent without this conditioning.

The real utility of the information-theoretical calculus for analyzing probabilistic dependence becomes evident when considering networks of information sources, channels, encoders and decoders. Precise definitions of all these devices together with the rules for their valid interconnection in neworks are required for the precise formulation of the calculus. Examples will be given in the presentation of this paper to illustrate the utility of the calculus in rather complicated networks.

Information-Theoretic Bounds in Authentication Theory[1]

Ueli M. Maurer
Department of Computer Science
ETH Zurich
CH-8092 Zurich, Switzerland

Abstract — **This paper gives a simplified treatment of, and new results on, information-theoretic lower bounds on an opponent's cheating probability in an authentication system with a given key entropy.**

I. Introduction

Authentication theory is concerned with providing evidence to the receiver of a message that it was sent by a specified legitimate sender, even in the presence of an opponent with unlimited computing power who can intercept and modify messages sent by the legitimate sender or send fraudulent messages to the receiver. Authenticity (like confidentiality) can be achieved by cryptographic coding when sender and receiver share a secret key.

Compared to Shannon's theory of secrecy, authentication theory is more subtle and involved. After some purely combinatorial results on authentication theory had been derived [1], Simmons [4] initiated a sequence of research activities on information-theoretic lower bounds in authentication theory (e.g., see [2], [3], [5], [6]).

II. Description of the Authentication Model

Consider a scenario in which a sender and a receiver share a secret key Z. The sender wants to send a sequence of messages $X_1, X_2, \ldots, X_n$, at some independent time instances, in an authenticated manner to the receiver. Each message X_i is authenticated separately by sending an encoded message Y_i which depends (possibly probabilistically) on Z, X_i, and possibly also on the previous messages $X_1, \ldots, X_{i-1}$. Based on Y_i, $Y_1, \ldots, Y_{i-1}$ and Z the receiver decides to either reject the message or accept it as authentic and, in case of acceptance, decodes Y_i to a message $\hat{X}_i$.

An opponent can use either of two different strategies for cheating. In an *impersonation* attack at time i, the opponent waits until he has seen the encoded messages $Y_1, \ldots, Y_{i-1}$ (which he lets pass to the receiver) and then sends a fraudulent message $\tilde{Y}_i$ which he hopes to be accepted by the receiver as the ith message. In a *substitution* attack at time i, the opponent lets pass messages $Y_1, \ldots, Y_{i-1}$, intercepts Y_i and replaces it by a different message $\tilde{Y}_i$ which he hopes to be accepted by the receiver and decoded to a message different from the one sent by the sender. There are three possible goals an opponent might persue in either of these two attacks:

- The receiver accepts Y_i as a valid message.
- The receiver accepts Y_i and decodes it to a message $\hat{X}_i$ *known* to the opponent. In other words, an opponent is only considered successful if he also guesses the receiver's decoded message $\hat{X}_i$ correctly.
- The receiver accepts Y_i and decodes it to a particular message $\hat{X}_i = x$ *chosen* by the opponent. Hence this type of attack depends on a particular value x.

[1]This work was supported by the Swiss National Science Foundation.

We will denote the maximal possible probabilities of success, for the three described scenarios, by $\hat{P}_I(i)$, $\tilde{P}_I(i)$ and $\hat{P}_I(i,x)$, respectively, for an impersonation attack at time i, and by $\hat{P}_S(i)$, $\tilde{P}_S(i)$ and $\hat{P}_S(i,x)$, respectively, for a substitution attack at time i.

III. Information-Theoretic Bounds

The literature on information-theoretic bounds in authentication theory is quite diverse because various different models are considered. Generally, the proofs are quite complicated and valid only for a restricted model while the results could actually be proven for a general model. For instance, some proofs only hold for deterministic encoding, for single (rather than a sequence of) messages, for a sequence of messages but with the restrictions that the encoding rule be the same for each message and that consecutive messages be distinct, or that the encoding rules do not depend on previous messages.

The goal of this paper is to derive various bounds in a coherent, more general setting, but by a simpler proof technique than those used before. In particular, we consider all three scenarios described above and our results could be generalized to a scenario where, for the sake of a smaller cheating probability, also a specified maximal probability of a decoding error for a correct message can be tolerated.

Some of the derived bounds are stated below. The first two bounds were also derived in [5] in a slightly less general form.

$$\begin{aligned}
\hat{P}_I(i) &\geq 2^{-I(Y_i;Z|Y_1\ldots Y_{i-1})} \\
\hat{P}_S(i) &\geq 2^{-H(Z|Y_1\ldots Y_i)} \\
\tilde{P}_I(i) &\geq 2^{-I(Y_i;Z|Y_1\ldots Y_{i-1}X_i)} \\
\tilde{P}_S(i) &\geq 2^{-H(Z|Y_1\ldots Y_iX_i)} \\
\hat{P}_I(i,x) &\geq 2^{-I(Y_i;Z|Y_1\ldots Y_{i-1},X_i=x)} \\
\hat{P}_S(i,x) &\geq 2^{-H(Z|Y_1\ldots Y_i,X_i=x)}
\end{aligned}$$

References

[1] E. N. Gilbert, F. J. MacWilliams, and N. J. A. Sloane, Codes which detect deception, *Bell Syst. Tech. J.*, Vol. 53, No. 3, 1974, pp. 405-424.

[2] J.L. Massey, Contemporary cryptology – an Introduction, in *Contemporary cryptology – the science of information integrity*, G.J. Simmons (Ed.), IEEE Press, 1992.

[3] U. Rosenbaum, A lower bound on authentication after having observed a sequence of messages, *J. of Cryptology*, Vol. 6, No. 3, 1993, pp. 135–156.

[4] G. J. Simmons, Authentication theory/coding theory, in *Advances in Cryptology – CRYPTO 84*, G.R. Blakley and D. Chaum (Eds.), Lecture Notes in Computer Science, No. 196. New York, NY: Springer, 1985, pp. 411–431.

[5] B. Smeets, Bounds on the Probability of Deception in Multiple Authentication, *IEEE Transactions of Information Theory*, Vol. 40, No. 5, 1994, pp. 1586-1591.

[6] M. Walker, Information-theoretic bounds for authentication schemes, *J. of Cryptology*, Vol 2, No. 3, 1990, pp. 131–143.

The Empirical Distribution of Good Codes

Shlomo Shamai (Shitz)[1] and Sergio Verdú[2]

[1]Department of Electrical Engineering, Technion-Israel Inst. of Technology, Haifa 32000, Israel sshlomo@ee.technion.ac.il
[2]Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, U.S.A. Verdu@Princeton.edu

Abstract — **Finding the input distribution that maximizes mutual information leads, not only to the capacity of the channel, but to engineering insights that tell the designer what good codes should be like. This is due to the folk theorem:** *The empirical distribution of any good code (i.e., approaching capacity with vanishing probability of error) maximizes mutual information.* **This paper formalizes and proves this statement.**

I. Introduction

The unique n-dimensional distribution that maximizes the n-block input-output mutual information of a binary symmetric channel (BSC) puts equal mass on all 2^n binary n-strings. Thus, common wisdom in information theory indicates that in order to approach the capacity of a BSC, a code must be such that the ensemble of its equiprobable codewords appears to be generated by a source of independent equally-likely bits. Formalizing and proving such a statement is not trivial as evidenced by the fact that the entropy rate of a source of pure bits is equal to 1 bit, whereas the entropy rate of the channel input induced by 2^{nR} equiprobable codewords is equal to R, and if the probability of error is to vanish, then $R \leq 1 - h(p) < 1$. Thus, convergence of the n-dimensional input distributions to a Bernoulli-1/2 source is ruled out. A good deal of the intuition on which the above common wisdom is grounded arises from the consideration of the input distributions of *random coding*, where not only do we average over equiprobable codewords, but over codebooks generated randomly according to the distribution maximizing mutual information. Then, the averaged input distributions of a random code are trivially equal to the capacity achieving input distributions. However, this trivial conclusion predicts nothing about the behavior of the input distributions of any particular code, which is the problem of interest.

It has been shown in [1] that for any finite-input channel that satisfies the strong converse, the *output* distribution induced by any good code sequence converges (in normalized divergence) to the (unique) output distribution induced by a capacity achieving input distribution. In certain cases (such as discrete memoryless channels with full-rank transition matrices [2]), such a result implies convergence of the input statistics. However, in general, such convergence does not follow directly from the convergence of output statistics.

II. Definitions

A. Empirical Distributions. For every codeword of a channel code we can find its first-order empirical distribution by computing the fraction of symbols in the codeword equal to each input letter. If for a given codebook we average the empirical distributions over equiprobable codewords we obtain the *first-order empirical distribution of the code.* Analogously, κ-th order empirical distributions can be defined by computing for each κ-string $\mathbf{v}$ the fraction of κ-strings within the codeword equal to $\mathbf{v}$. Averaging over equiprobable codewords results in the κ-th order empirical distribution of the code. Thus, for a code composed of M codewords of blocklength n, $\{z_{im},\ i = 1 \ldots n,\ m = 1, \ldots M\}$, the κ^{th}-order empirical distribution, $P^n_{\hat{X}(\kappa)}$, is defined as:

$$P^n_{\hat{X}(\kappa)} = \frac{1}{n-\kappa+1} \sum_{i=1}^{n-\kappa+1} P^n_{\hat{X}_i^{(\kappa)}}$$

where

$$P^n_{\hat{X}_i^{(\kappa)}}(a_1, \ldots, a_\kappa) = \frac{1}{M} \sum_{m=1}^{M} 1\{z_{im} = a_1\} \cdots 1\{z_{i+\kappa-1,m} = a_\kappa\}$$

B. Good Codes are channel codes whose rate is close to the channel capacity and whose decoding error probability vanishes with blocklength. More precisely, a *good code-sequence* for a channel with capacity C is a sequence of (n, M, λ_n) codes such that:

$$\lambda_n \longrightarrow 0\,,$$

$$\liminf_{n\to\infty} \frac{\log M}{n} = C\,.$$

III. Discrete Memoryless Channels

We have obtained results for a variety of channels, including channels with memory and continuous-alphabet channels. Our main result for discrete memoryless channels (DMC) is

Theorem 1 *Consider any good code sequence which does not use any symbol having zero mass under every input distribution that maximizes the single-letter mutual information. Then, the κ-order empirical distribution of such a code sequence satisfies:*

$$\lim_{n\to\infty} \min_{P_X \text{ s.t. } I(X;Y)=C} \mathbb{D}\left(P^n_{\hat{X}(\kappa)} \| P_X \times \cdots \times P_X\right) = 0\,.$$

where C is the channel capacity.

Note that the existence of a good code sequence satisfying the approximation property in Theorem 1 for any fixed κ is predicted by the optimality of constant-composition codes. But, in fact, this result holds for *any* good code sequence because of Theorem 1. A refinement of Theorem 1 entails letting κ grow with n. We have shown that any growth faster than $\log n$ destroys convergence.

Acknowledgements

This work was supported in part by a grant from the U.S.-Israel Binational Science Foundation. Fruitful discussions with Professor Amir Dembo are acknowledged.

References

[1] T. S. Han and S. Verdú, "Approximation Theory of Output Statistics," *IEEE Trans. Inform. Theory*, vol. 39, No. 3, pp. 752–772, May 1993.

[2] T. S. Han and S. Verdú, "Spectrum Invariancy under Output Approximation for Discrete Memoryless Channels with Full Rank," Problems of Information Transmission, pp. 101–118, April - June, 1993.

Asymptotic Behavior of the Lempel-Ziv Parsing Scheme and Digital Trees

Philippe Jacquet[1] and Wojciech Szpankowski[2]
INRIA, Rocquencourt, 78153 Le Chesnay Cedex, France
Dept. Computer Science, Purdue University, W. Lafayette, IN 47907, U.S.A.

Abstract — **In this work, for the memoryless source with unequal probabilities of symbols generation we derive the limiting distribution for number of phrases in the Lempel-Ziv parsing scheme. This proves a long standing open problem. In order to establish it we had to solve another open problem, namely, that of deriving the limiting distribution of the internal path length in a digital search tree.**

I. Introduction and Main Results

The primary motivation for this work is the desire to understand the asymptotic behavior of the fundamental parsing algorithm on words due to Lempel and Ziv [5]. It partitions a word into phrases (blocks) of variable sizes such that a new block is the shortest subword not seen in the past as a phrase. For example, the string 110010100010001000 is parsed into (1)(10)(0)(101)(00)(01)(000)(100).

We study the distribution of the number of phrases M_n constructed from a word of a fixed length n in a probabilistic framework. We assume that the word is generated by a probabilistic memoryless binary source. That is: *symbols are generated in an independent manner with "0" and "1" occurring respectively with probability* p *and* $q = 1 - p$. If $p = q = 0.5$, then we call it the *symmetric Bernoulli* model; otherwise we refer to the *asymmetric Bernoulli* model.

In order to study M_n, we reduce it to another problem on digital trees that is easier to handle. The reader is referred to [3] for a discussion and definition of digital trees. In short: the root of the tree is empty. All other phrases of the Lempel-Ziv parsing algorithm are stored in nodes. When a new phrase is created, the search starts at the root and proceeds down the tree, that is, symbol "0" in the input string means a move to the left and "1" means a move to the right. The search is complete when a branch is taken from an existing tree node to a new node that has not been visited before.

Observe that for fixed n the number of nodes in the associated digital tree is random and equal to M_n. We also consider a digital tree in which the number of nodes is fixed and equal to m, and we call such a model the *digital tree model.* For fixed m, we denote by $D_m(i)$ the length of the path from the root to the ith node (the ith depth). Then, the internal path length L_m becomes $L_m = \sum_{i=1}^{m} D_m(i)$.

In view of the above definitions, we note that M_n satisfies the following renewal equation $M_n = \max\{m : L_m = \sum_{k=1}^{m} D_m(i) \le n\}$, which directly implies that $\Pr\{M_n > m\} = \Pr\{L_m \le n\}$. Thus one can analyze M_n through L_m due to the following result of Billingsley [2]: If

$$\frac{L_m - \mu_m}{\sigma_m} \to N(0,1)\ , \tag{1}$$

then

$$\frac{M_n - n/(\mu_n/n)}{\sigma_n(\mu_n/n)^{-3/2}} \to N(0,1) \tag{2}$$

where $N(0,1)$ is the standard normal distribution, and μ_m and σ_m are positive constants.

Let $L_m(u) = Eu^{L_m}$ and $L(z,u) = \sum_{m=0}^{\infty} L_m(u)z^m/m!$ be generating functions of L_m and $L_m(u)$, respectively. We can show that $L(u,z)$ satisfies the following differential-functional equation for a memoryless source

$$\frac{\partial L(z,u)}{\partial z} = L(pzu,u)L(qzu,u) \tag{3}$$

with $L(z,0) = 1$.

Using the above differential-functional equation and (2), we prove the following theorem that directly extends the Aldous and Shields [1] results who established the limiting distribution of M_n only for the symmetric Bernoulli model.

Theorem . (i) *For a memoryless source the following weak convergence result holds*

$$\frac{M_n - EM_n}{\sqrt{varM_n}} \to N(0,1) \tag{4}$$

with $EM_n \sim \frac{nh}{\log n}$ *and* $varM_n \sim \frac{c_2 h^3 n}{\log^2 n}$ *where* $c_2 = (H-h^2)/h^3$ *with* $h = -p\log p - q\log q$ *being the entropy of the alphabet and* $H = p\log^2 p + q\log^2 q$. *Moreover, moments of* M_n *converge to the appropriate moments of the normal distribution. Finally,*

$$\Pr\{|M_n - EM_n| > \varepsilon EM_n\} \le A\exp\left(-a\varepsilon\sqrt{n}\right) \tag{5}$$

for some constants $A > 0$ *and* $\varepsilon > 0$.

Theorem above has plenty of applications in data compression (e.g., rate of convergence, etc). For example, using it we established in [4] the limiting distribution of the phrase length. Furthermore, using the large deviation result (5), we can obtain information about Lemepl-Ziv code redundancy, R_n. That is, $\Pr\{R_n > \varepsilon\} = \Pr\{M_n(\log M_n + 1) > n(h+\varepsilon)\} \le A\exp\left(-a\varepsilon\sqrt{n}\right)$ for $\varepsilon \ll h$.

[1] Supported by the Esprit Basic Research Action No. 7141.
[2] Supported by NSF Grants NCR-9206315 and CCR-9201078.

References

[1] D. Aldous and P. Shields, A Diffusion Limit for a Class of Random-Growing Binary Trees, *Probab. Th. Rel. Fields*, 79, 509-542 (1988).

[2] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York 1968.

[3] D. Knuth, *The Art of Computer Programming. Sorting and Searching,* Addison-Wesley (1973).

[4] G. Louchard and W. Szpankowski, Average Profile and Limiting Distribution for a Phrase Size in the Lempel-Ziv Parsing Algorithm, *IEEE Trans. Information Theory*, 41, 478-488 (1995).

[5] J. Ziv and A. Lempel, Compression of Individual Sequences via Variable-Rate Coding, *IEEE Trans. Information Theory*, 24, 530-536 (1978).

Empirical Context Allocation for Multiple Dictionary Data Compression

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

Abstract — **A class of multiple dictionary Lempel-Ziv algorithms is described, where a set of context dependent dictionaries are maintained, and a dictionary chosen based on empirical performance data. These algorithms are conceptually simpler than an earlier approach based on dynamic programming[1] and are also asymptotically optimal.**

It is well known[3] that the context of a symbol (the preceding few symbols) can be used to improve compression or prediction of the symbol. For example, the context algorithm[3, 4] chooses the estimated best context for compression via arithmetic coding. However, the most popular techniques are based on Lempel-Ziv coding. In LZ78, a tree structured dictionary is constructed using the source sequence, and then used for compression. Plotnik, Weinberger and Ziv[2] consider a source generated by a finite state Markov chain and show that maintaining separate dictionaries for each state of the source machine improves the rate of convergence of the algorithm.

In [1], a class of context dependent extensions to the Lempel-Ziv algorithm were described, in which multiple dictionaries were maintained, of which a subset (called the basis set, corresponding to a complete suffix tree of contexts) was chosen via dynamic programming to optimize an estimate of the compression achievable over the next phrase. This family of algorithms was shown to be asymptotically optimal, and showed promise of improved compression.

We here develop an alternative approach where the set of contexts selected at a given time need not, as in [1], correspond to a complete suffix tree. The method utilizes a more extensive set of performance estimates, which however is available via direct empirical observations for the proposed dictionary construction algorithms.

Associated with every context of length $\leq D$, we maintain a dictionary consisting of phrases seen in that context and the empirical performance of such dictionaries. For example, if $D = 3$, then, corresponding to the maximal depth context 010, we maintain a record of the performance of the dictionaries corresponding to context $\emptyset$ (the null context), 0, 10, and 010. These $D + 1$ numbers are updated each time the context is seen at the end of a phrase (not just when the context is actually used for compression). Compression of the next phrase is then via the dictionary corresponding to a current best empirical context. The decoder maintains the same estimates, and therefore knows the dictionary used.

Dictionary maintenance algorithms that we consider are closely related to those of [1]. For two of those algorithms, empirical performance measures are directly available as a consequence of the construction process. The two algorithms are (following the names in [1]):

- *Algorithm 2'– Multiple dictionaries*: Separate Lempel-Ziv trees are maintained for each possible context of depth upto D. Phrases are added to the corresponding dictionary every time a context is seen by means of constructs termed tokens added to the root of the LZ tree every time a context is seen at the end of a phrase, and then advanced through the tree using the subsequent symbols, ultimately being promoted to form a new node. When this occurs, the performance measures are updated.

- *Algorithm 3'–Compound dictionary*: In [1], it was suggested that it would be more efficient to view the multiple dictionaries as subtrees of a single larger dictionary, reached from the root via the appropriate context. This makes more efficient use of storage, and here too, tokens are used in the updating procedures.

Algorithms 2 and 2' have substantial overhead, which may be regarded as "wasted" since many of the dictionaries may not be used for compression. However, in Algorithm 3', dictionaries not used for compression also contribute to the growth of useful dictionaries, yielding better performance. Both the algorithms above are asymptotically optimal, as shown by the results of [1]. Experimental results on binary versions of ASCII files show that these new methods do better than standard Lempel-Ziv, and perform close to that of context allocation with dynamic programming.

References

[1] P. Franaszek, P. Tsoucas, and J.A. Thomas. Context allocation with an application to data compression. Technical Report 19022, IBM T.J. Watson Research Center, June 3 1993. Also presented at the IEEE International Symposium on Information Theory, Trondheim, Norway, June 27-July 1, 1994.

[2] E. Plotnik, M.J. Weinberger, and J. Ziv. Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the Lempel-Ziv algorithm. *IEEE Trans. Inform. Theory*, IT-38(1):66–72, January 1992.

[3] J. Rissanen. A universal data compression system. *IEEE Trans. Inform. Theory*, IT-29:656–664, July 1983.

[4] J. Rissanen. Complexity of strings in the class of Markov sources. *IEEE Trans. Inform. Theory*, IT-32:526–532, July 1986.

Universal Coding for Arbitrarily Varying Sources

Meir Feder and Neri Merhav
Department of Electrical Engineering - Systems, Tel Aviv University, Tel Aviv, 69978, ISRAEL
Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, ISRAEL

Abstract — **The minimum universal coding redundancy for finite-state arbitrarily varying sources, is investigated. If the space of all possible underlying state sequences is partitioned into types, then this minimum can be essentially lower bounded by the sum of two terms. The first is the minimum redundancy within the type class and the second is the minimum redundancy associated with a class of sources that can be thought of as "representatives" of the different types. While the first term is attributed to the cost of uncertainty within the type, the second term corresponds to the type itself. The bound is achievable by a Shannon code w.r.t an appropriate two-stage mixture of all arbitrarily varying sources in the class.**

We investigate the minimum attainable redundancy in universal coding for arbitrarily varying sources (AVS's). An AVS is a nonstationary memoryless source characterized by the probability mass function (PMF),

$$P(\mathbf{x}|\mathbf{s}) = \prod_{i=1}^{n} p(x_i|s_i), \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is an observed data sequence to be encoded, x_i taking on values in a finite set $\mathcal{X}$, and $\mathbf{s} = (s_1, \ldots, s_n)$ is an unknown arbitrary sequence of states corresponding to $\mathbf{x}$, where each s_i takes on values in a set $\mathcal{S}$. We shall assume, for the sake of simplicity, that the parameters of the AVS $\{p(x|s)\}_{x\in\mathcal{X}, s\in\mathcal{S}}$ are known.

The problem of universal coding for AVS's has relatively received only little attention. Berger [1, Sect. 6.1.2] and Csiszár and Körner [2, Theorem 4.3] have characterized the best attainable rate-distortion tradeoff for block-to-block (BB) codes where the average distortion is required to be within a prescribed level D for the *worst* possible state sequence. For the distortionless case ($D = 0$) the best attainable rate in this sense is given by the entropy of the worst memoryless source in the convex closure of $\{p(\cdot|s), s \in \mathcal{S}\}$, that is the maximum entropy attained among all mixtures $m(x) = \int_{\mathcal{S}} w(ds)p(x|s)$, w being a probability measure on $\mathcal{S}$. The reason for this worst case result is that both the rate is held fixed at each block and the distortion constraint must be met for every possible state sequence.

We show that one can improve upon this pessimistic result if variable-rate codes are allowed because then there is a potential freedom to "adapt" the rate to the underlying state sequence in some sense. Specifically, we show that for finite-state AVS's there exists lossless a block-to-variable (BV) code whose compression ratio is essentially the entropy of the memoryless source $m_{\mathbf{s}}(x) = \sum_{s\in\mathcal{S}} w_{\mathbf{s}}(s)p(x|s)$, where $w_{\mathbf{s}}(s)$ is the empirical probability (i.e., relative frequency) of $s \in \mathcal{S}$ along the underlying state sequence $\mathbf{s}$. This entropy is of course never larger than the maximum entropy mentioned above. It is therefore easy to see that the redundancy, namely, the excess rate beyond the per-letter entropy of the AVS given $\mathbf{s}$, is essentially equal to the mutual information $I_{w_{\mathbf{s}}}(S;X)$ associated with the joint PMF $w_{\mathbf{s}}(s)p(x|s)$. This quantity in turn agrees with that of [1] and [2] only if $\mathbf{s}$ maximizes the entropy.

Furthermore, $I_{w_{\mathbf{s}}}(S;X)$ is essentially a lower bound on the redundancy in a fairly strong sense. If we consider the set of all state sequences of a certain type class (i.e., the same empirical PMF $w_{\mathbf{s}}$) and hence yield the same $m_{\mathbf{s}}$, then by a direct application of [3, Theorem 1], for any uniquely decipherable code that is independent of $\mathbf{s}$, the redundancy is essentially never less than $I_{w_{\mathbf{s}}}(S;X)$ for *most* state sequences in this type class.

This bound is valid even if the type class in known a-priori. But if the type class is *not* known in advance intuition suggests that there must be an additional cost. We next demonstrate a coding scheme that is optimal in the sense of yielding the minimum attainable extra term, which in turn can be thought of as the redundancy associated with universal coding for a class of auxiliary sources that are "representing" the different type classes in a certain sense. Specifically, The proposed coding scheme can be interpreted as an hierarchical, two-step universal code, where the first step is to construct the best universal code within each type, and the second is to optimally integrate these codes by constructing another universal code for the class of the above mentioned auxiliary sources. The optimality of the proposed hierarchical code is in the sense that for any other code, most type classes have the property that except for a small minority of state sequences in the type class, the redundancy is essentially never less than the redundancy of the proposed code.

Finally, we point out that a natural subdivision of a class Λ of sources into subclasses $\Lambda_1, \Lambda_2, \ldots$, takes place in other situations as well. Another example is the class of all Markov sources, where Λ_i is the class of ith order Markov sources. The hierarchical universal coding approach demonstrated here, extends in the general case to a Shannon code w.r.t the double mixture, first over each Λ_i and then over $\{i\}$. Such a code was called "twice universal" in [4]. Similarly to Theorem 2, it can be shown that any other code cannot outperform the twice universal code, for "most" points in every Λ_i, except for a minority of classes Λ_i. Here by "most" we mean with high probabilty as measured by the mixture weights.

References

[1] T. Berger, *Rate Distortion Theory*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1971.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.

[3] N. Merhav and M. Feder, "A Strong Version of the Redundancy-Capacity Theorem of Universal Coding," to appear in *IEEE Trans. Inform. Theory*, May 1995.

[4] B. Ya. Ryabko, "Twice-universal coding," *Problems of Information Transmission*, pp. 173-177, July-Sept., 1984.

A TWO-STAGE UNIVERSAL CODING PROCEDURE USING SUFFICIENT STATISTICS

Toshiyasu MATSUSHIMA[1] and Shigeichi HIRASAWA

School of Science and Engineering, Waseda University 3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169 JAPAN

I. Introduction

A two-stage-procedure using the sufficient statistics of the parameters of the source models is proposed in this paper. In the procedure, the sufficient statistics calculated from a source sequence is transmitted at first stage. At second stage, the source sequence is encoded by using the conditional distribution given the sufficient statistics. Although the quantization is need to transmit the estimator vector in the previous two-stage codes [1][2], since the sufficient statistics is discrete random variable, the quantization is not need to transmit them. Moreover, the redundancy of the proposed code is equal to that of Bayes code[3][4] .

II. The previous two-stage codes

We assumed that a class of parameterized distributions of an information source is known but the parameters of the distribution function are unknown throughout this paper. Let $x_i \in A$ be a source symbol in a finite alphabet A. A source sequence is denoted by $x^n : x_1 x_2 \cdots x_n$. A parameterized distribution is denoted by $P(x^n|\theta)$ where $\theta \in \Theta$ is a real parameter vector in the parameter space Θ of the distribution.

In the coding procedures using MDL[1] or MML[2] criterion, at the first stage, the estimator $\hat{\theta}(x^n)$ of the parameter θ estimated from a source sequence x^n is encoded. At the second stage, the source sequence is encoded by using the estimator $\hat{\theta}(x^n)$. The code word length $L_M(x^n)$ of these procedures is represented by

$$L_M(x^n) = L(\hat{\theta}(x^n)) - \log P(x^n|\theta = \hat{\theta}(x^n)). \quad (1)$$

The first term of the right hand side $L(\hat{\theta}(x^n))$ represents the description length of the estimator vector $\hat{\theta}(x^n)$ itself. The second term is the ideal codeword length of a source sequence x^n encoded by $P(x^n|\theta = \hat{\theta}(x^n))$: the distribution whose parameter is substituted by the estimator $\hat{\theta}(x^n)$.

However, since the parameter vector is real, the quantization of the estimator vector $\hat{\theta}(x^n)$ is need to transmit it. MDL criterion was induced by considering the quantized scale of the estimator $\hat{\theta}(x^n)$ to minimize the total description length $L_M(x^n)$. MML criterion was also derived by studying encoding method of the estimator using the prior distribution $P(\theta)$.

III. Sufficient statistics code

The fundamental difference between sufficient statistics codes and the previous two-stage codes is to transmit the sufficient statistics $u(x^n)$ instead of the estimator $\hat{\theta}(x^n)$.

The sufficient statistics $u(x^n)$ satisfies the following equality.

$$H(\theta|u(x^n)) = H(\theta|x^n). \quad (2)$$

The above equality indicates that the sufficient statistics $u(x^n)$ includes all information with respect to θ in a source sequence x^n. Thus, there is no information loss with respect to θ by transmitting $u(x^n)$ instead of the estimator $\hat{\theta}(x^n)$.

In sufficient statistics codes, at the first stage, the sufficient statistics $u(x^n)$ is encoded and transmitted. At the second stage, the source sequence x^n is encoded by using $P(x^n|u(x^n))$: the conditional probability of x^n given $u(x^n)$.

The encoding probability $P(x^n|u(x^n))$ at the second stage essentially differs from $P(x^n|\theta = \hat{\theta}(x^n))$ used in the previous two-stage procedures. $P(x^n|\theta = \hat{\theta}(x^n))$ is given by substituting $\hat{\theta}(x^n)$ for θ in the source distribution $P(x^n|\theta)$. $P(x^n|\theta = \hat{\theta}(x^n))$ is different from the conditional probability $P(x^n|\hat{\theta}(x^n))$ under the condition that $\hat{\theta}(x^n)$ is estimated from a source sequence x^n. In sufficient statistics code, the conditional probability $P(x^n|u(x^n))$ under the condition that the sufficient statistics $u(x^n)$ is calculated from x^n, is used for the encoding probability.

The ideal codeword length $L_S(x^n)$ of sufficient statistics codes is given as follows:

$$L_S(x^n) = L(u(x^n)) - \log P(x^n|u(x^n)). \quad (3)$$

The first term of the right hand side of the above expression $L(u(x^n))$ represents the description length of the sufficient statistics $u(x^n)$ in first stage of the procedure. Although the quantization is need to transmit the estimator vector $\hat{\theta}(x^n)$ in the previous two-stage-code, since the sufficient statistics $u(x^n)$ is discrete random variable, the quantization is not need to transmit $u(x^n)$.

The second term is the ideal codeword length of the source sequence x^n in the second stage. The term is uniquely determined by the conditional probability $P(x^n|u(x^n))$. Then, the total code word length of sufficient statistics codes is depend on the coding method of $u(x^n)$.

Theorem 1 *The ideal code word length of sufficient statistics code $L_S(x^n)$ is identical with that of Bayes code, if the description length $L(u(x^n))$ of $u(x^n)$ as follows:*

$$L(u(x^n)) = -\log \int P(u(x^n)|\theta)P(\theta)d\theta. \quad (4)$$

Various type of sufficient statistics codes can be constructed by changing the prior distribution $P(\theta)$ as Bayes codes. Especially, the minimax redundancy codes are constructed by using the least favorable prior for the redundancy risk.

References

[1] J. Risannen. Stochastic complexity and modeling. *Annls of Statistics*, 14(3):1080–1100, 1986.

[2] C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *J. R. Stat. Soc. B*, 240–265, 1987.

[3] L. D. Davison. Universal noiseless coding. *IEEE Trans. Inf. Theory*, 19(6):783–795, Nov 1973.

[4] T. Matsushima, H. Inazumi, and S. Hirasawa. A class of distortionless codes designed by bayes decision theory. *IEEE Trans. Inf. Theory*, 37(5):1288–1293, Sep 1991.

[1] E-mail: toshi@matsu.mgmt.waseda.ac.jp

Adaptive limitation of the dictionary size in LZW data compression

K.Ouaissa, M.Abdat, P.Plume
Lab.Elect. & Comm., CNAM, Paris, France, *E-mail: ouaissa@cnam.fr*

Abstract — **Two modifications of the Lempel-Ziv-Welch (LZW) algorithm are presented to limit the dictionary size. First, a run-length encoding (RLE) is combined with the LZW algorithm, in order to pre-select the input data. Then, a dynamic update of the dictionary is performed by eliminating the free branches in the tree representing the dictionary.**

I. Introduction

The LZW technique is included in the V42 bis recommendation of the CCITT and it is widely used in communications. Basically, it is a lossless and a non statistic compression algorithm which maps variable length strings to fixed length indexes (codewords). It has the advantage of being adaptive and does not assume any advance knowledge of the source properties. It uses a dictionary which is built by performing a string matching after each source symbol occurrence. String of different lengths are represented by indexes in the dictionary, which is the same at the encoder and decoder. During the compression phase, the dictionary is built on the basis of the input symbols and the coder becomes more efficient with the growth of the table [1], [3]. However, once the dictionary is full, no adaptivity is provided any more. In order to be able to add a very long strings to the table, the algorithm needs a very large dictionary, the code words become very large and the compression ratio decreases. To counter this problem a compromise is necessary. In fact, the compression is optimised when the dictionary is a real mirror of the input statistics. With text or source files, long repetitive strings provide less information, and, thus, the corresponding space in the dictionary is not efficiently used. Therefore, the algorithm must continuously update the dictionary, without increasing its size. That can be achieved by the two following schemes:

-Combining the LZW algorithm with a run length encoding to avoid overloading the dictionary with long repetitive sequences (pre-selection).

-Eliminating less probable strings from the dictionary, in order to keep a sufficient level of adaptivity for the algorithm.

II. combining LZW and run-length encoding

The run-length encoding eliminates the repeated symbols from the input data. The number N of repetitions must be greater than a pre-defined threshold. It exchanges all the repetitions in the stream of data with a special sequence. The LZW algorithm can be introduced in the cascade as follows. The creation of a new string (Xi+y) in the dictionary is made by concatenating a unique character (y) with a string (Xi) present in the dictionary. The run-length encoding technique scans the input strings; if the input is a repetition of N symbols (N greater than a pre-defined threshold), without using the dictionary, the algorithm outputs a run-length encoding, to code the repetitions. Then, the LZW-RLE coder continues normally the coding process with the LZW algorithm. The compression ratios achieved respectively by the LZW encoder and the RLE are compared. According to our simulations, the threshold value N must be greater than 10.

III. Dynamic update

In the LZW algorithm, the update of the dictionary stops when the dictionary is full. For example, in the CCITT V42 bis recommendation [2], when the dictionary is full, the algorithm deletes the old dictionary (flush) and starts building a new one. The compression ratio decreases considerably after the dictionary flush. Instead of deleting the entire dictionary, it is proposed to delete just a section, namely all the free branches of the tree representing it. The procedure is as follows. While building the dictionary, the algorithm marks all the free branches, using a one row table. Once the dictionary is full, the flush phase deletes all the branches already marked. This technique keeps the very long strings, so that the statistical properties of the input are well known and the previously deleted branches are used to continue the update. The number of free branches deleted in each flush phase allows us to follow the evolution of the algorithm and estimate if it is better to delete only a part or all the dictionary. In fact, after several updates, the number of free branches tends to become a constant value. It corresponds to the saturation of the dictionary. At this point, deleting all the dictionary is the best solution.

IV. Results

The improvement in compression ratio, with the LZW-RLE coder is confirmed by several tests with standard files. It provides better performance during the learning phase with less complexity. The improvement is around 4 to 6 percent with respect to the LZW algorithm alone. The dynamic update acts once the dictionary is full. It provides an improvement of 4 percent with respect to the V42 bis method. The flush threshold value seems to be around 1000 free branches.

V. conclusion

In this paper, two modified algorithms based on combining the run-length encoding with (LZW) and the free branches deleting method have been analysed and simulated. A significant compression ratio improvement is achieved with the LZW-RLE coder, when repetition sequences are present in the file. The LZW-RLE coder does not affect the compression ratio in the case of normal files (files without repetitions). Application to speech and image coding leads to further refinements which are presently under study.

References

[1] P. PLUME, "Compression de donnees methodes, algorithmes et programmes detailles," *Eyrolles.* , 1993.

[2] CCITT, "Communication de donnees sur le reseau telephonique" *recommandation V42bis.* , Geneve, 1990.

[3] P. E. Bender, J. K. Wolf, "New asymptotic bounds and improvements on the Lempel-Ziv data compression algorithm " *IEEE Transactions on information theory,*, vol. 37, pt. I, pp. 379–423, 1991; pt. II, pp. 623–656, 1991.

Universal Coding of Integers and Unbounded Search Trees

R. Ahlswede[1], T. S. Han[2] and K. Kobayashi[3]

[1] Bielefeld Universität, Fakultät Mathematik, POB 100131, 33501 Bielefeld, Germany
[2] The University of Electro-Communications, Graduate School of Information Systems
and , [3] Department of Computer Science and Information Mathematics,
Chofugaoka 1-5-1, Chofu, Tokyo, 182, JAPAN

Abstract — **We study the universal coding problem for the integers, in particular, establish rather sharp lower and upper bounds for the Elias omega code and other codes. For these bounds, the so-called log-star function plays the central role. Furthermore, we invesigate unbounded search trees induced by these codes, including the Bentley-Yao search tree.**

1. Elias omega code and related codes

Let us denote the standard binary expression of positive integer $j \in \mathcal{N}^+$ as $(j)_2$. For example, $(13)_2 = 1101$. The binary expression of integer j to base 2^k is denoted by $(j)_{2,k}$. Next we express the floor function of log by $\lambda_2(j) = \lfloor \log_2 j \rfloor$. Moreover, λ_2^k is the k-fold composition of function λ_2.

Elias[1] introduced a universal code $\omega : \mathcal{N}^+ \to \{0,1\}^*$, called the ω-code, described by

$$\omega(j) = \begin{cases} 0 & , \text{if } j = 1 \\ (\lambda_2^{k-1}(j))_2 \cdots (\lambda_2(j))_2 (j)_2 0 & , \text{if } j \geq 2 \end{cases} \quad (1)$$

where $k = k(j)$ is the positive integer satisfying $\lambda_2^k(j) = 1$ (which exists for any $j \geq 2$). Then the codeword length of this prefix code ω is given by

$$c_E(j) = |\omega(j)| = \sum_{i \geq 1 : \lambda_2^i(j) \geq 0} (\lambda_2^i(j) + 1) \quad (j = 1, 2, \ldots). \quad (2)$$

Another class of universal codes introduced by Stout[3] is given, for any integer $d \geq 0$, by

$$S_d(j) = \begin{cases} (j)_{2,d} 0 & , \text{if } 0 \leq j < 2^d, \\ (\lambda_{[d]}^k(j))_{2,d} (\lambda_{[d]}^{k-1}(j))_2 \cdots & \\ \quad \cdots (\lambda_{[d]}^2(j))_2 (\lambda_{[d]}(j))_2 (j)_2 0 & , \text{if } j \geq 2^d, \end{cases} \quad (3)$$

for $j \in \mathcal{N} = \{0, 1, \ldots\}$, where

$$\lambda_{[d]}(x) = \lfloor \log_2 x \rfloor - d \quad (x > 0), \quad (4)$$

$\lambda_{[d]}^t$ is the t-fold composition of the function $\lambda_{[d]}$, and k is the positive integer satisfying $0 \leq \lambda_{[d]}^k(j) < 2^d$.

Stout has defined the code S_d only for $d \geq 2$. S_0 is identical to the code introduced by Levenshtein[4].

2. Bounds for the codeword lengths

In order to introduce the bound for $c_E(j)$, we define the **log-star** function $\log_2^*(x)$ for $x \geq 1$ as

$$\log_2^*(x) \equiv \log_2(x) + \log_2 \log_2(x) + \cdots + \log_2^{w^*(x)}(x) \quad (5)$$

where $\log_2^w(x)$ is the k-fold composition of the function $\log_2(x)$, and $w^*(x)$ is the largest positive integer satisfying $\log_2^w(x) \geq 0$. Therefore, $w^*(x) = 1, \log_2^*(x) = 0$ for $x = 1$.

Then we established upper and lower bounds for the length function $c_E(j)$.

□ **Theorem 1** For any real $x \geq 1$,

$$\log_2^*(x) < c_E(x) \leq \log_2^*(x) + w^*(x). \quad (6)$$

Here we have extended the domain of function $c_E(\cdot)$ to the set of real numbers through the extension of λ_2. Through a simple consideration, we can check that the upper bound is attained at the points $j_m = \exp_2^m(1)$ $(m = 0, 1, \ldots)$, where $\exp_2(x) = 2^x$ and $\exp_2^k(x)$ is the k-hold composition of function $\exp_2(\cdot)$. Moreover, the lower bound is also attained at the same points in the sense of

$$\lim_{x \uparrow j_m} c_E(x) = \log_2^*(j_m). \quad (7)$$

Therefore, the two bounds are best possible as far as we restrict the bounding functions to such smooth functions.

We can obtain same bounds for the codeword length of the Stout code S_d by a similar argument.

Furthermore, the unbounded search trees on $\mathcal{N}^+$ induced by the Elias omega code and Stout codes have a more beautiful recursive structures than Bentley-Yao search tree[2].

3. Modified log-star function

Define the **modified log-star** function by

$$\log_{r,\alpha}^*(x) = \log_r^*(x) - \alpha w_r^*(x) \quad (x \geq 1) \quad (8)$$

for integer $r \geq 2$ and real number α. Then, we have

□ **Lemma 1** For integer $r \geq 2$, set $\alpha_r^* = \log_r(\log_r e)$.

1) If $\alpha < \alpha_r^*$, then

$$\sum_{j=1}^{\infty} r^{-\log_{r,\alpha}^*(j)} < +\infty, \quad (9)$$

2) If $\alpha \geq \alpha_r^*$, then

$$\sum_{j=1}^{\infty} r^{-\log_{r,\alpha}^*(j)} = +\infty. \quad (10)$$

From the lemma, we can show the existance of better prefix codes than Elias omega code, and other known codes.

References

[1] P.Elias, "Universal codword sets and representation of the integers," *IEEE Trans. on Information Theory*, vol.IT-21, pp.194–203, 1975.

[2] J.L.Bentley and A.C.Yao, "An almost optimal algorithm for unbounded searching," *Information Processing Letters*, vol.5, no.3, pp.82–87, 1976.

[3] Q.F.Stout, "Improved Prefix Encodings of the Natural Numbers," *IEEE Trans. on Information Theory*, vol.IT-26, pp.607–609, 1980.

[4] V.I.Levenshtein, "On the redundancy and delay of decodable coding of natural numbers,"(in Russian) *Problems of Cybernetics*, vol.20, 173-179, 1968.

On The Context Tree Maximizing Algorithm

Paul A.J. Volf and Frans M.J. Willems
Information and Communication Theory Group, Eindhoven University of Technology

Abstract — **The context tree weighting algorithm was introduced at the 1993 ISIT. Here we are concerned with the context tree maximizing algorithm. We discuss several modifications of this algorithm.**

I. Introduction

In this paper we assume that the source has a tree structure. The context (e.g. the most recent symbols from the source sequence) selects one of the leaves. Symbols following this context are assumed to be independent. The tree structure is called the *model* of the source. A full tree with depth D and with symbol counts in its nodes and leaves is called a *context tree*. In [2] an one-pass algorithm, the *context tree weighting* algorithm was introduced. This method uses such a tree.

For the context tree weighting algorithm it was proved that the individual redundancy ρ of a source sequence x_1^T, with respect to a binary source with model $\mathcal{S}$ and with parametervector $\Theta_{\mathcal{S}}$ satisfies (the terms represent the model, parameter and coding redundancy respectively):

$$\rho(x_1^T|x_{1-D}^0,\mathcal{S},\Theta_{\mathcal{S}}) < (2|\mathcal{S}|-1)+(\frac{|\mathcal{S}|}{2}\log\frac{T}{|\mathcal{S}|}+|\mathcal{S}|)+2.$$

This holds for every model $\mathcal{S}$ and every parametervector $\Theta_{\mathcal{S}}$.

The *context tree maximizing* algorithm (see also [1]), a two-pass algorithm, fulfils the same upperbound, but at the same time, it will give a slightly longer codeword. During the first pass the counts in the tree will be updated. After the first pass the two-pass algorithm will determine the "best" model, and in the second pass it uses this model to compress the sequence. Two-pass algorithms can have distinct advantages. Most important is that their decoding complexity is considerably less than the complexity of the weighting algorithm.

II. The context maximizing algorithm

Just like the weighting algorithm, this algorithm uses the Krichevsky-Trofimov estimator for encoding memoryless sequences. This results in the following block probability for a sequence with a zeros and b ones (if $a>0$ and $b>0$) :

$$P_e(a,b)=\frac{\frac{1}{2}\cdot\frac{3}{2}\cdot\ldots\cdot(a-\frac{1}{2})\cdot\frac{1}{2}\cdot\ldots\cdot(b-\frac{1}{2})}{1\cdot2\cdot\ldots\cdot(a+b)}.$$

In every node of the context tree we compute the maximized probability according to the following formula. With D we denote the maximum level of the tree, and $l(s)$ is the depth of the context in node s. We define

$$P_m^s=\begin{cases}P_e(a_s,b_s) & \text{if } l(s)=D,\\ \frac{1}{2}\max(P_e(a_s,b_s),P_m^{0s}P_m^{1s}) & \text{if } l(s)<D.\end{cases}$$

One can find the model by walking depth-first through the tree. If the product of the maximized probabilities of the children is larger than the P_e in node s, then s must be an internal node of the model, else s is a leaf. The maximizing algorithm will find a model which minimizes the *description length* (MDL). The description length is the sum of the cost needed to describe the model (the factors $\frac{1}{2}$) and the cost of describing the data with this model (P_e).

III. The yoyo algorithm

The maximizing algorithm can be modified such that it produces a model with not more than C leaves (parameters), to limit the complexity of the decoder. We walk through the context tree again in a depth-first search way. In every node we compute a list which contains for all $c=1,C$ the maximized probability achievable with not more than c leaves. In each node the list can be computed by combining the estimated probability in that node with the lists from its two children.

For every total number of leaves one looks for the distribution of leaves over its two children that results in the highest product of the maximized probabilities. Finally one finds a list in the root with for every number of leaves up to C, the corresponding maximized probability.

To determine the list in the root one needs at most $D+1$ open lists. Once one knows the appropriate total number of leaves, one knows which distribution of the number of leaves over each child (of the root) resulted in this "optimal" solution. In this way the problem is reduced to two trees of depth $D-1$. If one applies this technique recursively, we will find the best constrained model.

IV. Model description on-the-fly

Instead of sending the model description first, followed by the code for the data, we now use a growing model. The decoder walks through the context tree as far the current model allows. If the new context passes an endpoint (leaf) of the current model, which is not known to be a leaf or internal node of the MDL model yet, and this new context differs from the previous ones that have passed this endpoint, then the decoder needs more information about the model. We must first tell him that the endpoint is a leaf or not. If not we should give him the same information about the next node on the context path, etc. This process ends when the current context diverges from the previous ones. The diverging node must be included. In this way the current model grows to the MDL-model.

In total the encoder now has to describe all internal nodes of the found model, plus all leaves (not at the maximum depth) which are followed by different context sequences.

With this technique we gain compared to the original two-pass algorithm. But the model costs in the weighting algorithm are similar. The maximizing algorithms can be modified such that the best "on-the-fly models" will be found.

V. Improved model description

In binary, but especially in non-binary trees, with on-the-fly model description, the number of children of a node that need specification is not known in advance. Using an estimator, e.g. P_e, to specify these children, we get improved compression.

References

[1] P. Volf and F. Willems. Context maximizing: Finding MDL decision trees. In *15th Symp. Inform. Theory Benelux*, pp. 192-200, Louvain-la-Neuve, Belgium, May 1994.

[2] F. Willems, Y. Shtarkov, and Tj. Tjalkens. Context tree weighting: A sequential universal source coding procedure for FSMX sources. In *IEEE ISIT*, page 59, San Antonio, Texas, Jan 1993.

A Fixed-to-Variable Variation of the Ziv-Lempel Coding

Ken-ichi Iwata[1], Tomohiko Uyematsu, and Eiji Okamoto

School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, 921-12, Japan

Abstract — **This paper presents a fixed-to-variable variation of the Ziv-Lempel code called "FVLZ code", and clarifies its asymptotic performance with respect to a non-probabilistic model for constrained sources proposed by Ziv and Lempel. It is shown that the FVLZ code has almost the same asymptotic performance as the Ziv-Lempel code.**

I. Ziv-Lempel Coding

In 1977, Ziv and Lempel proposed a universal coding algorithm called "LZ77 code"[1]. The LZ77 coding algorithm parses input data into a sequence of phrases with their length less than F, each of which excluding the last symbol is the longest matched string in a sliding window consisting of the previously encoded $N-F$ symbols. The phrases are represented by the position and length of the longest matched string in the window as well as the following un-matched symbol, and these triples are encoded into a codeword. Then, the window slides to the position just before the next symbol to be encoded.

II. Description of the Algorithm

We begin with the description of the FVLZ coding algorithm. Let A be a finite alphabet set with α elements, where $\alpha \geq 2$. Let C_i be the ith FVLZ codeword which is obtained by concatenating some intermediate codewords C_i^j $(1 \leq j \leq E(i))$ described later, i.e., $C_i = C_i^1 \ldots C_i^{E(i)}$. Let $d(C_i^j)$ denote the length of the input data encoded into the jth intermediate codeword C_i^j, and let $l_i^j = \sum_{k=1}^{j-1} d(C_i^j)$ where $l_i^1 = 0$. Assume that p indicates the number of encoded symbols. Then, the FVLZ coding algorithm can be described as follows:

Step 1 (Initialization) Sliding window is initialized in the similar manner as the LZ77 coding algorithm.

Step 2 (Encoding) Obtain the intermediate codeword C_i^j by using the LZ77 coding algorithm assuming that sliding window consists of the previously encoded $N-F+l_i^j$ symbols and the length of longest matched string is less than $F-l_i^j$. Then, the contents of C_i^j are represented with lengths specified in Table 1. If $p \bmod F = 0$, output the ith codeword $C_i = C_i^1 C_i^2 \cdots C_i^{E(i)}$, and refresh the sliding window by shifting F symbols to obtain the next FVLZ codeword. Repeat Step 2 until the whole input data is encoded. □

Table 1: Lengths of intermediate-codewords

Contents to be represented	Length
Starting position:	$\lceil \log(N-F+l_i^j) \rceil$
Longest length:	$\lceil \log(F-l_i^j) \rceil$
Last symbol (un-matched symbol):	$\lceil \log \alpha \rceil$

III. Example

Fig.1 shows an example of the FVLZ encoding for A={a,b}, N=16 and F=8. $B_i(1,N)$ in Fig.1A denotes a string in the sliding window used for the encoding of the ith FVLZ codeword C_i (i.e., $B_i(1,N)$ =abbbbaababbababa).

[1]e-mail: k-iwata@jaist.ac.jp

IV. Analysis

In this section, we clarify an asymptotic performance of the FVLZ code and compare it with that of the LZ77 code[1]. To this end, we employ a following model for constrained sources which was defined by Ziv and Lempel[1]. Let A^* denote the set of all strings of finite length over A. Given a string $S \in A^*$ of length $l(S)$ and a positive integer $m \leq l(S)$, and $S\{m\}$ denotes the set of all substrings of length m contained in S. Given a subset σ of A^*, and let $\sigma\{m\} = \{S \in \sigma | l(S) = m\}$. Assume that $\sigma(m)$ denotes the cardinality of $\sigma\{m\}$. Then, a subset σ of A^* is called a source, if the following three properties hold: 1) $A \subset \sigma$, 2) $S \in \sigma$ implies $SS \in \sigma$, 3) $S \in \sigma$ implies $S\{m\} \subset \sigma\{m\}$. With every source σ, we associate a sequence $h(1), h(2), \cdots$ of parameters, called the h-parameters of σ, where $h(m) = \frac{1}{m}\log(\sigma(m)) \quad m = 1, 2, \cdots$. Let the compression ratio ρ be

$$\rho = \frac{\text{total length of codewords}}{\text{encoded source length}}. \tag{1}$$

Now, we can state the following result.

Theorem 1 If the length of the sliding window N for a source with known h-parameters is chosen by $N = FM_F$ where $M_F = (F-1)\left\{\sum_{m=1}^{\lambda} \alpha^m + \sum_{m=\lambda+1}^{F-1} \sigma(F-1)\right\} + F$, $\lambda = \lfloor (F-1)h(F-1) \rfloor$. Then, the compression ratio ρ attainable by the FVLZ code satisfies

$$\rho \leq h(F-1) + \varepsilon(F), \tag{2}$$

where $\varepsilon(F) = (3 + \log(F-1) + 3\log F)/(F-1)$. □

By using Theorem 1, we can show the universality of the FVLZ code in the similar manner as Ziv and Lempel did for the LZ77 code in Ref.[1]. Since $\varepsilon(F)$ of Eq.(2) is equal to that of the LZ77 code up to the coefficient of the highest order, we can show that the FVLZ code has almost the same asymptotic performance as the LZ77 code has. Further, experimental results reveal that the FVLZ code and the LZ77 code provide almost the same performance from the viewpoint of compression ratio and encoding/decoding time, as well as it requires almost the same amount of memory as the LZ77 code.

References

[1] J. Ziv and A. Lempel: "A Universal Algorithm for Sequential Data Compression", *IEEE Trans. Inform. Theory*, vol. IT-23, no.3, pp.337–343, 1977.

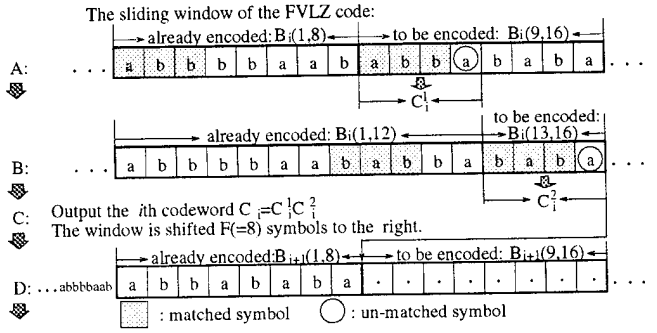


Figure 1: Encoding by FVLZ code with N=16 and F=8.

Proposal of partially decodable Ziv-Lempel code

Ken-ichi Iwata[1], Tomohiko Uyematsu, and Eiji Okamoto
School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, 923-12, Japan

Abstract — **This paper presents a new variation of the Ziv-Lempel code called "Partial Decodable Ziv-Lempel (PDLZ) code", which can decode a part of the encoded data from a sequence of codewords.**

I. Ziv-Lempel Coding

In 1977, Ziv and Lempel proposed a universal coding algorithm called "LZ77 code"[1]. The LZ77 coding algorithm parses input data into a sequence of phrases with their length less than L, each of which excluding the last symbol is the longest matched string in a sliding window consisting of the previously encoded $N-L$ symbols. The phrases are represented by the position and length of the longest matched string in the window as well as the following un-matched symbol, and these triples are encoded into a codeword. Then, the window slides to the position just before the next symbol to be encoded. Now, we define the matched relation as the relation between the ith symbol of the longest matched string in the sliding window and the ith symbol of the parsed phrase to be encoded. It is noted that if a_1 and a_2 are in matched relation, and a_2 and a_3 are in matched relation, then a_1 and a_3 are also in matched relation.

II. Description of the Algorithm

For each symbol in the sliding window, let a quotation symbol be the oldest symbol in matched relation with the symbol. Then, the quotation symbol has been encoded as an un-matched symbol. Let the quotation set be the set of the previously encoded K symbols. Then, the PDLZ coding algorithm can be described as follows:

Step 1 (Initialization) Sliding window is initialized in the same manner as the LZ77 coding algorithm.

Step 2 (Encoding) Obtain the next phrase to be encoded in the same manner as the LZ77 coding algorithm. Then, execute the following procedure:

Case 1: If the quotation symbols corresponding to the obtained phrase, are all in the quotation set, then the phrase is encoded into the ith codeword C_i in the same manner as the LZ77 coding algorithm.

Case 2: Otherwise, we divide the obtained phrase into some substrings, such that each last symbol in the substrings except for the last substring has the quotation symbol out of the quotation set. Then, each substring is encoded into the intermediate codeword C_i^j $j=1,2,\cdots$ in the similar manner as the LZ77 coding algorithm, and obtain C_i by concatenating C_i^j.

Refresh the sliding window to obtain the next codeword in the same manner as the LZ77 coding algorithm. Repeat Step 2 until the whole input data is encoded. □

Fig.1 shows an example of the PDLZ encoding for an input alphabet set A={a,b}, N=12, L=6 and K=6. For each symbol in the sliding window, the quotation window as shown in Fig.1(i) stores the position of the corresponding quotation symbol in terms of the length from the next symbol to be encoded.

[1]e-mail: k-iwata@jaist.ac.jp

III. Analysis

In this section, we clarify an asymptotic performance of the PDLZ code. To this end, we employ a following model for constrained sources which was defined by Ziv and Lempel[1]. Let A be a finite alphabet set with α elements, where $\alpha \geq 2$, and A^* denotes the set of all strings of finite length over A. Given a string $S \in A^*$ of length $l(S)$ and a positive integer $m \leq l(S)$, and $S\{m\}$ denotes the set of all substrings of length m contained in S. Given a subset σ of A^*, and let $\sigma\{m\} = \{S \in \sigma | l(S) = m\}$. Assume that $\sigma(m)$ denotes the cardinality of $\sigma\{m\}$. Then, a subset σ of A^* is called a source, if the following three properties hold: 1) $A \subset \sigma$, 2) $S \in \sigma$ implies $SS \in \sigma$, 3) $S \in \sigma$ implies $S\{m\} \subset \sigma\{m\}$. With every source σ, we associate a sequence $h(1), h(2), \cdots$ of parameters, called the h-parameters of σ, where $h(m) = \frac{1}{m}\log(\sigma(m)) \quad m = 1, 2, \cdots$. Let the compression ratio ρ be

$$\rho = \frac{\text{total length of codewords}}{\text{encoded source length}}. \tag{1}$$

Now, we can state the following result.

Theorem 1 Assume that for a source with known h-parameters, the length of sliding window N is chosen by

$$N = (L-1)\left\{\sum_{m=1}^{\lambda}(L-m)\alpha^m + \sum_{m=\lambda+1}^{L-1}(L-m)\sigma(L-1)\right\}+L,$$

where $\lambda = \lfloor (L-1)h(L-1) \rfloor$. Further, let K be specified by

$$K = (N-L)^{1+\xi}, \qquad \xi > 0. \tag{2}$$

Then, the compression ratio ρ attainable by the PDLZ code satisfies

$$\rho \leq \{h(L-1) + \varepsilon_1(L)\}\{1+\varepsilon_2(L)\}, \tag{3}$$

where $\varepsilon_1(L) = (3 + 3\log(L-1) + \log(L/2))/(L-1)$ and $\varepsilon_2(L) = 2^{\xi}/(L-1)^{2\xi-1}L^{\xi}$. □

Theorem 1 implies the following corollary.

Corollary 1 If $K > (N-L)^{4/3}$ then the PDLZ code is a universal code in the sense of Ziv and Lempel[1]. □

References

[1] J. Ziv and A. Lempel: "A Universal Algorithm for Sequential Data Compression", *IEEE Trans. Inform. Theory*, vol. IT–23, no.3, pp.337–343, 1977.

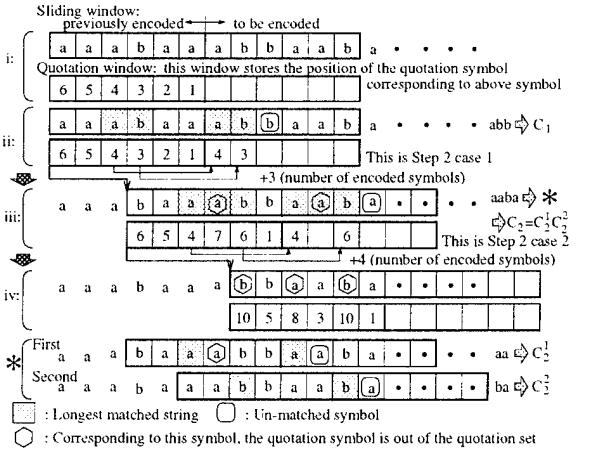


Figure 1: PDLZ encoding with N=12, L=6 and K=6.

Coded Multicarrier Code Division Multiple Access

Douglas N. Rowitch and Laurence B. Milstein[1]

Dept. of Elect. and Comp. Eng., University of California, San Diego
La Jolla, California 92093

Abstract — **This paper presents a multicarrier signaling technique for an asynchronous Direct Sequence (DS) Code Division Multiple Access (CDMA) system which employs linear convolutional codes to achieve frequency diversity performance gains in excess of path diversity gains realized in conventional single carrier RAKE DS CDMA systems.**

I. Overview

DS CDMA is a popular signaling technique in which binary data sequences for multiple access users are modulated by unique spreading signature sequences having bandwidth much greater than that of the data. Waveforms are transmitted simultaneously over the same frequency band and are distinguished at the receiver via a correlation operation against the spreading code of the user-of-interest. We consider a slowly varying, Rayleigh fading multipath channel, where the spread bandwidth exceeds the coherence bandwidth of the channel, and, thus, the signals are said to fade in a frequency selective manner. In such systems, a RAKE receiver is often employed to combine the energy received over several resolvable propagation paths.

We present an alternative system where the available frequency bandwidth is decomposed into M distinct sub-bands, each of an bandwidth equal to the coherence bandwidth of the channel. The sub-channels, therefore, tend to fade non-selectively, and are assumed to fade independently. In short, we exchange path diversity for frequency diversity, wherein forward error correction may be utilized without the penalty of bandwidth expansion.

II. Summary

The data sequence for a given user is input to a rate 1/M convolutional encoder (where M is the number of carriers) and each of the M outputs are multiplied by a spreading sequence which, in turn, modulates the M carrier tones. The receiver utilizes coherent BPSK detection and weights the outputs of each correlator in an optimum fashion. These outputs are then used to calculate branch metrics in a soft decision Viterbi decoder. Whereas the conventional DS CDMA system experiences path diversity on the order of the number of resolvable paths, the coded multicarrier DS CDMA system experiences frequency diversity on the order of the number of carriers plus an effective diversity improvement on the order of the minimum free distance of the convolutional code [1]. The diversity gains realized make for significant improvements in user capacity, while preserving the desirable properties exhibited in DS CDMA systems: robustness to fading, tolerance to multiple access interference, and a narrowband interference suppression effect [2].

The performance of the coded multicarrier system is compared to that of a conventional single carrier system in the presence of additive white Gaussian noise, multiple access interference, and Gaussian partial-band interference. It can be shown that the outputs of the M sub-channel correlators are approximately conditionally Gaussian, conditioned on the respective channel fade amplitudes [3]. We derive the optimal correlator weights and branch metrics for the soft decision decoder using standard methods [1].

To obtain an upper bound on the average probability of bit error, we assume that the all-zero path is sent and consider the event that some competing path is selected. This is accomplished by developing a convolutional code generating function evaluated in terms of an exponential upper bound on the probability of a pairwise error event [1]. Since the variances of sub-channel correlator outputs may be different, due to partial-band interference, we consider the pairwise error event of a competing path containing precisely d_i code bit errors in the i^{th} bit location (i.e., i^{th} sub-channel). It can be shown that the Chernoff bound on this probability is

$$P_2(d_1,..,d_M) \le \prod_{i=1}^{M} \left(\frac{1}{1+\bar{\gamma}_i} \right)^{d_i},$$

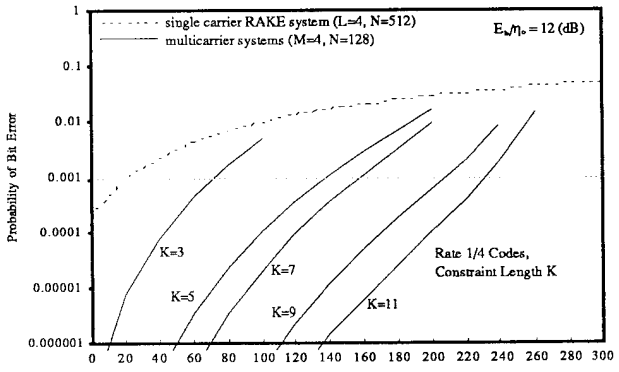
where $\bar{\gamma}_i$ is the average signal-to-noise ratio of the i^{th} channel. It is then straightforward to develop a generating function for a particular convolutional code which enumerates not just the number of code bit errors over a path, but the location (i.e., sub-channel) of those bit errors, whereupon the probability of bit error may be union bounded as

$$P_b \le \left. \frac{dT(D_1,..,D_M,N)}{dN} \right|_{N=1,\ D_i=\frac{1}{1+\bar{\gamma}_i},\ i=1,..,M}.$$

To analyze and compare these systems, we selected raised-cosine chip wave-shaping filters with 50% excess bandwidth. Single carrier RAKE system performance is taken as equivalent to that of 4^{th} order path diversity reception using maximal-ratio combining [1]. The multicarrier system employs 4 carriers, and thus, rate 1/4 codes of varying constraint lengths [4]. We hold total system bandwidth, information rate, and energy-per-bit constant. The figure below depicts the upper bound on the BER for multicarrier systems as a function of the number of muliple access users for E_b/η_o fixed at 12 (dB). At a BER of 10^{-3}, significant capacity gains are realized as an increasing function of the code constraint length.

References

[1] J. Proakis, *Digital Communications*. McGraw-Hill, New York, 1989.
[2] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*. McGraw-Hill, New York, 1994.
[3] S. Kondo and L. B. Milstein, "On the Performance of Multicarrier DS CDMA Systems," submitted to *IEEE Trans. Commun.*
[4] K. J. Larsen, "Short Convolutional Codes with Maximal Free Distance for rates 1/2, 1/3, and 1/4," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 371-372, May 1973.



[1] This work was supported in part by the National Science Foundation under Grant NCR-9213140.

The Performance of Voice and Data Communications in a Mobile Cellular CDMA System

John H. Gass, Jr., Daniel L. Noneaker, and Michael B. Pursley[1]
Dept. of Elec. and Comp. Eng., Clemson University, Clemson, SC, USA 29634-0915

Abstract — **The purpose of this paper is to examine the effects of the power control technique, the coding, and the interleaving depth on the performance of code-division multiple-access (CDMA) systems with different chip rates and rake receivers with different numbers of taps. We consider the implications of the results for the support of voice and data services in a cellular CDMA system.**

Direct-sequence (DS) spread spectrum CDMA is a leading candidate for use in mobile cellular systems and personal communication systems. Important characteristics of a CDMA system include the chip rate, the power-control technique, the forward error-correction (FEC) code, the depth of code-symbol interleaving, and the number of taps in the rake receiver. Though the development of emerging CDMA systems has focused primarily on the support of voice communications, the increasing demand for packet data services points to the need for systems that efficiently support both voice and data traffic.

The effect of *near-far interference* [1] in a cellular CDMA system can be reduced by adapting the power of each transmitter to the channel response or the interference environment. In a full-duplex voice connection, the forward (base station to mobile) link can serve in part as a feedback channel for power-control commands from the base station. This is referred to as *closed-loop power control.* We consider the effect of feedback delay on the performance of a CDMA system with closed-loop power control, and the effect is examined for several channels and for systems of different chip rates and different numbers of taps in the rake receiver.

In contrast, data traffic on the reverse link is likely to be bursty. In many instances, it is not practical to provide feedback during the transmission of a data packet. As a result, the mobile must determine *a priori* the appropriate power level for the entire packet. This is referred to as *average power control.* Some compensation for rapid fading can be obtained by using FEC coding together with interleaving as a form of time diversity. We consider the effect of coding and interleaving on the performance of a CDMA system with average power control, and we examine its effectiveness for different channels and for systems with different chip rates and different numbers of rake receiver taps.

The ability of the receiver to resolve multipath components of the received signal depends on the chip rate of the DS signal. Our channel model reflects this phenomenon and allows for tractable analysis of receiver performance. Each channel is a special case of the Gaussian wide-sense-stationary uncorrelated-scattering channel, and it is described in detail in [2]. A closed-from expression is derived in [3] for the probability of error at the input to the decoder for a CDMA system that employs closed-loop power control and rake reception. The performance of the system is assumed to be limited by multiple-access interference, and the composite interference is modeled as additive white Gaussian noise. The result is employed here to determine two quantities of interest – the spectral efficiency [3] of the cell and the average signal-to-interference ratio (SIR) that is required to achieve a target error probability. For a given traffic mix and collection of channels, the relationship between required SIR and spectral efficiency varies with the chip rate, the power-control feedback delay, the FEC code, and the interleaving depth.

In contrast to the result in [3], no closed-form expression can be obtained for the probability of error at the *output* of the decoder. Chernoff-bound techniques can be used to evaluate the performance of coding and finite interleaving depth, but the bounds fail to converge for many circumstances of interest in mobile communications. Thus, we employ simulations to examine the effect of persistent fading on the performance of CDMA systems with average power control, convolutional coding, and interleaving. The receiver that is considered employs Viterbi decoding.

We have obtained numerical results for many circumstances that are encountered in mobile communications. It is shown that the performance of a CDMA system with a low chip rate is more sensitive to channel and system parameters than is a CDMA system with a high chip rate. The rake receiver is necessary for adequate performance of a low-chip-rate system under many more circumstances than for a high-chip-rate system. The best choice of chip rate for a system with closed-loop power control depends on the ratio of the maximum Doppler spread to the feedback delay, and it also depends on the allowable number of taps. For a CDMA system employing average power control, coding, and interleaving, a high chip rate provides performance superior to that of a low chip rate in most circumstances.

[1] This research was supported in part by the Holcombe Endowment at Clemson University and in part by the Army Research Office under grants DAAH04-94-G-0154 and DAAH04-93-G-0253. J. H. Gass is the recipient of a National Science Foundation Graduate Research Fellowship.

References

[1] A. J. Viterbi, A. M. Viterbi, and E. Zehavi, "Performance of power-controlled wideband terrestrial digital communication," *IEEE Trans. Commun.*, vol. COM-41, pp. 559–569, Apr. 1993.

[2] D. L. Noneaker and M. B. Pursley, "On the chip rate of CDMA systems with doubly selective fading and rake reception," *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 853–861, June 1994.

[3] J. Gass, D. L. Noneaker, and M. B. Pursley, "On the capacity of a power-controlled mobile cellular CDMA system," *Proc. 1995 Vehicular Technology Society Conference*, paper E4.7, July 1995.

Successive Cancellation in Fading Multipath CDMA Channels[1]

MAHESH K. VARANASI

ECE Dept, University of Colorado, Boulder, Colorado 80309. *varanasi@spot.colorado.edu*

Summary– Given an ordered J group partition of the K simultaneously transmitting users of a CDMA channel, a sequential group detector consists of J group detectors that are connected sequentially. The j^{th} group detector uses the decisions from the previous $j-1$ group detectors and cancels the inter-user interference from those users before it makes joint decisions for the j^{th} group. This successive interference cancellation scheme was introduced in [1] for the Gaussian CDMA channel. This paper consists of extending that idea to the Frequency-Selective Rayleigh Fading (FSRF) CDMA channel (described in [2]). The two group detectors (I and II) introduced in [2] for the FSRF-CDMA channel are considered as the basic building blocks. The resulting sequential group detectors can be regarded as members of two distinct classes (each class parametrized by the ordered partition) of multiuser detectors that satisfy a wide range of complexity constraints. In particular, each of the two sequential group detectors has a time complexity per symbol (TCS) of $O(\sum_{j=1}^{J} M^{K_j})$ for M-ary signalling, where K_j is the j^{th} group size. The optimum multiuser detector has a fixed TCS of $O(M^K)$. The simplest case corresponds to the degenerate ordered partitions consisting of K groups of size 1 each. For this choice, the two sequential group detectors reduce to two distinct decorrelating decision feedback detectors. These special cases can be seen as two distinct generalizations (to the FSRF-CDMA channel) of the multiuser detector by the same name for the Gaussian channel that was proposed in [3] and for which the analysis can be found in [1]. A succinct indicator of the average BER over high SNR regions for the FSRF-CDMA channel is defined via the asymptotic efficiency in [2]. In this work, upper and lower bounds on the asymptotic efficiency for the two sequential group detectors are derived. Minimax criteria under which these detectors are optimal are specified. The following numerical example illustrates the vast improvements achievable by the sequential group detector based on the group detector II of [2] over the detector proposed in [4].

Numerical Example– Consider the six user direct-sequence spread-spectrum system employing Gold sequences of length 31 of [2] operating in a fading multipath environment with four paths for each user. The users are numbered according to decreasing average power ratios (with respect to the minimum power) given in order as $[10.0, 2.5, 2.0, 1.5, 1.25, 1.0]$. Suppose that the performance of a single-user RAKE receiver for the last (weakest) user in the hypothetical single-user scenario, where all the other users are absent, is considered acceptable. Equivalently, the effective SNR (ESNR) to minimum actual SNR ratio (henceforth referred to relative ESNR) for every user has to be no less than 1. The linear suboptimum detector of [4] results in relative ESNRs for the six users given in order as $[1.40, 0.37, 0.53, 0.21, 0.18, 0.26]$. As for sequential group detection, it turns out that the decorrelating decision-feedback detector suffices. The resulting upper bounds for the relative ESNRs for the six users are $[1.40, 1.22, 1.12, 1.06, 1.09, 1.0]$ and the lower bounds are $[1.40, 1.22, 1.12, 1.06, 1.06, 1.0]$. The minimum specification is met. Moreover, note that the upper and lower bounds coincide for all but the fifth user. Now suppose that the power ratios are made less disparate by reducing them for the odd-numbered users so that the power distribution for the six users is $[2.5, 2.5, 1.5, 1.5, 1.0, 1.0]$. The relative ESNRs for the six users for the linear suboptimum detector are $[0.35, 0.37, 0.39, 0.21, 0.15, 0.26]$. The upper bounds for the relative ESNRs for the six users for the decorrelating decision feedback detector are $[0.35, 1.22, 0.84, 1.06, 0.87, 1.0]$ with a lower bound of 0.35 for all the users. The wide gap between the upper and lower bounds for users 2 to 6 suggests that error propagation is a severe problem inspite of the users being arranged in decreasing order of powers. However, consider a sequential group detector with an ordered group partition $\{1,2\}\{3,4\}\{5,6\}$ consisting of three groups of size two each. In this case, relative ESNRs for the six users are equal to $[1.16, 1.22, 1.115, 1.114, 1.0, 1.0]$ (the upper and lower bounds coincide for every user in this case). The minimum requirement is thus met. Moreover, error propagation has little effect on the performance. The lower complexity of the sequential group detector however is achieved at the expense of approximately 3 dB loss for the strongest users and nearly 1.25 dB for the users of intermediate power relative to the optimum detector.

[1]This work was supported by NSF Grant NCR-9406069.

References

[1] M. K. Varanasi, "Group Detection for Synchronous Gaussian Code-Division Multiple-Access Channels," *IEEE Trans Inform. Th.*, **41**:4 (July, 1995).

[2] M. K. Varanasi, "Group Detection for Synchronous CDMA Communication Over Frequency-Selective Fading Channels," *Proc 31st Annual Allerton Conf on Communication, Control, and Computing*, Sep 29-Oct 1, 1993, pp. 849-858.

[3] A. Duel-Hallen, "Decorrelating Decision Feedback Multiuser Detector for Synchronous Code-Division Multiple-Access Channel," *IEEE Trans Commun.* **COM-41**:2, 285-290, February 1993.

[4] Z. Zvonar and D. Brady, "Suboptimum Multiuser Detector for Frequency-Selective Rayleigh Fading Synchronous CDMA Channels," *IEEE Trans Commun.* **43**:2/3/4 (Feb/Mar/April 1995) pp. 154-157.

Adaptive Interference Suppression for DS-CDMA with Impulsive Noise

Narayan B. Mandayam
Wireless Information Network Laboratory (WINLAB), Dept. of ECE, Rutgers University, Piscataway, NJ 08901

Abstract — **We develop an adaptive interference suppression scheme for DS-CDMA systems in the presence of impulsive noise. This scheme is realized by deriving an IPA based stochastic gradient algorithm that minimizes the average probability of error for such systems. The resulting detector outperforms the conventional matched filter detector for such systems.**

I. Introduction

Recently, there has been much work done on deriving adaptive linear detectors for DS-CDMA systems corrupted by additive Gaussian noise ([1] and the references within). However, such communication systems are often interfered with by noises other than the classical white Gaussian noise, and in here we consider DS-CDMA systems corrupted by natural impulsive noise sources, such as those found in low-frequency atmospheric channels, and for channels corrupted by man-made impulsive sources such as those occurring in urban or military radio networks. The conventional correlation receiver has been shown to experience a degradation in performance in impulsive noise (relative to the Gaussian noise model) even when the user's codes are chosen to have low cross-correlations [2]. On the other hand, when the multiple access interference dominates, the linear correlator in the impulsive noise channel is not near near-far resistant (similar to the Gaussian Channel). In this paper, we develop an adaptive linear detector, for such impulsive noise channels, which directly minimizes the average probability of bit-error. The approach is similar to that used in [3], where we develop an infinitesimal perturbation analysis (IPA) based stochastic gradient algorithm for achieving minimum probability of bit-error. The adaptive interference rejection scheme is shown to have a very simple recursive structure (thereby by allowing easy implementation), and the conditions for convergence of this algorithm are presented.

II. System Description

We will consider a K-user DS-CDMA system where the received signal in the channel is the sum of the transmissions due to the K users and additive channel noise. The received signal due to the transmission of the k^{th} user at any receiver is given as

$$\rho_k(t) = \sqrt{2P_k} \sum_{i=-\infty}^{\infty} b_{i,k}\, a_k(t - iT - \tau_k)\cos(\omega_c t + \phi_k),$$

where $b_{i,k} \in \{-1, +1\}$ is the i^{th} bit of the k^{th} user, T is the bit-period, P_k, ϕ_k, and τ_k are the power, carrier phase and delay of the k^{th} user, respectively. The carrier frequency is denoted by ω_c, and $a_k(t)$ is the spreading waveform of the k^{th} user. The received signal in the channel due to the K users and additive noise is given as

$$r(t) = \sum_{k=1}^{K} \rho_k(t) + \eta(t),$$

where $\eta(t)$ is assumed to be the additive impulsive noise that is characterized by the first order probability density function

$$f_{\eta[n]}(x) = (1-\epsilon) f_n(x) + \epsilon f_I(x),$$

where $\epsilon \in [0, 1]$, and f_n and f_I are pdf's [2]. The *nominal* density function f_n is usually taken to be a Gaussian density representing the background noise. The impulsive component of the noise is represented by f_I which is taken to be more heavily tailed than f_n. The above model represents an approximation to the canonical Class A interference model studied by Middleton and Spaulding.

III. Adaptive Interference Suppression

In [2] a conventional correlator was used for detection of the desired user's bits. We are interested in finding the best set of correlation sequences $\underline{h}$, such that the average probability of bit-error is minimized. Following the approach in [3], we develop an IPA based stochastic gradient algorithm that yields the optimum linear detector for this system. Therefore, we require the vector $\underline{h}^*$ such that

$$\underline{h}^* = \arg\{\min_{\underline{h}} \bar{P}_e\}. \tag{1}$$

We now adaptively update the vector $\underline{h}$, using a stochastic algorithm given as

$$\underline{h}_{i+1} = \underline{h}_i - \gamma_i \nabla_{\underline{h}_i} P_e(\underline{h}_i, \underline{s}_i), \tag{2}$$

where the gradient $\nabla_{\underline{h}} P_e$ is estimated using infinitesimal perturbation analysis, and $\underline{s}_i$ is the vector of transmitted signals for the i^{th} iteration, i.e., the i^{th} bit period. It is shown that the algorithm allows a very simple recursive structure owing to the analyticity of the Q function. The conditions for convergence of the algorithm are presented as well. The performance of the adaptive linear detector is seen to be uniformly better than that of the linear correlator even under the extreme cases when either the multiple access interference or the impulsive noise is dominant.

References

[1] S. Verdu, "Adaptive Multiuser Detection", *Proceedings of IEEE International Symposium on Spread Spectrum Theory and Applications*, Oulu, Finland, July 1994.

[2] B. Aazhang, and H. V. Poor, "Performance of DS/SSMA Communications in Impulsive Channels-Part I: Linear Correlation Receivers" *IEEE Trans. Commun.* vol. 35, no. 11, Nov., 1987, pp. 1179-1188.

[3] N. B. Mandayam, and B. Aazhang, "An Adaptive Multiuser Interference Rejection Algorithm for Direct-Sequence Code Division Multiple Access", *Proceedings of the International Symposium on Information Theory* , Trondheim, Norway, June 1994.

Error-and-Erasure Decoding of Convolutional Coded DS/SSMA Communications in AWGN and Rayleigh Fading Channels

Jin Man Kwon and Sang Wu Kim
Depart of Electrical Engineering
Korea Advanced Institute of Science and Technology
Taejon 305-701, Korea
email: swkim@itl.kaist.ac.kr

Abstract — **We examine the performance of convolutional coded DS/SSMA communication system with error-and-erasure decoding in AWGN and multipath Rayleigh fading channel. The demodulator makes a three-level-decision {-1,1,?} based on the channel state information(CSI), where ? represents an erasure. The CSIs considered are the matched filter output and the fading amplitude. The optimum decision threshold that minimizes BER is found to be almost equal to the threshold that maximizes the channel cut-off rate R_0. A simple parallel decision scheme is proposed to give a performance which is very close to the optimum decision scheme. The performance improvement made by using the CSI is investigated.**

I. Introduction

It is well known that soft decision decoding requires 2-3dB less in signal-to-noise ratio over the hard decision decoding in AWGN channels [1]. However, soft decision decoding requires real arithmetic operations, which are much more complex than binary operations involved in hard decision decoding. Clark and Cain has pointed out that erasing unreliable symbols based on channel state information (CSI) and performing error-and-erasure decoding is an effective method to provide a good trade off between system performance and implementation complexity[2]. In this paper we analyze the performance of convolutional coded DS/SSMA communication systems employing error-and-erasure decoding and binary PSK modulation with several demodulation schemes.

II. Demodulation Schemes

We consider several demodulation schemes that make a three-level-decision {-1,1,?} based on the CSI. In AWGN channel, we use the matched filter output as a CSI, which is most convenient, useful, and easy to get. If the absolute value of the matched filter output is larger than a threshold, the demodulator makes a decision {-1,1} based on the matched filter output, otherwise the demodulator erases the corresponding symbol. In multipath Rayleigh fading channel we use the matched filter output and/or the fading amplitude as a CSI. For the case of demodulator using only the fading amplitude, we consider a demodulator that makes a hard decision if the fading amplitude is larger than a threshold, otherwise erases the symbol. We assume the fading amplitude information is available at the demodulator. We also consider a demodulator that uses both the matched filter output and the fading amplitude. In this case if the fading amplitude is below a threshold(Ω_c) or the matched filter output is below a threshold(Ω), the demodulator erases the symbol, otherwise makes a hard decision based on the matched filter output.

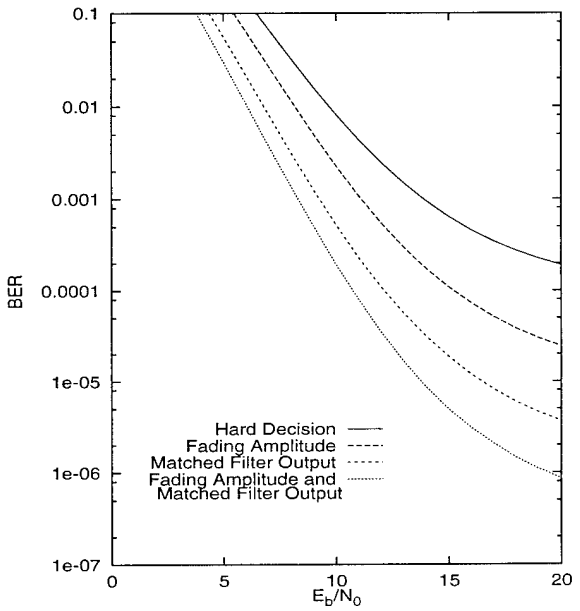


Fig. 1: BER vs. E_b/N_0 : code rate = 1/2, constraint length = 7, convolutional code, number of user = 30, 128 *chips/coded bit*, Rayleigh fading channel($\sigma^2 = 1/2$)

III. Discussions

We have investigated the optimum erasure threshold that minimizes BER. It is found that the erasure threshold that maximizes the channel cut-off rate R_0 is almost optimal and the optimum erasure threshold increases as the traffic increases. Based on this observation, we propose a simple parallel decision scheme that changes the erasure threshold according to the channel traffic. We found that the parallel scheme yields a performance close to that with the optimum decision scheme. We have also examined how much the performance improvement can be made by using the CSIs in Rayleigh fading channel. Fig.1 shows the BERs with different CSIs. We can see that the erasure based on the fading amplitude information alone gives a higher BER than with matched filter output alone. However, when the fading amplitude information is combined with the matched filter output, the fading amplitude information gives a gain of 1.0-2.5dB in E_b/N_0 at the BER of 10^{-3}, and even a higher gain can be obtained for lower BER.

References

[1] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, NY: John Wiley and Sons, 1965.

[2] C. C. Clark and J.B. Cain, *Error Correcting Coding for Digital Communications*, pp. 354-357 Plenum Press, New York, 1981.

Diversity Performance of $\frac{\pi}{4}$ - DQPSK for the Reverse Link in a DS-CDMA Cellular System

Leonard E. Miller and Jhong S. Lee

J. S. Lee Associates, Inc., Rockville, Maryland USA

Abstract — **The performance is evaluated for a hypothetical CDMA digital cellular telephone system whose reverse link uses $\pi/4$-DQPSK modulation and equal gain RAKE diversity combining. The results are shown numerically in comparison with those for Qualcomm's (IS-95) CDMA cellular system, which uses 64-ary orthogonal modulation on the reverse link.**

INTRODUCTION

The mobile-to-base (reverse) link of the North American IS-95 DS-CDMA cellular system employs an M-ary orthogonal modulation using Walsh-Hadamard sequences with QPSK phase coding with M = 64. For lack of the carrier phase reference-providing pilot signals, which are used for the forward (base-to-mobile) links, the system employs noncoherent demodulation for the reverse links for each of the L-path diversity receptions in its RAKE system.

In this paper we suggest and investigate an alternative scheme for the reverse link modulation and multipath receptions: the information sequence is to be $\pi/4-$DQPSK modulated after inserting the DS-CDMA spreading sequence in the I and Q channels prior to the pulse shaping and summation, and the demodulation is done with partially coherent differential detection for each multipath component before diversity combining.

We show a closed form error probability expression for the $\pi/4-$DQPSK reverse link with L independent multipath diversity receptions in Rayleigh fading and CDMA interference. The results are evaluated with capacity and processing gain values as parameters. The exact closed form expression for the performance of the system is based on the authors' previous work [1].

SUMMARY OF ANALYTICAL RESULTS

Resolution of multipath signal components separated in time delay by more than the chip period of the DS-CDMA SS sequence is possible, and the paths can be combined to provide diversity. The unconditional probability of error for the reception of one of L fading multipath components is found to be

$$p_2(e)=\frac{1}{2}\left[1-\frac{\rho_L\cos(\pi/4)}{\sqrt{\left[\rho_L\cos(\pi/4)\right]^2+\frac{1}{4}+\rho_L}}\right] \qquad (1)$$

where

$$\rho_L \triangleq \frac{\overline{\mathrm{E}}_b/L\mathcal{N}_0}{1+(\overline{\mathrm{E}}_b/\mathcal{N}_0)\cdot\frac{Md}{PG\cdot FG_s}} \qquad (2)$$

and where M is the number of multiple access users, PG is the spread spectrum processing gain, F is the frequency re-use factor, d is a voice activity factor, and G_s is the sector antenna gain.

For a receiver combining the $L>1$ paths, assuming independent fading in the multiple paths, the probability of error can be shown to be

$$P_L(e)=\left[p_2(e)\right]^L\cdot\sum_{k=0}^{L-1}\binom{L-1+k}{k}\left[1-p_2(e)\right]^k. \qquad (3)$$

The values of d, M, F, PG, and G_s determining the amount of interference can be traded off to achieve the desired system performance. The corresponding error-expression for the diversity combining of M-ary orthogonal signals is [2]

$$P_L(e)=\frac{\mathrm{M}/2}{\mathrm{M}-1}\cdot\frac{1}{\Gamma(L)}\sum_{n=1}^{\mathrm{M}-1}\binom{\mathrm{M}-1}{n}\frac{(-1)^{n-1}}{(1+n+n\rho_L)^L}\times\sum_{k=0}^{n(L-1)}\beta_k(n)\,\Gamma(L+k)\left(\frac{1+\rho_L}{1+n+n\rho_L}\right)^k, \qquad (4)$$

in which the value of $\beta_k(n)$ is the coefficient of x^k in the expansion

$$\left(\sum_{k=0}^{L-1}\frac{x^k}{k!}\right)^n=\sum_{k=0}^{n(L-1)}\beta_k(n)\,x^k. \qquad (5)$$

It is found that, in the fading environment, the proposed reverse link modulation, which is less complex than the IS-95 64-ary orthogonal modulation, actually outperforms the latter when $L\le 2$ paths are combined. We consider examples of capacity and processing gain, and implications for the error correcting codes needed for meeting operational requirements for voice or data.

REFERENCES

[1] L. E. Miller and J. S. Lee, "New Closed Form Expressions for Differentially Detected $\pi/4-$DQPSK System Performance in AWGN and Rayleigh Fading," *Proc. 1994 IEEE Int'l Symp. on Info. Thy.*, p. 89, Trondheim, Norway, 27 June-1 July 1.

[2] J. G. Proakis, *Digital Communications*, McGraw-Hill, 1983.

On the Expected Value of the Average Interference Parameter of Code-Sequences in CDMA Systems

Hans Dieter Schotten

Institut für Elektrische Nachrichtentechnik, RWTH Aachen, Melatener Strasse 23, D-52056 Aachen, Germany
Phone: +49-241-807680, Fax : +49-241-8888196, EMail: schotten@ient.rwth-aachen.de

Abstract — **In this paper, the average interference parameter (AIP) of polyphase code-sequences in DS-CDMA systems is investigated. The expected value and the variance of the AIP for randomly chosen cyclic shifts of the code-sequences are derived.**

I. Introduction

The performance of direct-sequence code-division multiple-access (DS-CDMA) systems depends on the correlation properties of the used code-sequences. The most common criteria to describe the correlation behavior are the periodic peak correlation parameter, describing the worst-case behavior, and the average interference parameter (AIP) on which the signal-to-noise ratio depends. Usually, families of code-sequences for these systems are constructed considering the periodic peak correlation parameter. In a second step, cyclic shifts of these sequences which result in an optimum AIP are sought. (The periodic peak correlation parameter remains unchanged if the sequences are cyclically shifted.) Since not all combinations of shifts can be examined, simplified search methods, e.g. based on the sidelobe energy, are applied. To compare the performance of these techniques and to derive bounds on the achievable AIP, the expected value and the variance of the AIP for randomly chosen shifts are needed. For this reason, these values will be derived for some of the most important families of code-sequences.

II. Investigated Families

We consider families $\mathcal{F} = \{S_k(n) \mid 1 \le k \le K\}$ consisting of K sequences of length N $(0 \le n \le N-1)$. The elements of the sequences are roots of unity. The aperiodic correlation function is defined by $C_{SR}(m) = \sum_{n=0}^{N-1-m} S^*(n)R(n+m)$ $(m \ge 0)$ and the periodic correlation function by $\tilde{C}_{SR}(m) = \sum_{n=0}^{N-1} S^*(n)R(n+m \bmod N)$. The maximum magnitude of the periodic crosscorrelation values and the autocorrelation sidelobes is the periodic peak correlation parameter $\tilde{\theta}$.

Three types of large families of code-sequences have been investigated: *Prime-phase code-sequences* (e.g. Gold-, Kasami-, and Kumar-Moreno-families) are constructed in the Galois-field $GF(p^r)$ using an additive character [1]. Because of their practical importance, *quadriphase code-sequences* and other prime-power-phase sequences are considered. The construction of these sequences in Galois-rings $GR(p^a, r)$ is described in [2]. The *third family* is constructed by multiplying all sequences of families of type 1 or 2 with $\exp(j2\pi kn/N)$ with $k = 0...N-1$.

III. Interference Parameters

We consider an asynchronous phase shift keying DS-CDMA system for K users. The signal-to-noise ratio of these systems can be expressed in terms of the total interference parameter TIP [3] which is defined as (at receiver i)

$$\mathrm{TIP}_i = 1/(3N^3) \sum_{k \ne i} 2\mu_{S_iS_k}(0) + \mu_{S_iS_k}(1),$$

where $\mu_{S_iS_k}(l) = Re\left(\sum_{\nu=1-N}^{N-1} C^*_{S_kS_k}(\nu) C_{S_iS_i}(\nu+l)\right)$. Usually, the average interference parameter AIP $= 2\mu(0) + \mu(1)$ with $\mu(t) = 1/(K(K-1)) \sum_{i=1}^{K} \sum_{k=1,k\ne i}^{K} \mu_{S_iS_k}(t)$ is used as measure for the average system performance. To simplify the notation, we define the sum $\mathcal{S}(\mathcal{F},\nu) = \sum_{S\in\mathcal{F}} \tilde{C}_{SS}(\nu)$.

IV. Expected Values of the AIP

Since the AIP depends on the cyclic shift of the code-sequences, we suppose that the cyclic phase of the sequences is picked at random with each of the shifts being equally likely to be chosen [4]. Then, the expected value of $\mu(0)$ can be derived: $E[\mu(0)] =$

$$N^2 + \frac{2}{N^2M(M-1)} \sum_{\nu=1}^{N-1} (N-\nu)^2 \left[\mathcal{S}(\mathcal{F},\nu)\mathcal{S}^*(\mathcal{F},\nu) - \overline{\mathcal{S}}(\mathcal{F},\nu,\nu)\right].$$

The expected value of $\mu(1)$ and the variance of the AIP can be expressed in terms of other sums $(\overline{\mathcal{S}}(\mathcal{F},\nu,\nu+l) = \sum_{S\in\mathcal{F}} \tilde{C}^*_{SS}(\nu)\tilde{C}_{SS}(\nu+l))$. These sums have been derived for all families described in section II, too.

V. Results

We have investigated families of size $M \approx (N+1)^t$ with $t \ge 1$. Typical periodic peak correlation parameters $\tilde{\theta}$ are $\tilde{\theta} \le 2t\sqrt{N+1}+1$ or $\tilde{\theta} \le 2^{t-1/2}\sqrt{N+1}+1$ if binary sequences are considered or $\tilde{\theta} \le t\sqrt{N+1}+1$ for sequences with larger phase alphabet. In both cases, we found $\mathrm{E}(\mu(1)) << \mathrm{E}(\mu(0))$ and hence

$$\mathrm{E(AIP)} \approx 2\mathrm{E}(\mu(0)) = 2N^2 - \frac{2}{3}\frac{(2N-1)(N-1)}{K-1}.$$

Obviously, E[AIP] does not depend on the size of the phase-alphabet and is nearly independent of the size of the family. For the described linear families, the $E[AIP]$ becomes $2N^2$ - the expected value for random sequences - if N tends to infinity. For the variance, we found noticeable differences depending on the investigated families. Using these results, the known numerical results on the AIP of linear code-sequences for different selection criteria of cyclic shifts (e.g. LSE/AO, MSE/AO) can be explained. Moreover, bounds on the achievable AIP for all linear families are derived.

References

[1] Kumar P.V., Moreno O.: Prime-Phase sequences with periodic correlation properties better than binary sequences. IEEE Trans. IT-37 (1991), 603 – 616.

[2] Kumar P.V., Helleseth T., Calderbank A.R., Hammons. A. J.: Large families of quaternary sequences with low correlation. Proc. IEEE Int. Symp. on Inform. Theory, Trondheim, 1994. 71.

[3] Pursley M.B.: Performance evaluation for phase-coded spread-spectrum multiple-access communication – part I: System analysis. IEEE Trans. Commun. COM-25 (1977), 795–799.

[4] Sarwate D.V.: Mean-square correlation of shift-register sequences. Proc. IEE 131 F (1984), 101–106.

Synchronous Frequency-Hopped CDMA Using Wavelets

Fred Daneshgaran[a] and Marina Mondin[b]

[a] Elec. & Comp. Eng. Dept., California State University, Los Angeles, CA (USA)
[b] Dip. di Elettronica, Politecnico, Torino (ITALY)[1]

Abstract — **In this paper we present the Wavelet Orthogonal Frequency Division Multiplexing (WOFDM) that in conjunction with Frequency Hopping can be used for Synchronous Code Division Multiple Access (FH/S-CDMA). A Low Probability of Intercept (LPI) modulation scheme based on a pseudo random selection of basis functions for modulation spanning the same frequency channel is also described.**

I. Introduction

In [1] the use of scaling functions and wavelets, multiplicity-M wavelets and wavelet packets to modulate different information signals on adjacent channels with overlapping spectra was proposed. In this paper we demonstrate an application of this technique for multiple access communication [2].

The envisioned Frequency-Hopped Synchronous Code Division Multiple Access (FH/S-CDMA) scheme is for a multipoint to point fully synchronized communication. In this environment, wavelets provides a great flexibility in controlling the data rate and hence the power, making the proposed CDMA scheme inherently adaptive.

II. Orthogonal Frequency Channelization

The basic techniques to subdivide a given frequency band into orthogonal subchannels spanned by basis functions derived from the scaling functions and wavelets are described in [1]. This defines the WOFDM modulation scheme, which possesses the following characteristics: (1) orthogonal channels are spanned by translates of a single envelope function. The translation step size is directly related to the BandWidth (BW) of the subchannel; (2) the channels overlap in frequency but remain orthogonal with proper synchronization; (3) there is great flexibility in how the available BW is channelized, and this channelization has a tree structure. It is therefore possible to accommodate variable rate data modulation by routing data to different nodes of the tree structure that have different data rate capacities; (4) this switching induces some transient InterSymbol Interference (ISI).

III. FH/S-CDMA with Wavelets

The described WOFDM scheme can be employed for multiple access communications using frequency hopping, where a given information sequence can be hopped by *routing* the data in this sequence to the input of the filter generating the desired frequency channel.

The key features of this scheme are: (1) there is no need for a programmable frequency synthesizer; (2) the size of the hopping BW is related to the information data rate. Changes in this data rate are accommodated by routing the data to the appropriate internal nodes of the tree structure. The protocol for how the variable data rate is to be accommodated should be established from the outset and programmed into the operation of the connection network; (3) multicarrier modulation is possible with the proposed technique. Note that in the proposed scheme a high degree of security may be afforded to the communication system using a relatively small number of orthogonal frequency channels due to the combinatorial power of the connection network; (4) the hopping rate relative to the data rate is directly controlled by the rate at which the connection machine changes its patterns relative to the maximum rate each channel can be utilized; (5) aside from carrier synchronization needed to perform the down conversion, clock synchronization and PN code synchronization are needed for proper operation.

Direct Sequence (DS) spectrum spreading could be incorporated into the design by forming the product between the spreading code and the information sequence *prior* to modulating the wavelet filters. In this process what controls the BW of each hopping channel is the PN code rate used for the DS component.

IV. Low Probability of Intercept

We previously noted that the switching of frequency channels employed in order to accommodate variations in source data rate causes transient ISI [3]. This transient ISI can be used to introduce a novel LPI modulation scheme. More specifically, suppose a given frequency band spanned by a shift orthogonal function is channelized in a variety of ways. Each such channelization corresponds to a different distribution of dimensions in the time-frequency plane. A modulator can be state dependent and use a given distribution of dimensions for modulation in accordance with a PN code known to the transmitter and receiver. Suppose the modulator state varies rapidly so that a given distribution of dimensions is not used for more that a few symboling intervals. An unintended receiver with perfect knowledge of the waveforms used by the transmitter and perfect knowledge of symbol timing may still be unable to recover the symbols since it perceives a sequence with very high randomly fluctuating ISI [3].

The above procedure can be embedded in the FH/S-CDMA, and the multicarrier modulation scheme proposed here, and two PN codes could be used by each information source, one controlling the operation of the switching network used to frequency hop the spectrum of the transmitted signal, and the other used to select which distribution of dimensions in the time-frequency plane is to be used by the modulator.

[1] This work was partially supported by M.U.R.S.T.

References

[1] F. Daneshgaran, M. Mondin, "Wavelets and Scaling Functions as Envelope Waveforms for Modulation," *IEEE-SP Int. Symp. on Time-Frequency and Time-Scale Analysis*, Philadelphia (USA), October 25-28, 1994.

[2] F. Daneshgaran, M. Mondin, "Orthogonal Frequency Division Multiplexing and its Application to Frequency-Hopped CDMA," *Proc. of the 29th CISS*, Philadelphia, MA (USA), March 1995.

[3] F. Daneshgaran, M. Mondin, "Multidimensional signaling for bandlimited channels," *Proc. of ISIT 95*, Whistler, B.C., (Canada), September 17-22, 1995.

Near-Orthogonal Coding for Spread Spectrum and Error Correction

Karen W. Halford, Yaron Rozenbaum[1], and Maïté Brandt-Pearce[2]
Dept. of Electrical Engineering, Univ. of Virginia, Charlottesville, VA 22903

Since the spreading operation and error correction must share the bandwidth available in a CDMA system, it is appropriate to approach these problems jointly. Several papers have addressed this problem with promising results [1, 2]. Hui has shown that under certain assumptions, the system performs better when more bandwidth is devoted to error correction [2]. Giallorenzi [3] shows that combining error correction decoding and multiuser detection significantly improves system performance. We extend this research by considering not only simultaneous *despreading* and *decoding*, but also the simultaneous *encoding* and *spreading*.

We consider a coded asynchronous CDMA system over an AWGN channel with constant information rate, R_I. Each user's transmission rate is $R_{tx} = R_I \cdot Q \cdot N$ where $1/Q$ is the convolutional code rate and N is the spreading factor. The receiver matches to each signature sequence and performs maximum likelihood sequence detection.

For fixed R_I and R_{tx} we optimize the Asymptotic Multiuser Coding Gain (AMCG) with respect to Q and N. The AMCG relates the energy gain for high SNR to the single user uncoded antipodal system, i.e. η in the expression $\mathcal{P}_e = \mathcal{Q}\{\sqrt{2E_b/N_o\eta}\}$ where $\mathcal{Q}(\cdot)$ is the Marcum-Q function, E_b is the information bit energy, and N_o is the one sided noise density. We have extended this measure derived in [3] which considers $Q = 2$ and fixed N to arbitrary Q and N.

The probability of error for the k^{th} user can be bounded by $\mathcal{P}\{\eta_k(\bar{e})=\eta_{k,min}\}\mathcal{Q}\{\sqrt{(2E_{bk}/N_o)\eta_k(\bar{e})}\} \leq \mathcal{P}_e(k) \leq \sum_{\bar{D}\in\mathcal{C}}\sum_{\bar{e}}\mathcal{P}\{\bar{D}\}\mathcal{Q}\{\sqrt{(2E_{bk}/N_o)\eta_k(\bar{e})}\}$ where $\mathcal{C}$ is the codebook , $\bar{e}$ is any valid error sequence for the codeword $\bar{D}$, $\eta_k(\bar{e})$ is the energy gain of user k when the error event $\bar{e}$ occurs, $\eta_{k,min} = \min_{\bar{e}}\{\eta_k(\bar{e})\}$ and E_{bk} is the energy of user k. For high SNR, $\eta_{k,min}$ will dominate, hence it is the AMCG. In the 2 user case the AMCG is bounded by $\min\left\{f\left(\sqrt{E_2/E_1}, d_f, \xi\right), d_f/Q\right\} \leq \eta_{k,min} \leq \eta_k(\bar{e})$ for some valid $\bar{e}$ where E_1 and E_2 are the two user's energies, ξ is the sum of the magnitude of the two partial crosscorrelation, d_f is the free distance of the convolutional code and $f(\sqrt{E_2/E_1}, d_f, \xi) = 1/2[d_f(1+E_2/E_1) - 2\xi\sqrt{E_2/E_1}(d_f+1)]$.

These bounds for two users are computed for 3 different length M-sequences in Fig.1 (a) and (b). Plots (a) and (b) represent $R_{tx} = 32R_I$ and $R_{tx} = 64R_I$ respectively. These bounds were computed using the maximum partial crosscorrelations over all delays between the two users. Fig.1 (a) shows that when the partial crosscorrelations approach 1, the system with the lower coding rate may not show any improvement. However, when the crosscorrelations are high but less than 1, as in Fig.1 (b), the lower rate codes perform as well as the single user detector, i.e. ACMG=ACG=d_f/Q for all E_2/E_1.

Since high crosscorrelations between signature sequences can prevent expected coding gains, we propose spreading and despreading in the frequency domain which was considered for optical systems in [4]. Because delays appear as phase

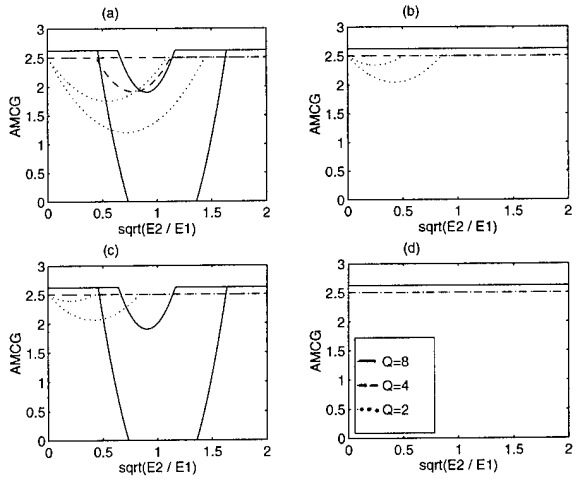


Fig. 1: In (a), (c), and (d) for Q=8,4,2: d_f=21, 10, and 5, and N=4, 8 and 16, and for (b) d_f=21,10, and 5 and N=8,16 and 32. The crosscorrelations ξ for Q=8,4,2 are (a): 1.0, 0.71, 0.6, (b): 0.71, 0.6, 0.35, (c): 1.0, .43, .35 and (d): 0,0,0.

factors in the frequency domain, we can find sequences for an asynchronous system that are both short and have sufficiently low crosscorrelations to allow coding gains.

In this system the encoder multiplies each encoded bit by the inverse FFT of a signature sequence that has low crosscorrelation in the frequency domain. The decoder matches the Fourier transform of each received symbol to the signature sequence and sends the output to a maximum likelihood sequence decoder. In Fig.1 (c) and (d), we show the bounds for the AMCG for two frequency domain codes with constant rates, $R_{tx} = 32R_I$. Fig.1 (c) and (d) are computed assuming worst case interference for frequency domain M-sequences and Hadamard codes, respectively. These codes show a great improvement over the time domain codes, and, in fact, the Hadamard sequence achieves the single user ACG for all E_2/E_1. Although the Hadamard sequences outperform the M-sequences, there are fewer available Hadamard sequences for a given sequence length.

The asymptotic multiuser coding gain of a CDMA system can achieve the single user coding gain when the crosscorrelations between users are low. However, since low crosscorrelations between short signature sequences are difficult to obtain in the time domain in an asynchronous system, frequency domain signature sequences are a viable alternative.

[1] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J.Selected Areas Comm.*, vol. 8, no.4, pp. 641-649, May 1990.

[2] J. Hui, "Throughput analysis for code division multiple accessing of the spread spectrum channel," *IEEE J. Selected Areas Comm.*, vol. SAC-2, no.4, pp. 482-486, July, 1984.

[3] T. R. Giallorenzi, *Multiuser Receivers for Coded CDMA Systems.* PhD thesis, Univ. of Virginia, Charlottesville, VA, 1994.

[4] J. A. Salehi, A. M. Weiner, and J.P. Heritage. Coherent ultrashort light pulse code-division multiple access communication systems. *IEEE J. of Lightwave Tech.*, 8, no. 3:478–491, 1990.

[1]Y. Rozenbaum is with Plantronics, Inc., Santa Cruz, CA

[2]Supported in part by NSF grant #ECS-9409452.

Unveiling turbo codes: some results on parallel concatenated codes

Sergio Benedetto and Guido Montorsi[1]
Dipartimento di Elettronica, Politecnico di Torino – Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract — **We propose an analytical method to upper bound the bit error probability of parallel concatenated block and convolutional codes.**

I. Introduction

The so called *turbo codes* [1], which in the following we will call parallel concatenated convolutional codes (PCCC), consist of two linear, generally simple convolutional codes (the *constituent codes*, CC) linked by an interleaver as shown in Fig. 1. In [1], PCCC's with appropriate choices of the CC's and of the interleaver have been shown to yield coding gains close to those predicted by the Shannon limit, yet keeping the complexity of an "ad hoc" iterative soft-decoding procedure significantly low and comparable to that of the CC's. These results have been further reinforced by [2]. Despite the astonishing performance of the turbo codes, however, neither serious attempts toward a theoretical explanation of the codes behavior/performance nor a sufficient comprehension of the role and relative importance of the ingredients of a PCCC have appeared in the literature so far. In this paper, we propose an analytical method to upper bound the error probability of a PCCC, and use it to shed light on important issues raised by these new coding schemes.

II. An analytical upper bound to the bit error probability of PCCC's

Fig. 1 shows clearly the discouraging complexity of the attempts trying to obtain the weight enumerating function of a PCCC, especially when the length N of the interleaver is large (say 1000-10000) as it should be to yield good performance. The only viable solution to the problem seems to pass through an appropriate and meaningful way of making independent the weights of the parity checks generated by the first and second encoders. To this end, we define a *uniform interleaver* as a probabilistic device which maps a given input information sequence of length N and weight w into all distinct $\binom{N}{w}$ permutations with equal probability $1/\binom{N}{w}$. Use of this device, instead of the actual interleaver, makes the weight enumerating functions $A_w^{C_1}(Z)$ and $A_w^{C_2}(Z)$ of the parity checks generated by the two encoders, conditioned to a given weight w of the input sequence, independent. As a consequence, the conditional weight enumerating function of the parity check bits of the whole PCCC $A_w^{C_P}(Z)$ can be easily obtained as

$$A_w^{C_P}(Z) = \frac{A_w^{C_1}(Z) \cdot A_w^{C_2}(Z)}{\binom{N}{w}},$$

and, from it, an upper bound to the bit error probability can be written in the form

$$P_b(e) \leq \sum_{w=1}^{N} \frac{w}{N} W^w A_w^{C_P}(Z)\Big|_{W=Z=e^{-R_c E_b/N_0}},$$

[1]This work was supported by European Space Agency and by CNR under Progetto Finalizzato Trasporti, sub-project Prometheus.

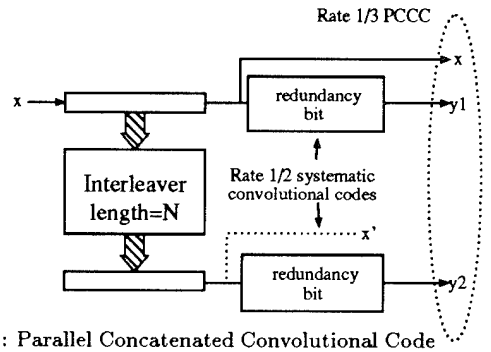


Fig. 1: Parallel Concatenated Convolutional Code

where R_C is the rate of the PCCC. Previous results refer to an $(N-L, 3N)$ block code equivalent to the PCCC and obtained from it considering input information sequences of length $N-L$ and codewords of length $3N$, where L is the constraint length of the CC's, generated by terminating trellises of the two CC's. Extensions to the case of continuous PCCC can be done [3].

III. The role of interleaver and CC's

Use of the uniform interleaver permits a separation of the effects of the interleaver length and of the CC's on the performance of the PCCC. Using our analytical tools, we see that, for large N and in the limits of the validity of the upper bounds, the interleaver provides an *interleaevr gain* which decreases the bit error probability by a factor $1/N$. Moreover, we prove that this gain can be obtained only if the CC's are recursive convolutional codes, and that this is due to the particular weight profile of them, characterized by the fact that input sequences of weight $w = 1$ do not produce error events of finite lengths. Finally, by extensive simulations, we validate the upper bounds based on the uniform interleaver, showing that an interleaver chosen as a random permutation is likely to yield bit error probabilities very close to those anticipated by the bounds.

As to the role of the recursive CC's (defined by the generating function $(1, n(D)/d(D)$ for the case of rate $1/2$), we have shown that a reasonable design criterion consists in choosing the polynomial $d(D)$ defining the feedback connections as a primitive polynomial, and that the choice of the numerator $n(D)$ should aim at maximizing the weight of the parity checks for input information sequences of minimum weight $w = 2$.

References

[1] Claude Berrou, Alain Glavieux, and Punja Thitimajshima. "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes". In *Proceedings of ICC'93*, Geneve, Switzerland, May 1993.

[2] D. Divsalar and F. Pollara. "Turbo Codes for PCS Applications". In *Proceedings of ICC'95*, Seattle, Washington, June 1995.

[3] Sergio Benedetto and Guido Montorsi. "Unveiling turbo-codes: some results on parallel concatenated coding schemes". Submitted to IEEE Transaction on Information Theory, January 1995.

Unit-Memory Hamming Turbo Codes

Jung-Fu Cheng Robert J. McEliece
Electrical Engineering Dept., California Institute of Technology, Pasadena, CA 91125, USA

I. Introduction

Several parallel concatenated coding schemes (turbo codes) based on multi-memory (MM) convolutional codes (more specifically, a $(2, 1, 4, 7)$ code) were recently proposed to achieve near Shannon-limit error correction performance with reasonable decoding complexity [1]-[3]. On the other hand, in many cases of interest, unit-memory (UM) codes have been demonstrated to have larger free distances than the MM codes with the same rate and the same number of memory elements [4]. In this paper, new turbo codes based on the $(8, 4, 3, 8)$ UM Hamming code [4] will be developed and shown to possess better performance potential in some senses. The standard turbo decoding algorithms, however, do not appear to achieve this potential.

II. Encoder

An equivalent systematic recursive generator matrix for the UM Hamming code can be obtained by first properly permuting the columns and then multiplying on the left by the inverse of the left-most 4×4 sub-matrix of the original generator matrix:

$$G = [I|P] = \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{1}{1+D} & 1 & \frac{D}{1+D} & 1 \\ 0 & 1 & 0 & 0 & 1 & \frac{1}{1+D} & 1 & \frac{D}{1+D} \\ 0 & 0 & 1 & 0 & \frac{D}{1+D} & 1 & 1 & \frac{1}{1+D} \\ 0 & 0 & 0 & 1 & 1 & \frac{D}{1+D} & \frac{1}{1+D} & 1 \end{bmatrix}$$

The corresponding encoder can be implemented with three memory elements. The encoder for the UM turbo (UMT) code is similar to those for the MMT codes [1]-[3], except that there are multiple inputs to the encoder of the component codes. The trellis is terminated using the method of [3]. Since the systematic bits from the second encoder are discarded, the overall code rate is $K/3(K+4)$, where K is the interleaver size.

III. The MAP Algorithm for Multi-Input Recursive Trellis Codes

In this section, a modified MAP algorithm is presented to deal with multiple inputs. Let the state of the encoder for the (n, k, ν) code at time t be $S_t \in \{0, 1, \ldots, 2^\nu - 1\}$, for $t = 0, \ldots, L = K/k$, where the initial and final states, S_0 and S_L, are known. The input block $\mathbf{u}_t = (u_{t,1}, \ldots, u_{t,k})$ causes a transition from S_{t-1} to S_t, and the corresponding output codeword $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,n})$ is observed over an AWGN channel as $\mathbf{y}_t = (y_{t,1}, \ldots, y_{t,n})$, for $t = 1, \ldots, L$. The log likelihood ratios of the *a posteriori* probabilities can be computed as:

$$\Lambda(u_{t,j}) = \log \frac{\sum_s \sum_{s'} \gamma^{+1}_{t,j}(s', s)\, \alpha_{t-1}(s')\, \beta_t(s)}{\sum_s \sum_{s'} \gamma^{-1}_{t,j}(s', s)\, \alpha_{t-1}(s')\, \beta_t(s)}$$

$$\alpha_t(s) = \frac{\sum_{s'} \Gamma_t(s', s)\, \alpha_{t-1}(s')}{\sum_s \sum_{s'} \Gamma_t(s', s)\, \alpha_{t-1}(s')}, \quad \text{for } t = 1, \ldots, L$$

$$\beta_t(s) = \frac{\sum_{s'} \Gamma_{t+1}(s, s')\, \beta_{t+1}(s')}{\sum_s \sum_{s'} \Gamma_{t+1}(s', s)\, \alpha_t(s')}, \quad \text{for } t = L-1, \ldots, 0$$

where, if the transition $s' \to s$ is allowed by input $u_{t,j} = i$,

$$\gamma^i_{t,j}(s', s) = \Pr\{u_{t,j} = i\} \Pr\{\mathbf{y}_t \,|\, S_t = s, u_{t,j} = i, S_{t-1} = s'\}$$

$$\Gamma_t(s', s) = \sum_{\mathbf{i}\,:\, s' \to s} \Pr\{\mathbf{u}_t = \mathbf{i}\} \Pr\{\mathbf{y}_t \,|\, S_t = s, \mathbf{u}_t = \mathbf{i}, S_{t-1} = s'\}$$

IV. Decoder and Performance

The decoder structure used is similar to that in [2] except that the MAP algorithm in III is applied instead. Numerical results are shown in Fig. 1 and summarized as follows:

- The minimum distance of the $(60, 16)$ UMT code with the best known interleaver is 14. Maximum-likelihood decoding simulation of this code shows a gain of 0.5 dB over the $(80, 16)$ MMT code [3] which has the same minimum distance. The use of turbo decoding introduces a loss of about 1.5 dB.
- For large block lengths, simulation results show that the turbo decoding algorithm converges faster than that for MMT codes, but the performance is not as good. Comparing these with the transfer bounds computed with a double recursion method and a random averaging argument [5], a gap of coding gain with turbo decoding as in the previous case can be observed again.

References

[1] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo-codes," *Proc. IEEE Int'l Conf. Comm.'93.*

[2] P. Robertson, "Illuminating the structure of code and decoder of parallel concatenated recursive systematic (turbo) codes," *Proc. IEEE Global Telecomm. Conf.'94.*

[3] D. Divsalar and F. Pollara, "Turbo Codes for Deep-Space Communications," *JPL TDA Progress Report 42-120*, Feb. 1995.

[4] K. Abdel-Ghaffar, R. J. McEliece and G. Solomon, "Some partial-unit-memory convolutional codes," *JPL TDA Progress Report 42-107*, Nov. 1991.

[5] S. Benedetto and G. Montorsi, "Performance evaluation of turbo-codes," *IEE Elec. Letters,* Feb. 1995.

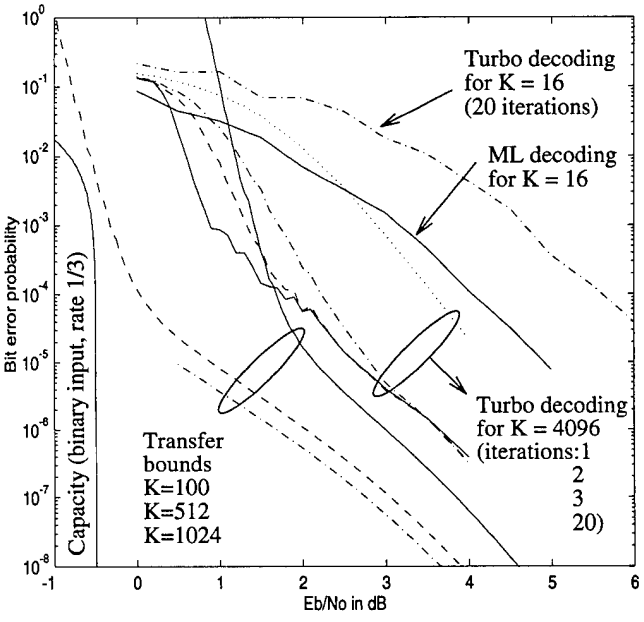


Figure 1: Performance of unit-memory turbo codes.

Distance Spectrum of the Turbo-codes

R. Podemski (#), W. Hołubowicz(#), C. Berrou(*), A. Glavieux(*)

(#) - Franco-Polish School of New Information and Communication Technologies, ul.Mansfelda 4, 60-854 Poznań , POLAND

(*) - Ecole Nationale Superieure des Télécommunications de Bretagne, Technopole de Brest Iroise, BP 832, 29285 Brest Cedex, FRANCE

Abstract - **In this paper an analytical approach to newly invented Turbo-Codes (TC) is presented. That approach is based on evaluating the properties of TC by means of the Minimum Hamming Distance (MHD) and the Hamming Distance Spectrum (HDS). An algorithm for computing HDS is presented and numerical results are discussed. The concept of basic return-to-zero sequence is introduced. It is shown how basic return-to-zero sequences can be used in the algorithm for computing HDS and how it can justify the properties of TC. Numerical results of computing MHD and HDS for different TCs are presented and verified by simulations.** [1]

I. INTRODUCTION

TC seem to be very attractive for applications in practical communication systems, since their error performance is close to the Shannon limit [1, 2]. During the last two years some modifications of the originally proposed parameters of both the encoder and decoder of the turbo-code, have been proposed, which lead to the improvement of the turbo-code performance. In most cases the performance of turbo-codes has been evaluated by means of simulation. In this paper we show that the properties of turbo-codes can be predicted by means of MHD and HDS. We describe a method to efficiently calculate the MHD and HDS of the turbo-codes, provided that the interleaver size is not larger than 16x16. This procedure is a modification of the well known Fano-algorithm. We introduce the concept of basic return-to-zero sequence. We shown how basic return-to-zero sequence can be used in the Fano algorithm for computing HDS. We show also how the properties of are correlated with basic return-to-zero sequences can justify the properties of TC.

II. DESCRIPTION OF THE SYSTEM

The scheme of a turbo-encoder is given in Fig. 1 [1]. Turbo-encoder consists of two Recursive Systematic Coders (RSC), Interleaver (I) and puncturing circuit. Both RSC encoders are identical rate-1/2 convolutional encoders. In our study we have considered RSC encoders: (23,35), (7,5), (5,7), (15,17), (5,7), (1,1)[2]. The puncturing pattern used by us is following: we transmit bit Y0 without any change, alternatively every second bit Y1, Y2 is punctured Thus the overall rate of the TC is 1/2 and the transmitted sequence is: Y0, Y1, Y0, Y2, Y0, Y1 ... .

III. AN ALGORITHM FOR COMPUTING HDS

The algorithm used by us for computing HDS of the turbo-codes is the modified Fano algorithm. The modification is following: we use the fact that in order for a turbo-coder to return to the all-zero-state, both RSC encoders must come to the zero state. So, instead of applying to the input of the turbo-code arbitrary binary sequences, we feed it only with some selected sequences which are known to force RSC1 to come to the zero state, so called *return-to-zero sequences*. *Basic return-to-zero sequences* are defined as those return-to-zero sequences which are not a linear combination of other return-to-zero sequences. We have proven that for any recursive code there exists only one basic return-to-zero sequence. For example, for RSC (5,7) the basic return-to-zero sequence is $\underline{x}$=[101], for RSC (7,5) it is equal to $\underline{x}$=[111].

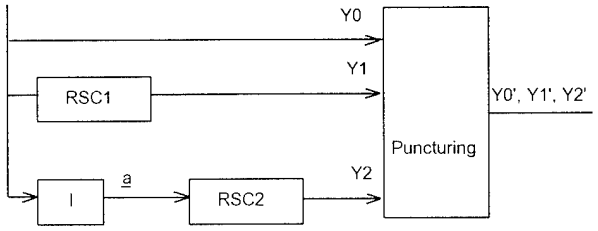


Fig. 1. The scheme of the turbo-encoder.

IV. BASIC RETURN-TO-ZERO SEQUENCE

Basic return-to-zero sequence can always help in rejecting "bad" RSC encoders. We have shown that for any TC (whatever the size or kind of interleaving is) with RSC (5,7) one basic return-to-zero sequence can drive TC to the all-zero-state. For such TC when we use the puncturing pattern presented in Fig. 2 the weight of an output sequence of TC is always equal to 5. Thus for that particular RSC code and puncturing pattern any changes in the size of the interleaver or introducing non-uniformity to the interleaver, will not lead to the increase of MHD.

V. CONCLUSIONS

We have computed HDS for a range of TC, for different RSC codes, different interleavers (sizes up to 16x16, both uniform and non-uniform). We have also verified our results by simulation. Conclusions of our study are the following:

- Simulation results show that analytical approach by using MHD and HDS can be used for evaluating the properties of turbo-codes. For example for TC with RSC (7,5) and the interleaver I=8x8 the difference between simulation and analytical results for BER=10^{-5} is 0.22 dB for uniform interleaving and 0.7 dB for non-uniform one.
- The number of elements in the HDS which must be taken into account does not exceed 3 (sometimes one spectrum element is sufficient). For example for TC with RSC (7,5) and uniform interleaver I=8x8 the difference between BER curves for 1 and 3 (or more) elements is 0.4 dB for BER=10^{-4} and 0 dB for BER=10^{-6}. There is no difference in BER between 3 or more elements.
- BER of the turbo-code can be increased by:
  - *increasing the constraint length of the RSC code*. For example for BER=10^{-6} and uniform interleaver I=8x8 the TC with RSC (23,35) is better than TC with RSC (15,17) by 3.2 dB, and outperforms TC with RSC (1,1) by about 5.8 dB.
  - *increasing the size of the interleaver*. For example for TC with RSC (7,5) for BER=10^{-6} TC with I=8x16 outperforms TC with I=8x8 by 2.4 dB and TC with no interleaving by 1.6 dB.
  - *introducing non-uniformity to the interleaver*,
- For any Recursive Code there exists only one basic return-to-zero sequence. By analyzing the properties of basic return-to-zero sequence we may find "bad" codes. The problem which is still open is how basic return-to-zero sequence can be used for designing TC which would possess very good properties (i.e. large MHD value).

REFERENCES

[1] C.Berrou, A.Glavieux, P.Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding:Turbo-Codes", ICC'93 Symposium, Geneva, May 1993,

[2] Patents No 9105279 (France), No 92460011.7 (Europe), No 07/870,483 (USA)

[1] This work was partially sponsored by the following grant of the National Committee for the Scientific Research: KBN-8S50401905.

[2] Generating polynomials are given in the octal notation.

Low-Rate Turbo Codes for Deep-Space Communications

D. Divsalar and F. Pollara[1]

Jet Propulsion Laboratory, MS 238-420, 4800 Oak Grove Drive, Pasadena, California, 91109, USA

Abstract— **We develop b/n multiple turbo codes and an iterative turbo decoding scheme based on an approximation to the optimum bit decision rule (MAP). For random interleaver size of 16384 bits, a bit error probability of 10^{-5} at a required E_b/N_o of about 0.8 dB from the binary-input channel capacity for rate b/n was obtained for various turbo codes. Examples are given for rate b/n =1/2, 1/3, 1/4 and 2/6 turbo codes using component codes with up to 16 states.**

I. Introduction

Turbo codes were recently proposed by Berrou, Glavieux and Thitimajshima [1]. We propose rate b/n codes that consist of the parallel concatenation of q systematic recursive convolutional codes, with random interleavers of size N between rate b/n_i, encoders, such that $n = \sum_{i=1}^{q} n_i$. Encoding and decoding is done block by block. Encoders are forced to the all-zero state at the end of each block by a simple termination method [4].

II. Turbo Decoding for Multiple Codes

Let u_k be a binary random variable taking values in $\{0, 1\}$, representing the sequence of information bits $\mathbf{u} = (u_1, \ldots, u_{Nb})$. This sequence is partitioned into N groups of b bits representing input symbols. Bit-by-bit, rather than symbol-by-symbol, interleaving is performed.

The modified MAP algorithm [5] provides the log likelihood ratio $L_k = \log \frac{P(u_k=1|\mathbf{y})}{P(u_k=0|\mathbf{y})}$ given the received symbols $\mathbf{y}$, where

$$L_k = \log \frac{\sum_{\mathbf{u},u_k=1} P(\mathbf{y}|\mathbf{u}) \prod_{j\neq k} P(u_j)}{\sum_{\mathbf{u},u_k=0} P(\mathbf{y}|\mathbf{u}) \prod_{j\neq k} P(u_j)} + \log \frac{P(u_k=1)}{P(u_k=0)} \quad (1)$$

Consider the parallel concatenation of q codes. The combination of permuter and systematic recursive convolutional code is considered as a block code with input $\mathbf{u}$ and output $\mathbf{x}_j$, $j = 1, 2, \ldots, q$. The components of $\mathbf{x}_j$ may be binary or non-binary. For the non-binary case multilevel modulation is used, resulting in *turbo trellis coded modulation* (TTCM). The corresponding received sequences are $\mathbf{y}_j$.

The optimum bit decision rule (MAP) for data with uniform probabilities is

$$L_k = \log \frac{\sum_{\mathbf{u},u_k=1} \prod_{j=1}^{q} P(\mathbf{y}_j|\mathbf{u})}{\sum_{\mathbf{u},u_k=0} \prod_{j=1}^{q} P(\mathbf{y}_j|\mathbf{u})} \quad (2)$$

An approximation to $P(\mathbf{y}_j|\mathbf{u})$ was used in [4] to obtain (2) as $L_k = \sum_{j=1}^{q} \tilde{L}_{jk}$, where $\tilde{L}_{jk}$'s are iterative solutions to a set of non-linear equations that can be efficiently computed using the MAP algorithm with pre-interleaving and post-deinterleaving as

$$\tilde{L}_{jk}^{(m+1)} = \log \frac{\sum_{\mathbf{u},u_k=1} P(\mathbf{y}_j|\mathbf{u}) \prod_{i\neq k} e^{u_i \sum_{l=1,l\neq j}^{q} \tilde{L}_{li}^{(m)}}}{\sum_{\mathbf{u},u_k=0} P(\mathbf{y}_j|\mathbf{u}) \prod_{i\neq k} e^{u_i \sum_{l=1,l\neq j}^{q} \tilde{L}_{li}^{(m)}}} \quad (3)$$

for $k = 1, 2, \ldots, Nb$ and $j = 1, 2, \ldots, q$. Then $\tilde{L}_{jk} = \lim_{m\to\infty} \tilde{L}_{jk}^{(m)}$. All initial conditions are set to zero, i.e. $\tilde{L}_{jk}^{(0)} = 0$.

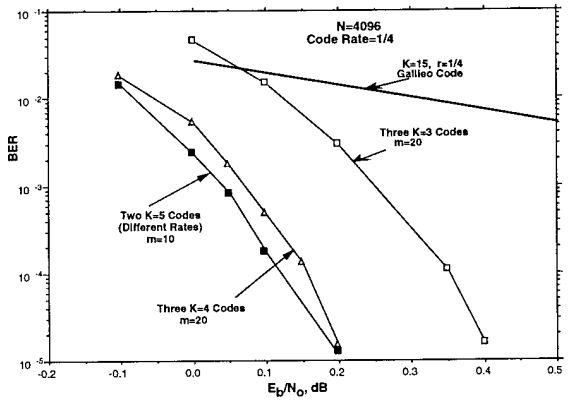
[1]The research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

III. Performance

The bit error rate performance of these codes was evaluated by using transfer function bounds [3] [2]. In [2] it was shown that transfer function bounds are very useful for signal-to-noise ratios above the cutoff rate threshold and that they cannot accurately predict performance in the region between cutoff rate and capacity. In this region, the performance was computed by simulation.

The figure below shows the performance of turbo codes with the following generators: For two $K = 5$ constituent codes, $(1, g_b/g_a, g_c/g_a)$ and (g_b/g_a), with $g_a = (37)_{octal}$, $g_b = (33)_{octal}$ and $g_c = (25)_{octal}$; For three $K = 3$ codes, $(1, g_b/g_a)$ and (g_b/g_a) with $g_a = (7)_{octal}$ and $g_b = (5)_{octal}$; For three $K = 4$ codes, $(1, g_b/g_a)$ and (g_b/g_a) with $g_a = (13)_{octal}$ and $g_b = (11)_{octal}$.

Further results at BER=10^{-5} were obtained for two constituent codes with interleaving size $N = 16384$ as follows. For a rate 1/2 turbo code using two codes, $K = 2$ (differential encoder) with (g_b/g_a) where $g_a = (3)_{octal}$ and $g_b = (1)_{octal}$, and $K = 5$ with (g_b/g_a) where $g_a = (23)_{octal}$ and $g_b = (33)_{octal}$ the required bit SNR was 0.85 dB. For rate 1/3, we used two $K = 5$ codes, $(1, g_b/g_a)$ and (g_b/g_a) with $g_a = (23)_{octal}$ and $g_b = (33)_{octal}$ and obtained bit SNR= 0.25 dB. For rate 1/4, we used two $K = 5$ codes with $(1, g_b/g_a, g_c/g_a)$ and (g_b/g_a) with $g_a = (23)_{octal}$, $g_b = (33)octal$ and $g_c = (25)_{octal}$ and obtained bit SNR = 0 dB. For a rate 2/6 turbo code each constituent code is constructed from two parallel $K = 3$ codes $(1, g_{b1}/g_a, g_{c1}/g_a)$ and $(1, g_{b2}/g_a, g_{c2}/g_a)$ where the output of g_{b1}/g_a is added to the output of g_{b2}/g_a and the output of g_{c1}/g_a is added to the output of g_{c2}/g_a. $g_a = (7)_{octal}$, $g_{b1} = (6)_{octal}$, $g_{c1} = (1)_{octal}$, $g_{b2} = (7)_{octal}$, $g_{c2} = (4)_{octal}$. The resulting code has 16 states with two inputs and four outputs. The second code is identical to the first one but not using the systematic bits. BER=10^{-5} was obtained at bit SNR=0.2 dB.



References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in Proc. ICC '93.

[2] D. Divsalar, S. Dolinar, R.J. McEliece, F. Pollara, "Transfer Function Bounds on Turbo Code Performance", JPL TDA Prog. Rep. 42-122, Aug 15, 1995.

[3] S. Benedetto and G. Montorsi, "Performance evaluation of turbo-codes", Electronics Letters, Feb. 2, 1995, Vol. 31, No. 3, pp. 163-165.

[4] D. Divsalar and F. Pollara, "Multiple Turbo Codes for Deep-Space Communications", JPL TDA Prog. Rep. 42-121, May 15, 1995.

[5] J. Hagenauer and P. Robertson, "Iterative (Turbo) decoding of systematic convolutional codes with the MAP and SOVA algorithms", Proc. of the ITG conference "Source and channel coding", Oct. 1994, Frankfurt.

'Turbo' Coding for Deep Space Applications

Jakob Dahl Andersen (jda@it.dtu.dk)
Institute of Telecommunication, Technical University of Denmark, DK-2800 Lyngby, Denmark

Abstract - **The performance of the 'turbo' coding scheme is measured and an error floor is discovered. These residual errors are corrected with an outer BCH code. The complexity of the system is discussed, and for low data rates a realizable system operating at E_b/N_0 below 0.2 dB is presented.**

I. Introduction

Recently it has been discovered that a very good performance can be achieved with iterative decoding of a parallel concatenation of small convolutional codes [1]. This coding scheme is named 'turbo' coding. The basic idea is to encode the information sequence twice, the second time after a pseudo-random interleaver, and to do iterative decoding on the two encoded sequences in two decoders. The system can be regarded as a kind of product code. Due to the information exchange among the two decoders the decoding algorithm must provide soft output. We use the MAP algorithm [2] which actually calculates the a posteriori probability of each information bit. The convolutional codes are used in a recursive systematic form since it gives an improved performance with this system.

II. The Error Floor

The first simulations were based on the recursive systematic code $(1, 1+D^4/1+D+D^2+D^3+D^4)$. We use the same code for both encoders but for the second one the information sequence is not transmitted. This gives an overall rate of 1/3. We use a block length of 10384 information bits. For all simulations presented in this paper all numbers including the channel input are represented as floating point values.

As seen from Figure 1, the results achieved with this system are very promising since the Bit Error Rate (BER) after 18 iterations is close to 10^{-5} already at 0.2 dB. Unfortunately the BER decreases very slowly for improved SNR. What we see are many frames with only a few bit errors. This is due to the low free distance of this coding scheme. The free distance of this system might be as low as 10. The actual profile depends on the specific interleaver.

A search for better interleavers might give improved performance. However, the main problem is combinations of two low weight words for the basic code. Consequently the performance with interleaver structures like block interleavers is quite poor, and a search among the random interleavers can only remove a couple of the worst low weight patterns.

III. The Extended 'turbo' Coding Scheme

An obvious way to remove the error floor (or saddle) is to use an outer code. Since the bursts consist of very few bit errors, we will use a (10384,10000) BCH code capable of correcting 24 errors. This outer code corrects all the residual errors, but we loose 0.16 dB due to the decreased rate. With this system the Probability of Frame Loss (PFL) is below 10^{-4} at 0.4 dB.

Improved performance can be achieved with a system based on rate 1/3 codes with only 8 states. This gives rate 1/5 for the 'turbo' coding scheme. In this case we have also used the outer BCH code.

With this system we have simulated 25,000 frames without frameloss at 0.1 dB. This means that the 90% confidence level for the PFL is below 10^{-4}. The BER is shown in Figure 1.

IV. Complexity

The performance must of course be compared to the complexity. We have estimated the number of operations needed in the MAP algorithm for recursive systematic codes and conclude that this is about 4 times the number of operations in a Viterbi decoder. This means that the number of operations for 18 iterations with M=3 codes is in the order of 2^{12}. We believe that with a logarithm quantization an 8 bit representation is sufficient for the internal representation in the MAP decoder. With this quantization and channel input quantized in 16 levels we expect a performance degradation about 0.1 dB.

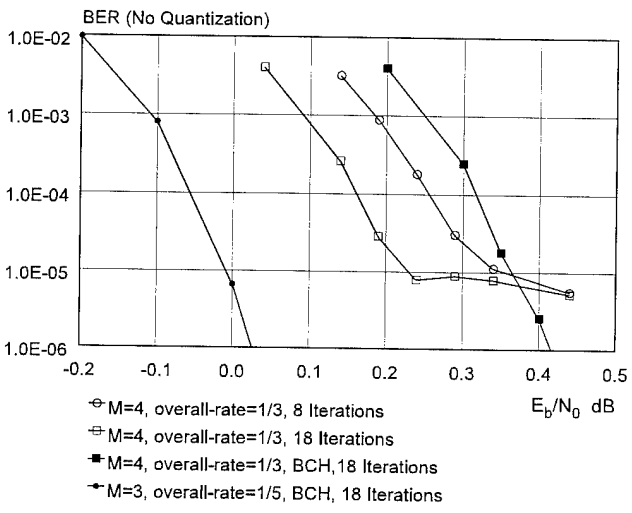
For low data rates the 'turbo' coding scheme can be implemented with only one MAP decoder (used 2×18 times), and the decoder for the BCH code can be implemented on a DSP. Further the calculations inside the MAP decoder can be serialized, using the same hardware for each state.

This means that for data rates below 100 kbit/s the complexity of this system is moderate, and the extended 'turbo' coding scheme might be an alternative to ordinary concatenated systems.

References.

[1] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-correcting Coding and Decoding: Turbo-codes(1)", *Proc. ICC '93*, May 1993, pp. 1064-1070.

[2] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate", *IEEE Transactions on Information Theory*, Vol. IT-20, March 1974, pp. 284-287.



Interleaver Design for Three Dimensional Turbo Codes

Adrian S. Barbulescu and Steven S. Pietrobon
Satellite Communications Research Centre, University of South Australia, The Levels SA 5095, Australia

Abstract — **A new bandwidth efficient interleaver is described for turbo codes when used to decode short frames of data using the MAP algorithm. Applications in rate compatible turbo codes and encryption are presented.**

I. Introduction

It is well known that the interleaver design is the key to achieve the best performance for turbo codes [1] . For very large frame sizes, random interleavers are near optimum. For small frame sizes – for which the interleaver depth is less than ten times the constraint length of the component convolutional code – a random interleaver is not the best choice. In the following, we consider a three dimensional turbo–code (3D–TC) shown in Figure 1 which has the feedback polynomial equal to all ones.

II. Design Criteria

In order to use a *maximum a posteriori* (MAP) decoding algorithm [2], the initial and the final state of each one dimensional encoder should be fixed for all three coded sequences. This could be achieved by appending three different "tails", one for each coded sequence which will reduce the bit rate. A new interleaver type called a "simile" interleaver was described in [3] for a two–dimensional turbo–code which needs only one "tail" to be appended. A similar method will be used to create a "simile" interleaver for a 3D–TC.

We denote v the encoder memory size of each one dimensional encoder. We can rearrange the whole block of N information bits in mod (v + 1) sequences. The important advantage in doing this is that from the point of view of the final encoder state, the order of the individual bits in each sequence does not matter as long as they belong to the same sequence. The "simile" interleaver has to perform the interleaving of the bits within each particular sequence in order to drive the encoder into the same state as without interleaving. In [3] we described a particular block helical interleaver. This can be extended to 3D–TC by assuming that the number of columns is a multiple of (v + 1). The information sequence is stored row–wise and the two interleaved sequences start from the left corners: bottom left corner and up the diagonal for interleaver I^a and top left corner and down the diagonal for interleaver I^b.

A second criteria is needed if the coded bits are punctured: each information bit should have associated with it, after puncturing, one and only one coded bit. In this way the correction capability of the code is uniformly distributed over all information bits. This type of interleaver was introduced in [4] for a two dimensional turbo–code and was called an "odd–even" type of interleaver.

Using a block helical interleaver, if the number of columns is a multiple of the dimension order, which is 3 for 3D–TC, we can multiplex the coded bits of the straight sequence whose index in time modulo 3 is zero with the interleaved I^a coded bits whose index in time modulo 3 is one and with the interleaved I^b coded bits whose index in time modulo 3 is two. In this way all information bits have associated with them one and only one coded bit.

III. Applications

The coding gain can be varied without changing the convolutional code. In a good channel a rate half turbo code composed of the uncoded sequence {x} and the punctured sequence $\{y/y^a\}$ can be used. If the channel becomes noisier a rate third code can be obtained by transmitting {x}, {y} and $\{y^a\}$ sequences. It was shown in [5] that the probability of error is proportional with N^{-1}. For an even worse channel a rate quarter code can be used by transmitting the $\{y^b\}$ sequence which would make the probability of error proportional with N^{-2} and so on. As in the case of rate compatible convolutional codes, the same turbo decoder can be used in all cases.

In Figure 1 we use the sixteen state turbo code (v=4) [1]. Each interleaver is made from five pseudo random interleavers with different generator polynomials which can start from different states produced by a long pseudo random generator. The outputs of the turbo encoder are buried in noise whose variance is known and can be changed each frame or even in each interleaved sequence. The long pseudo random generator which generates the starting states of the interleavers together with the variance of the noise are the keys to the proposed encryption system. We assume these keys to be secret and known at the receiver end. This principle is similar with that for CDMA transmissions.

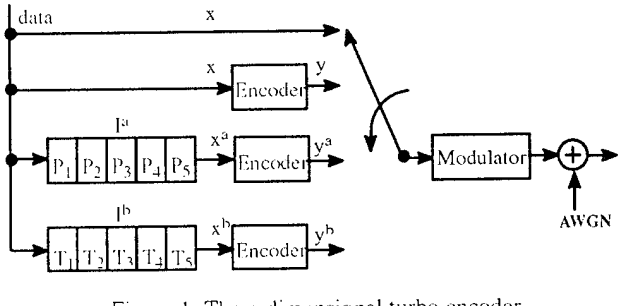


Figure 1. Three dimensional turbo encoder

This work was supported by the Institute for Telecommunications Research, University of South Australia.

References

[1]. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error–Correcting coding and decoding: Turbo–Codes," *ICC 1993*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[2]. S. S. Pietrobon and A. S. Barbulescu, "A simplification of the modified Bahl decoding algorithm for systematic convolutional codes," *Int. Symp. Inform. Theory & its Applic.*, Sydney, Australia, pp. 1073 – 1077, Nov. 1994.

[3]. A. S. Barbulescu and S. S. Pietrobon, "Terminating the trellis of turbo codes in the same state," *Electron. Lett.*, vol 31, pp. 22–23, Jan. 1995.

[4]. A. S. Barbulescu, and S. S. Pietrobon, "Interleaver design for turbo codes," *Electron. Lett.*, vol 30, pp. 2107–2108, Dec. 1994.

[5]. S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," submitted to *IEEE Trans. Inform. Theory*., Jan. 1995.

Weight Distributions of Turbo–Codes

Yuri V. Svirid

Chair for Communications, Technical University of Munich, Arcisstr. 21, D-80290 Munich, Germany, *and*
Dept. for RTS, Belarusian State University of Informatics and Radioelectronics, P. Brovka str. 6, 220027 Minsk, Belarus

Abstract — **An optimal interleaving between two component encoders of a turbo–code is proposed. For any real constructable interleaver the optimality criterion is given. For component codes (CC) with known weight distribution (WD) the WD of the turbo–code with perfect interleaving is calculated. As CC's the random codes and terminated convolutional codes are considered. It is shown that the often observed "break" in the performance curves for turbo–codes is a result of their "broken" WD.**

I. Introduction

Any codeword of the recently introduced turbo–codes [1] has the following structure: $[\mathcal{I}|\mathcal{I}\Lambda|\mathcal{I}'\Lambda]$, where $\mathcal{I}$ is the k–tuple of the information bits, Λ is the $k \times r$ binary matrix, and $\mathcal{I}'$ is a version of $\mathcal{I}$ with interleaved (permutated) coordinates. As CC's both systematic block codes and convolutional codes with terminated encoders have been in use until now. The rate of the whole code in both cases is $R = k/(k+2r)$. The linearity of turbo–codes is shown in [2].

II. Optimal Interleaving and WD of Whole Code with Known WD's of Component Codes

Dispose all $2^k - 1$ nonzero codewords of one CC into k groups so that each ith ($i = \overline{1,k}$) group consists of $\binom{k}{i}$ codewords of weight i in the information part. Note, that *if the information vector $\mathcal{I}$ belongs to the ith group, then the permutated vector $\mathcal{I}'$ will be in this group too.*

The aim of interleaving is to produce (by manipulating the weights of the second redundancy part) the whole codewords with the overall weights as large as possible. It means that within each group the first redundancy part with small weight should be associated after interleaving with a second redundancy part with large weight and vice versa.

Let the WD of CC be known in the form $A(i,j)$, which denotes the number of codewords with Hamming weight i of the information bits and weight j of the redundancy bits. Within each group dispose the codewords with non–decreasing weights of the redundancy part so that for any l holds: $j(i,l+1) \geq j(i,l)$, where $j(i,l)$, $l = \overline{1,\binom{k}{i}}$, is the weight of the redundancy part of the lth codeword in the disposed ith group. Note, that for any i and l the numbers $j(i,l)$ are determined by $A(i,j)$. The lth codeword of the turbo–code in this group has then weight

$$W(i,l) = i + j(i,l) + j(i, \tbinom{k}{i} - l + 1). \qquad (1)$$

Counting all codewords, from (1) we immediately obtain the WD of the turbo–code in the form $A(i,j)$, which yields also the number $A(w)$ of codewords with weight w.

An interleaving, which leads to the same WD of the turbo–code as can be obtained from (1), will be called a *fully optimal interleaving* (f.o.i.).

Viewing $W(i,l)$ for each i as a random variable of l, the criterion for the optimal interleaving can be formulated as a problem of minimizing its variance: $\sigma_l^2\{W(i,l)\} \longrightarrow \min$.

III. Turbo–Codes with Random CC's

The random $(k+r,k)$ code has the WD $A(w) = \binom{k+r}{w}/2^r$, which is obtained from the equation between the probability of occurrence of $(k+r)$–tuple and of codeword both of weight w. However, for applying (1) the WD in the form $A(i,j)$ is required. From a similar equation for each group we get:

$$A(i,j) = \tbinom{k}{i}\tbinom{r}{j}\tfrac{1}{2^r}. \qquad (2)$$

Because of Vandermonde convolution: $\sum_{i+j=w} \binom{k}{i}\binom{r}{j} = \binom{k+r}{w}$, the code with WD (2) is a random code too.

Combining (2) and (1), we see that for each group i (due to the symmetry $\binom{r}{j} = \binom{r}{r-j}$) each parity–weight j is associated after f.o.i. with a second parity–weight $r-j$. Thus, $\forall i,l$: $W(i,l) = i + r$. Furthermore, $A(i,r) = \binom{k}{i}$, $A(i, j \neq r) = 0$ and the WD of the whole code is: $A(0) = 1$, $A(w) = \binom{k}{w-r}$ for $r < w \leq k+r$, and $A(w) = 0$ otherwise. Hence, the minimum distance is $r+1$, which increases with increasing k.

IV. Convolutional Codes as CC's

WD of these terminated codes for great k and rate $R = 1/2$ can be written as $A(i,j)/\binom{k}{i} = \binom{r}{j}\rho_{i,t}^j(1-\rho_{i,t})^{r-j}$, where $\rho_{i,t} = (1-(1-2i/k)^{J(t)})/2$, for feed–back encoders $J(t)$ is a linear function of the time $t = \overline{1,k}$ and for feed–forward encoders $J(t)$ it is a constant J equal to the number of nonzero terms in the generator polynomial ($\rho_{i,t} = \rho_i$ in this case). According to the DeMoivre–Laplace theorem the right–hand side of the last WD can be approximated by a Gaussian distribution: $A(i,j)/\binom{k}{i} \approx \exp\left(-(j-\mu_i)^2/(2\sigma_i^2)\right) / \left(\sigma_i\sqrt{2\pi}\right)$, with mean $\mu_i = r\rho_i$ and variance $\sigma_i^2 = r\rho_i(1-\rho_i)$, where for feed–back encoders ρ_i is the time average of $\rho_{i,t}$.

Due to the symmetry of the Gaussian distribution around its mean, one sees that after applying the f.o.i. rule (1) all codewords of the turbo–code within the ith group have weight $W(i,l) \approx i + 2\mu_i$, while the total number of codewords in this group is $\binom{k}{i}$. In case of feed–forward encoders $W(i,l) \approx i + 2Ji$ for small and large i and $W(i,l) \approx i + r$ for i near to $k/2$ (which corresponds to random codes). Hence, the minimum distance is $1+2J$. For feed–back encoders the minimum distance increases with increasing k and $W(i,l) \approx i + r$ for all i except very small ones. Codes with these encoders are thus near to random codes. The great distinction between values $W(i,l)$ and between number of codewords for small and central i results into a "break" in the performance curves.

Using the proposed WD's, one can obtain the bounds on error rate for turbo–codes like union bounds in [2].

References

[1] C. Berrou, A. Glavieux, P. Thitimajshima, "Near Shannon limit error–correcting coding and decoding: Turbo–codes (1)," *International Conference on Communications (ICC'93)*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[2] Yu. V. Svirid, "Additive upper bounds for turbo–codes with perfect interleaving," *EIDMA Winter Meeting on Coding Theory, Information Theory and Cryptology*, Veldhoven, The Netherlands, p. 35, December 19–21, 1994.

Threshold Decoding of Turbo–Codes

Yuri V. Svirid*, Sven Riedel

Chair for Communications, Technical University of Munich, Arcisstr. 21, D-80290 Munich, Germany
*also with Belarusian State University of Informatics and Radioelectronics, P.Brovka str. 6, 220027 Minsk, Belarus

Abstract — **The idea of iterative decoding of two–dimensional systematic convolutional codes — so–called turbo–codes — is extended to threshold decoding, which is presented in "Soft–In/Soft–Out" form. The computational complexity of the proposed decoder is very low. Surprisingly good simulation results are shown for the Gaussian channel.**

I. Preliminaries

We restrict ourselves to binary data. A convolutional encoder with rate $R_c = k/(k+1)$ produces the output bits $x_u^{(1)}, \ldots, x_u^{(k+1)}$ at time $u = 0, 1, 2, \ldots$. During the transmission the noise sequence $e_u^{(1)}, \ldots, e_u^{(k+1)}$ corrupts the coded bits. This sequence is statistically independent from digit to digit. Thus, we receive the sequence $\hat{x}_u^{(i)} = x_u^{(i)} \oplus e_u^{(i)}$, $1 \le i \le k+1$, where $\oplus$ denotes the modulo–two addition. We assume that an error has occurred, if $e_u^{(i)} = 1$, and $e_u^{(i)} = 0$ otherwise.

For threshold decoding it is important to provide information about the error symbol $e_u^{(i)}$. The a posteriori log-likelihood ratio for this symbol can be calculated as $L(e_u^{(i)}|y_u^{(i)}) = \ln \frac{P(e_u^{(i)}=0|y_u^{(i)})}{P(e_u^{(i)}=1|y_u^{(i)})} = 4\frac{E_s}{N_0} a \cdot |y_u^{(i)}| + L(e_u^{(i)})$, where $y_u^{(i)}$ is the matched filter output associated with the binary value $x_u^{(i)}$, E_s/N_0 is the signal–to–noise ratio, a is the fading amplitude, and $L(e_u^{(i)})$ is the a priori log-likelihood ratio for symbol $e_u^{(i)}$.

Following [1], we shall use a special operation $\boxplus$, which denotes $L(v_1) \boxplus L(v_2) = L(v_1 \oplus v_2)$ for log-likelihood ratios of statistically independent binary random variables v_1 and v_2.

II. Soft–In/Soft–Out Threshold Decoding

Soft–In threshold decoding is well–known as A Posteriori Probability (APP) decoding [2]. The objective of Massey's decoder is to maximize the probability $P(e_0^{(i)} = \xi | \{A_j^{(i)}\})$ that the error symbol $e_0^{(i)}$, $1 \le i \le k$, has a certain value $\xi \in \{0, 1\}$ under the condition that we have a set $\{A_j^{(i)}\}$, $1 \le j \le J$, of parity checks orthogonal on $e_0^{(i)}$. Each parity check $A_j^{(i)}$ can be calculated as modulo–two sum of $e_0^{(i)}$, a special selection of error symbols $e_s^{(\alpha)}$, $1 \le \alpha \le k$, $s \in S_j^{(i,\alpha)}$, associated with the information bits $x_s^{(\alpha)}$, and the error symbols $e_{s'}^{(k+1)}$, $s' \in S_j^{(i,k+1)} \subset \bigcup_{\alpha=1}^{k} S_j^{(i,\alpha)}$, associated with the parity check bits $x_{s'}^{(k+1)}$. The sets $S_j^{(i,\alpha)}$ and $S_j^{(i,k+1)}$ consisting of integers are depending on the generator polynomials of the code. The soft output of the decoder can be written as

$$L(e_0^{(i)}|\{A_j^{(i)}\}, y_0^{(i)}) = \underbrace{\sum_{j=1}^{J}(1-2A_j^{(i)})w_j^{(i)}}_{\text{extrinsic}} + \underbrace{4\frac{E_s}{N_0} a \cdot |y_0^{(i)}|}_{\text{channel}} + \underbrace{L(e_0^{(i)})}_{\text{a priori}},$$

where

$$w_j^{(i)} = \boxplus_{\alpha=1}^{k} \boxplus_{s \in S_j^{(i,\alpha)}} L(e_s^{(\alpha)}|y_s^{(\alpha)}) \boxplus \boxplus_{s' \in S_j^{(i,k+1)}} L(e_{s'}^{(k+1)}|y_{s'}^{(k+1)}).$$

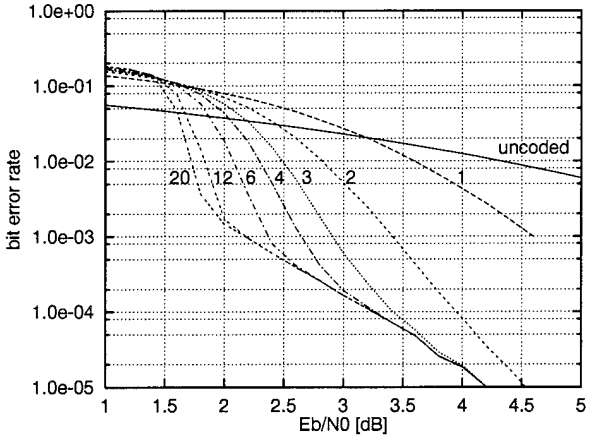
The $\boxplus$ operation can be approximated by sign and minimum operations. The value $1 - 2A_j^{(i)}$ is equal to $+1$ or -1. Thus, we need only compare operations and additions to calculate the extrinsic term.

III. Iterative ("Turbo") Decoding

We can split the soft output into three terms, namely into the so–called extrinsic information representing the influence of the error bits orthogonal on the current bit $e_0^{(i)}$, the soft output of the channel, and the a priori value $L(e_0^{(i)})$. If a priori information about the error bits is available, it is also used in calculating the weights $w_j^{(i)}$. Only the extrinsic value (the information produced by the previous decoder) should be passed on as new a priori value to the next decoder. The structure of the codec (with a random interleaver between two encoders for self–orthogonal codes) corresponds to [3].

IV. Simulation Results

The plots in the Figure show the achieved bit error rates using up to 20 iterations over a Gaussian channel (code rate $\approx 1/2$, length of interleaver 9990, two component codes with $J = 3$). 16 $\boxplus$ operations and 8 additions are needed per information bit and iteration for calculating the soft output.



The "break" in the curves after enough iterations is the result of the weight distribution of the used feed–forward component codes [4].

Acknowledgements

The authors would like to thank Prof. J. Hagenauer for motivation and support.

References

[1] J. Hagenauer, "Soft is better than Hard," *Communications and Cryptography — Two Sides of One Tapestry,* edited by R. E. Blahut, D. J. Costello, Jr., U. Maurer, and T. Mittelholzer, pp. 155–171, Kluwer Academic Publishers, 1994.

[2] J. L. Massey, "Threshold Decoding," Cambridge, Ma, M.I.T. Press, 1963.

[3] C. Berrou, A. Glavieux, P. Thitimajshima, "Near Shannon limit error–correcting coding and decoding: Turbo–codes (1)," *International Conference on Communications (ICC'93),* Geneva, Switzerland, pp. 1064–1070, May 1993.

[4] Yu. V. Svirid, "Weight distributions of turbo–codes," accepted for *1995 IEEE Int. Symposium on Information Theory.*

An Efficient Reservation Connection Control Protocol for Gigabit Networks

Emmanouel A. Varvarigos and Vishal Sharma[1]
Department of Elect. & Comp. Eng., University of California,
Santa Barbara, CA 93106-9560, USA

Abstract — **The *Efficient Reservation Virtual Circuit* (or ERVC) protocol is a novel connection control protocol designed for constant-rate delay-insensitive traffic in gigabit networks. In the ERVC protocol, session durations are recorded and capacity is reserved only for the duration of the session, starting at the time it is actually needed. The protocol also has the "reservation ahead" feature, which allows a node to calculate the time at which the requested capacity will be available and reserve it in advance, thus avoiding wasteful repetition of the call setup phase. In addition, the protocol is robust to link and node failures, and allows soft recovery from processor failures.**

I. Introduction

The ERVC protocol is one of the two candidate protocols that we are considering for implementation in the 40 Gbit/s ATM-based fiber-optic Thunder and Lightning network currently being developed at UCSB. In designing the connection and flow control algorithms for this network our objectives were to ensure lossless transmission, efficient utilization of capacity, minimum pre-transmission delay for delay-sensitive traffic, and packet arrival in correct order. To meet these objectives, we have proposed the ERVC protocol for constant-rate traffic, and the Ready-to-Go Virtual Circuit (or RGVC) protocol for best-effort traffic and traffic with little delay tolerance. The RGVC protocol, described in [1], uses back-pressure and requires buffering at intermediate nodes, whereas the ERVC protocol, described in [2], uses reservations and requires little buffering at intermediate nodes.

II. Why the ERVC protocol ?

In standard reservation schemes (abbreviated SRVC) the capacity required by a session at an intermediate node is reserved starting at the time the setup packet arrives at the node. This is inefficient since the capacity reserved will actually be used at least one round-trip delay after the arrival of the packet at the node. This is because the setup packet has to travel from the intermediate node to the destination, an acknowledgement has to be sent to the source, and the first data packet of the session has to travel to the intermediate node. Over long transmission distances, the round-trip propagation delay may be comparable to, or even larger than, the holding time of a session. In particular, if a typical session requests capacity r bits/sec, and transfers a total of M bits over a distance of L kilometers, then the maximum percentage of time that the capacity is efficiently used in a SRVC protocol is

$$e = \frac{\frac{M}{r}}{\frac{2Lc}{\eta} + \frac{M}{r}}, \qquad (1)$$

where $c/\eta = 5\,\mu$s/km is the propagation delay in the fiber. Typical values of these parameters for the Thunder and Lightning network are $r = 10$ Gbit/s, $M = 0.5$ Gbit, and $L = 3000$ km (coast-to-coast communication), which yields $e = 0.625$. In contrast, the efficiency factor e for the ERVC protocol can be as large as $e = 1$, independently of the parameters r, L, and M.

The "reservation ahead" feature of the ERVC protocol allows sessions to reserve capacity in advance for use at a later time. Thus, if capacity is available for a session starting at a time that is within the delay that the session can tolerate, the call is accepted on its first attempt. This feature, therefore, avoids unnecessarily prolonged call setup phases, reduces a session's susceptibility to blocking, and leads to efficient utilization of the available capacity.

III. Basic description of the protocol

In the ERVC protocol, each network node keeps track of the *utilization profile* of each outgoing link, which describes the residual capacity available on the link as a function of time. The utilizatio profile is stored as a linked-list of records, and is updated efficiently. Each intermediate node reserves the required capacity starting at the time at which this capacity will actually be used (which is at least one round-trip delay after the arrival of the setup packet at the node), and for time equal to the session duration. If the session duration is unknown, it is treated as infinite, and capacity is reserved for that session for an unspecified duration (as in standard reservation schemes). If the capacity is not available at the time requested, the setup packet may make a reservation starting at the first time the capacity becomes available, if the session can tolerate the delay. Since, capacity is blocked for other sessions only for the duration of the call and is available for the remaining time, this allows a considerably greater number of sessions to be served. It also avoids the wasteful repetition of the call setup process, because it enables a session to reserve the required capacity in its first attempt, possibly at a time later than the requested time. If adequate capacity is available at every intermediate node, the source eventually receives an acknowledgement from the destination and begins transmitting data. If the time at which adequate bandwidth first becomes available is exceeds the delay tolerance of the session, the call is blocked and is reattempted later, probably via a different path. The ERVC protocol requires a pre-transmission delay at least equal to the round-trip propagation delay between the source and the destination (as all reservation protocols do).

[1]Research supported by ARPA under Contract DABT63-93-C-0039

References

[1] E. A. Varvarigos and V. Sharma, "An efficient reservation connection control protocol for gigabit networks," submitted *IEEE/ACM Trans. on Networking*, January 1995.

[2] E. A. Varvarigos and V. Sharma, "A loss-free connection control protocol for the Thunder and Lightning network," submitted *Globecom'95*, March 1995.

Guaranteeing Spatial Coherence in Real-time Multicasting

Max R. Pokam & G. Michel[1]
Laboratoire de Génie Informatique, Grenoble, France

Abstract — **We introduce the spatial coherence quality of service requirement for real-time point-to-multipoint communications in distributed systems. The notions of multicast end-to-end delay and global jitter are defined and their relationships with the spatial coherence are described.**

I. Introduction

As there is an increasing effort among communications system designers to provide communications applications with more and more elaborate services, coming to a real-time multicasting application in which a message sent from a source to a set of sinks is required to meet specified time and geographical (spatial) constraints, the underlying communications systems should allow spatial coherence quality of service requirement. We improve the *steadiness* and *tightness* metrics, defined as functions of maximum and minimum individual point-to-point delays [1]. to provide spatial coherence guarantee. The next section introduces the notions of *multicast end-to-end delay* and *global jitter* and then gives their relationships to the *spatial coherence*. In section III, three deterministic scheduling policies for point-to-point real-time communications are graded with respect to their suitability to spatial coherence.

II. Multicast End-to-end Delay, Global Jitter and Spatial Coherence

Given a data packet transmitted over a multipoint connection, the *multicast end-to-end delay* is defined as an N-dimensional vector $\vec{d} = (d_1, d_2, ..., d_N)$, where N is the number of elements in the recipient set, and d_i is the *ith* individual end-to-end delay. The scalar value of the multicast end-to-end delay is derived form the modulus of vector $\vec{d}$ as $d = \frac{1}{\sqrt{N}}\sqrt{\sum_{k=1}^{N}(d_k)^2}$. It is a scalar function of variables $d_1, d_2, ..., d_N$. The infinitesimal variation in the value of d is then derived as :

$$\delta d = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{d_k}{d} . \delta d_k \qquad (1)$$

In the above equation, the term δd_k of the righthand part is the individual delay jitter for sink k (j_k). The lefthand part, δd, is the global delay jitter that takes into account all the individual delay jitters of the multicast connection. It will further denoted as j_S. Equation 1 is then rewritten as $j_S = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \frac{d_k}{d} . j_k$. The spatial coherence is defined as a measure of the skew among the time instants at which a message transmitted over a real-time multicast connection is received at the different sinks. The spatial coherence is achieved when individual end-to-end delays have an equal value, in which case $fracd_k, d = 1$ for all k, $1 \leq k \leq N$. Hence, in order to guarantee spatial coherence, the ratio $fracd_k, d$ must be kept as close a possible to unity. In other words, the following double inequality should hold :

$$1 - \zeta \leq \frac{d_k}{d} \leq 1 + \zeta \qquad (2)$$

[1]This work was done in the framework of the IMAG project RACINES

Where ζ is a positive scalar very close to zero. From the above definition of the global jitter, and imposing a bound J_S on it, we derive equation 3.

$$\frac{1}{\sqrt{N}}(1-\zeta)\sum_{k=1}^{N} j_k \leq J_S \leq \frac{1}{\sqrt{N}}(1+\zeta)\sum_{k=1}^{N} j_k \qquad (3)$$

Assuming that ζ is close to zero, the above equation simplifies to:

$$J_S\sqrt{N} = \sum_{k=1}^{N} j_k \qquad (4)$$

From which the bounds on individual delay jitters can be solved.

III. Deterministic Scheduling Policies

We consider three deterministic scheduling policies for point-to-point real-time communications : 1)- the Earliest Due Date for Jitter (*EDD-J*) [2], 2)- the Stop & Go (*S & G*) [3] and, 3)- the Hierarchical Round Robbin (Φ]em HRR) [4]. Each mechanism is graded, in the range 0 to 3, according to three criteria: a)- the suitability to guarantee throughput or bit rate, b)- the suitability to guarantee end-to-end-delay and, c)- the suitability to spatial coherence as a result of the previous two criteria. The scores are presented in the following table.

	Throuput	*Delay*	**Spatial Coherence**
EDD-J	1	3	3
S & G	3	3	3
HRR	3	0	1

IV. Conclusion

From three examples of real-time point-to-point scheduling techniques, we showed how spatial coherence is achievable form the observance of individual end-to-end delay and jitter bounds. Thus the research results in real-time point-to-point communications can easily be extended to address the issue of spatial coherence quality of service requirement of real-time mult-casting applications. The case of statistical traffics and statistical multicast real-time requirements can be dealt with in an approach similar to the one we used for deterministic traffics and requirements.

References

[1] Sape Mullender (Editor). Distributed Systems. *Addison Wesley, Second Edition.*

[2] Dinesh Verma, Hui Zhang, and Domenico Ferrari. Delay Jitter control for real-time communication in packet switching network. In *IEEE Tricomm'91*, April 1991.

[3] S. J. Golsetani. Congestion-free transmission of real-time traffic in packet networks. In *IEEE INFOCOM'90*, pages 527-536, June 1990.

[4] C.R. Kalmanek, H. Kanakia, and S. Keshav. Rate controlled servers in very high speed networks. In *Globecom'90*, December 1990.

Peakedness of Stochastic Models for High-Speed Network Traffic

Brian L. Mark[1]
Dept. Elect. Eng., Princeton University,
Princeton, New Jersey, U.S.A.

David L. Jagerman, G. Ramamurthy
C&C Research Laboratories, NEC USA Inc.,
Princeton, New Jersey, U.S.A.

Abstract — ***Peakedness* was originally developed by teletraffic engineers as a tool for characterizing call arrival processes at a trunk group. We generalize the peakedness theory to include a class of stochastic models used in studies of high-speed networks and apply it to the approximate analysis of a statistical multiplexer.**

I. Introduction

In networks based on the Asynchronous Transfer Mode (ATM), information is transmitted asynchronously over high-speed links in the form of 53-byte units called *cells*. Accurate traffic characterization is a crucial step in performing network resource allocation and dimensioning.

II. Generalized Arrival Process

Define a *rate process* $\{R_t, t > 0\}$ to be a strictly stationary random process with finite, nontrivial first two moment measures. The process $\{R_t, t > 0\}$ is to be understood in the generalized function sense with the interpretation that $R_t dt$ represents the *amount* of work arriving in the infinitesimal interval $[t, t+dt)$. The *generalized arrival process* is then defined by

$$N_t = \int_0^t R_\tau d\tau, \tag{1}$$

where N_t represents the amount of work arriving in the interval $(0, t]$.

The standard arrival process defined as a *stationary point process* is a special case with

$$R_t = \sum_{i=1}^{\infty} b_i \, \delta(t - T_i), \tag{2}$$

where b_i is the number of arrivals at the ith arrival epoch, T_i, and $\delta(t)$ is the Dirac delta function. Another special case is the *discrete-level fluid process* with

$$R_t = \sum_{i=1}^{\infty} f_i \, \mathrm{rect}\left(\frac{t - T_i}{T_{i+1} - T_i}\right), \tag{3}$$

where f_i is the fluid flow rate, T_i is the epoch of the ith transition and $\mathrm{rect}(t) = u(t) - u(t-1)$, where $u(t)$ is the unit step function.

III. Generalized Peakedness

We introduce a concept of peakedness for a general arrival process as defined by (1). The arrival process is offered to an *infinite server system* which is represented by an i.i.d. process, $\{D_t, t > 0\}$, with marginal cdf F. Define

$$S_t = \int_0^t 1_{\{D_u > t-u\}} R_u du, \tag{4}$$

with the following interpretation: In the interval $[u, u + du)$, $R_u du$ units of work are offered to a new server, introduced at time u, which removes this work from the system after a duration D_u. Then S_t represents the amount of work present in the system at time t. The *peakedness functional* with respect to the service time cdf F is defined by

$$z[F] = \lim_{t\to\infty} \frac{\mathrm{Var}[S_t]}{E[S_t]}. \tag{5}$$

For the case of an orderly point process, the definitions (4) and (5) reduce to the standard concept of peakedness.

The following result of Eckberg [1] extends to our generalized notion of peakedness:

$$z[F] = 1 + \frac{\mu}{\lambda} \int_{-\infty}^{\infty} [k(x) - \lambda\delta(x)] F^*(x) dx. \tag{6}$$

Here, F^* is the autocorrelation function of $F^c = 1 - F$, $\mu^{-1} = \int_0^\infty F^c(x)dx$ is the mean service time, $\lambda = E[R_t]$ is the mean arrival rate, and $k(\tau) = \mathrm{Cov}(R_{t+\tau}, R_t)$ is the covariance function of the rate process.

IV. Application

The generalized peakedness can be obtained in closed form via (6) for a large class of stochastic traffic models, including the popular Markov modulated fluid models. In particular, the *peakedness function* of a Markov on-off fluid with peak rate r, mean *on* time β^{-1} and mean *off* time α^{-1} with respect to constant service time distribution is given by

$$z_{const}(\mu) = \frac{2r\beta}{(\alpha+\beta)^3}[\alpha + \beta + \mu(1 - e^{-(\alpha+\beta)/\mu})]. \tag{7}$$

Peakedness can also be estimated empirically through measurements of an actual traffic stream and then used to construct a stochastic traffic model.

Lee and Mark [2] propose a method for approximating a general arrival process with a more computationally tractable superposition of two types of on-off Markov fluid sources by matching central moments of the rate process R_t and an index of dispersion measure. Since the peakedness function contains strictly more information about the arrival process than the index of dispersion, a more accurate traffic characterization can be achieved by using the peakedness function (7) to perform the match. We demonstrate the effectiveness of our approach with an application to the analysis of a statistical multiplexer.

[1]The first author has been supported by an NSERC Postgraduate Scholarship.

References

[1] A. E. Eckberg, "Generalized Peakedness of Teletraffic Processes," *Proc. 10-th International Teletraffic Congress*, Montreal, Canada, 1983.

[2] H. W. Lee and J. W. Mark, "ATM Network Traffic Characterization Using Two Types of On-Off Sources," *INFOCOM '93*, pp. 152-159, 1993.

Fault Detection in Communication Protocols using Signatures

G. Noubir, K. Vijayananda, P. Raja
Swiss Federal Institute of Technology, Lausanne,
Computer Engineering Department, EPFL-DI-LIT,
noubir, vijay, raja@di.epfl.ch

Abstract — **Run-time fault detection in communication protocols is essential to detect faults that cannot be detected during the testing phase. In this paper, we use a polynomial-based signature function to detect run-time faults in communication protocols.**

I. Introduction

Signature Analysis [2] and FSM methods [4, 1] are two popular methods that are used to verify the control flow of programs. Run-time fault detection in communication protocols is essential to detect faults that arise due to coding defects, memory problems and external disturbances. In this paper, we summarize the results presented in [3]. We propose a new signature function which is based on polynomials, to detect run-time faults in communication protocols. Every state has a signature which represents the signature of all paths leading to that state and this is stored in a static signature table. The run-time path is transformed into a number (signature) using the signature function and compared with the static signature table for its correctness. While the FSM table has at least two dimensions, the static signature table has only one dimension.

II. Signature Generation

Let $A = (Q, \Sigma, \delta)$be a FSM with a state S_0 such that it has a predefined signature equal to zero and all the other states are reachable from S_0. The signature function is a polynomial with the values of states and events as coefficients and maps every path beginning at state S_0 into a value from an algebraic field F. For any two paths C_1 and C_2, the signature function must satisfy: $\exists p < 1; Prob[Signature(C_1) = Signature(C_2) | C_1 \neq C_2] < p$, where p is defined as the aliasing probability of the signature function. We use three kinds of signature depending on the availability of the state and event information. They are full-path, event, and state signatures. The polynomials associated with these signatures are given below. The signature is computed by evaluating the polynomial at a given point x_0.

Full-path: $P_C(x) = \sum_{i=0}^{n-1}(s_i x^{2(n-i)} + e_i x^{2(n-i)-1}) + s_n$

State: $P_C(x) = \sum_{i=0}^{n-1} s_i x^{n-i} + s_n$

Event: $P_C(x) = \sum_{i=0}^{n-1} e_i x^{n-i-1}$

where: s_i: state value, e_i: event value, n: length of the state path and x: a number from a given Galois field F. The following theorem gives an upper bound on the aliasing probability of the signature function.

Theorem 1 *Prob[Signature(C_1)=Signature(C_2) | $C_1 \neq C_2$] = $\frac{1}{|F|}$*

Corollary 1 *The probability that an illegal path is undetected is equal to* $\frac{1}{|F|}$

[0]This work was partially supported by the Swiss PTT project F& E N°309.

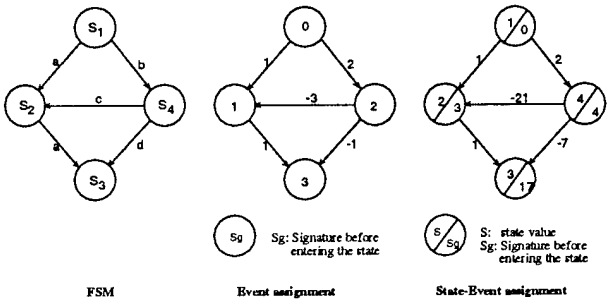


Fig. 1: Event-State assignment example.

When many paths lead to the same state, they are called *parallel paths*. Parallel paths should result in the same signature. This will reduce the complexity of signature verification. This constraint is used in generating the system of linear equations which can be used to assign values to the states and events (state-event assignment problem) [3].

III. Example

We explain the fault detection technique using the FSM shown in Fig. 1. Solving the state-event assignment problem for $x = 2$, we have $S_1 = 1$, $S_2 = 2$, $S_3 = 3$, $S_4 = 4$, $a = 1$, $b = 2$, $c = -21$, and $d = -7$. The initial state (S_1) has a signature value equal to 0. The signature is computed for $x = 2$. Consider the path $S_1 b S_4 c S_2 a S_3$. The step-wise computation of full-path signature is shown in Tab. 1.

Path	Computation	Signature
$S_1 b$	$(0 * 2 + 1) * 2 + 2$	4
$S_1 b S_4 c$	$((4 * 2 + 4) * 2 + (-21))$	3
$S_1 b S_4 c S_2 a$	$((3 * 2 + 2) * 2 + 1)$	17

Tab. 1: Full-path signature

IV. Conclusion

We have presented a signature-based method for detecting run-time faults in communication protocols. This technique has been applied to detect faults in protocols like ABP and TP4 [3].

Acknowledgements

We thank Prof. H. J. Nussbaumer for his valuable comments.

References

[1] A. Bouloutas G.W. Hart and M. Schwartz. Fault Identification Using a FSM Model with Unreliable Partially Observed Data Sequence. *IEEE trans. on comm.*, July 1993.

[2] Régis Leveugle. *Analyse de Signature et Test en Ligne Integré sur Silicium.* PhD thesis, Institut National Polytechnique de Grenoble, January 1990.

[3] G. Noubir and K. Vijayananda. Signature-based Fault Detection for Communication Protocols. Tech. Rep. 94/93, DI-LIT, Swiss Federal Institute of Technology at Lausanne, Dec. 1994.

[4] C. Wang and M. Schwartz. Fault Detection with Multiple Observers. *IEEE Trans. on Networking*, 1(1), 1993.

Sporadic Information Sources

Urs Loher
Signal and Information Processing Laboratory
Swiss Federal Institute of Technology
CH–8092 Zürich, Switzerland

Abstract — **Message arrivals encountered in digital transmission over most real communication channels are not independent but appear in clusters. We propose a model of such a bursty $K-$ary source using a Markov chain with two states. It is shown that the protocol information of this sporadic source can be drastically reduced on the one hand by not encoding intermessage information (e.g., the starting point of a packet) and on the other hand by buffering and reordering messages. Trade–offs between reduced protocol information and message delays are also considered.**

Summary

Messages such as commands, inquiries, file transmissions, and the like, traveling through a network, are extremely bursty. A model of a bursty $K-$ary source using a Markov chain with two states "quite" (or "idle") and "busy" (sometimes also called "active") is proposed as a sufficiently realistic model for many such sources. In the "quiet" state, the source transmits no (message) information, while in the "active" state, the source acts as a $(K-1)-$ary discrete memoryless source (DMS). The transition probabilities between states describe the sporadic nature of the source. Let p and q denote the probability of changing from the quiet to the busy state and from the busy to the quiet state, respectively. With this, the information rate in the steady–state, defined as the entropy per source letter [bits/time unit], U, can be calculated to be

$$H_\infty(U) = \underbrace{\frac{p}{p+q}\log(K-1)}_{\text{message information}} + \underbrace{\frac{p}{p+q}h(q)+\frac{q}{p+q}h(p)}_{\text{protocol information}} \quad (1)$$

where

$$h(p) = -p\log p - (1-p)\log(1-p)$$

is the binary entropy function.

Since each message symbol contains $\log(K-1)$ bits of information and since the source is producing message symbols during a fraction $p/(p+q)$ of time, the first term on the right side of (1) may be interpreted as the entropy of the messages. Similarly, the second term may be viewed as the entropy in the message length and the third term as the entropy in the length of the quiet periods. The information in the source output consists of two parts: a message part and a protocol part. Although such a separation seems reasonable intuitively, it is by no means entirely apparent that message information and protocol information can be separated completely from one another and considered independently. We show that a significant fraction of the channel capacity must be used for protocol information when either the expected message length is short ($q \gg 0$), or the quiet sequences are much longer than the message sequences ($q/p \gg 1$), or the signalling alphabet is small.

Whereas message information must generally be encoded losslessly, it is usually not necessary to encode all protocol information. For instance in a packet–switching network, messages are generally delayed by varying amounts in passing through the network in different ways and their arrival order may be changed. One can save protocol information by not resolving intermessage time delays. If we are not interested in "full–reconstructability" of the entire source output including the messages in their original order and/or the exact length of quiet periods, then we can use less than an average of $h(q) = H(L)/E(L)$ bits per message symbol to indicate the length L of the messages and/or less than $q/p \cdot h(p)$ bits per message symbol to indicate the lengths of the quiet periods. It is precisely the possibility of reordering and buffering the messages that permits a decrease in the amount of protocol information to be transmitted. We present both coding strategies that maintain messages in their original order going through the network and coding strategies that ignore message order. Unfortunately, the reduction in protocol information by the latter strategies is gained mostly at the cost of an enlarged message delay. One of the most important performance measures, however, is the average (or the maximum) delay required to deliver a message from the origin to the destination. We analyze the trade–off between the maximum tolerable delay and the amount of protocol information that must be sent. It is shown that the minimal necessary protocol information required to encode the message length decreases exponentially fast with increasing delay. Examples are given to illustrate and to compare the various strategies. Finally, certain generalizations of the concept of sporadic sources are devised for some related applications.

Acknowledgements

The author is very grateful to J.L. Massey for his help and many fruitful discussions.

References

[1] Gallager, R.G., *Information Theory and Reliable Communication,* John Wiley & Sons, Inc., 1968

[2] Bertsekas, D. and Gallager, R.G., *Data Networks,* Prentice–Hall, Inc., 1992

An Analysis Approach for Cell Loss Rate of Shared Buffer ATM Switching

Zhao Yu-biao, Yu Jian-ping, and Liu Zeng-ji
National Key Lab. of ISN, Xidian University, Xi'an 710071, P.R.China

Abstract — A novel approach is presented for analyzing cell loss rate of shared buffer ATM switching. It provides a new means to solve problems in more complex queueing system. It is an accurate alogrithm instead of conventional methods by employing a one-step transition matrix.

SUMMARY

ATM is a promising tansport and switch technique for a future B-ISDN. One of major areas under study of ATM switching system is switch architectures. Among various kinds of ATM architectures, shared buffer ATM switching is the best choice in terms of cell loss rate, throughput and swithching delay[1].

The relation between cell loss rate and shared buffer size is analyzed in some literatures. Those results are not accurate because the number of total cells that arrive at each time slot destined for the individual output ports are not independent. Since the total number of cells arriving at each time slot is no larger than switch input ports, those cells do not switch for the other output ports, if some cells destine for some certain output ports. The negative correlation causes the sum of the queues for the output ports to be stochastically smaller than what this sum would be were the queues to be independent. Based on this opinion, an accurate approach is developed for analysis of cell loss rate in shard buffer ATM switch. Outline of this analytic method is addressed as follows.

The shared buffer switch has N input ports and N output ports. At each time slot, cells arrive at each input link according to a Bernoulli process with probability $p < 1$. Each cell is uniformly to be destined for any of the N output ports. And, at each input ports, successive cells that do arrive are independently destined for their respective output ports.

Let a_i represent the probability of i arriving cells to the switching at each time slot. Based on the assumption above, a_i is a binomail -distributed. That is

$$a_i = C_N^i \; p^i (1-P)^{N-i}$$

It is assumed that the state of Markov chain is respresented by the number of cells in the switching. Then the probabilty transition matrix of arriving cells regardless of leaving cells is

$$P_a = \begin{bmatrix} a_0 & a_1 \cdots & a_N & & \\ & a_0 & a_1 \cdots & a_N & \\ & & \cdots & & \\ & & a_0 & a_1 \cdots & a_N \\ & & & \cdots & \end{bmatrix}$$

At each time slot, the number of cells which are transmitted to output links is that of output ports which have queueing cells. Let b_{ni} be the probability for i output ports which have queueing cells when the total number of cells is n in the switching. Therefore, the following equation can be derived by using Markov chain. That is

$$b_{ni} = \frac{\sum_{j=0}^{i-1} C_{i-1}^j (-1)^j (i-j)^{n-1}}{(i-1)!} \cdot \frac{N!}{N^n (N-i)!}$$

The leaving cells probability transitoin matrix regardless of arriving cells is

$$P_b = \begin{bmatrix} 1 & & & & \\ b_{11} & & & & \\ b_{22} & b_{21} & & & \\ & \cdots & & & \\ b_{NN} & b_{NN-1} & \cdots & b_{N1} & \\ & b_{NN} & b_{NN-1} & \cdots & b_{N1} \\ & & \cdots & & \end{bmatrix}$$

From the analyzing above, we can derive the realistic probability transtion matrix for the switching system. That is

$$P = P_b \cdot P_a$$

In order to solve the steady probability from this matrix, let the biggest number of state be large enough such that the difference, due to the finite state instead of the infinite state, is negligibly small, then the steady probability distribution can be easily obtained by formal ways. That is the accurate relation between cell loss rate and buffer size in the switching.

Reference

[1] Yasuro Shobatake,et al.,"A one-chip scalable 8*8 ATM switch LSI employing shared buffer architeture," IEEE J-SAC Vol.9, No. 8, Oct. 1991

A Scheme to Adopt Dynamic Selection of Error-Correcting Codes in Hybrid ARQ Protocol

Liu Zhong, Gu Xuemai, Guo Qing and Jia Shilou
Dept. Radio Eng., Harbin Institute of Technology

Harbin, P.R.China, 150001

Abstract — **Automatic Repeat_Request(ARQ) has been widely used for its high reliability and convenience in implementation. But its low performance is shown when channel is in noisy state. This paper presents an adaptive error control scheme with combination of FEC and ARQ. An encoder and a decoder for large constraint length convolution codes are constructed by TMS320E25 microprocessors to implement error control. Based on channel condition, the system can modify diffuse convolution codes constraint length automatically. Such an adaptive error control system combined with an ARQ system based on HDLC protocol is efficient to transmit data in high speed under bad radio channels.**

I .FEC/ARQ Adaptive Error Control Mode

The adaptive error control system block diagram is shown in Fig. 1, where the CCU(Communication Control Unit) implements HDLC protocol and system control.

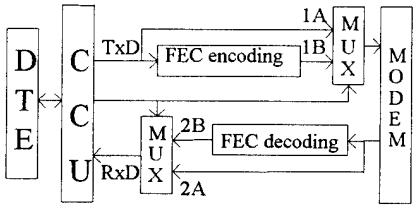


Fig. 1 System block diagram

Assuming the employed codes are denoted as C_1、C_2、$\cdots C_6$, where C_2、C_3、$\cdots$、C_6 are self-orthogonal diffuse convolution codes (2,1,4X+2), X is scaled as 32、64、128、256、512 and C_1 can be expressed as CRC implying that the transmitting data is encoded only by cyclic redundant check bits and ARQ protocol is employed. The C_2、C_3、$\cdots$、C_6 can correct 2X bits burst errors and 2 bits random errors. The rate is 0.5 only, and the decoding operation brings about delay. With the growth of the diffuse length, the decoding delay and the protection bits will rise, and the throughput will decrease. Thus, the focus problem is to determine dynamically which of the available codes to achieve the highest throughput for each channel status.

The system uses two frame structures: one is named as special frame, Fs, which begins with a special frame flag 0FFH, followed by diffuse length index; the other is named as data frame, F_d, which begins with a data frame flag 00H, and followed by encoded data

In the scheme, the channel condition is indicated by the probability of error frame. The procedure of the scheme can be described briefly as follow: suppose now that the C_k is used and the data packet(raw data) length is L bits, the transmitter counts the successful transmission frames per M frames(including the retransmission frames, but excluding Fs frames), let the result is denoted as V, if $V < N$(N is threshold), then the channel is in worse condition, and the transmitter attempt to adopt C_{k+1} and transmits a Fs frame to the receiver, then makes statistics of the successful transmission from the beginning; if $N < V < M$, C_k is suitable for the channel condition; if V=M, the transmitter checks whether or not the 3 × M transmission is successful without retransmission continuously, if not, C_k can be employed without changing, else the transmitter will attempt to adopt C_{k-1} and transmits a Fs frame to the receiver. In this way, the code can be selected dynamically to achieve the highest throughput.

II . The FEC Board and Operation Principle

The encoding and decoding is accomplished by a single board(FEC board) which contains two TMS320E25 and peripheral interface unit. Because the algorithm is executed by software, so the circuit is simple, the board size is small and it is convenient to change code from one to another. The 4K×16EPROM on chip is sufficient to contain five subroutines corresponding to the available codes C_2、$\cdots$、C_6. When CCU interrupts FEC board and then sends a code index to FEC board, the TMS320E25 executes the corresponding subroutine. The ARQ protocol is accomplished by CCU, simultaneously, the CCU makes statistics of successful frame and takes a selection of codes, and then conveys the index number of the selected code to FEC board.

The convolution code synchronization is achieved by use of frame synchronization. The Barker(11) is used as synchronization code, and five Barker(11) construct a synchronization code group to ensure at least one of the five codes not disturbed. The TMS320E25 on the FEC board makes correlation calculation to decide whether the receiving frame is in synchronization or not. Followed the synchronization code group, a NOT Barker(11) is arranged to indicate the end of synchronization head, the continued is encoded data. At the last part in transmitting frame, several protection bits is added to ensure the data remained in buffer(corresponding to shift-registers in hardware design) to be decoded completely .

III. Discussion and Testing Result

The system performance depends on the parameters L、M、N. With the growth of L, the probability of frame failing transmission will rise on fading channel condition. The larger M is , the slower the system sensitivity to channel condition is. The larger N(N < M) is, the more frequent code adjustment is. By practices, we have obtained some valuable data about the optimal parameters over mobile channel.

For testify the whole efficiency of this system, we make some practices on the following condition: Rc(channel data speed)=32Kb/s, L=1024, M=5, N=3. Let burst error probability be denoted as P_r, $Pr = \tau / 2T$, where τ : burst error lasting time, T : burst error appearing period. Let P_g express probability of random error and $P = P_r + P_g$ express probability of burst and random error combination. We transmitted a file sized 1920K bits in several simulative channels and obtained some practice data listing in Table 1、Table 2、and Table 3. In the following tables, T_1 expresses the consumed time in ARQ mode without error-correcting; T_2 expresses the consumed time in the mode described in this paper. From the result, it can be seen that the system performance is equivalent to ARQ system on unmixed burst error channel, while on other feature channels, the system is much superior to ARQ system .

Tabel 1 T=1S

τ	P_r	T_1(s)	T_2(s)
2	1×10^{-3}	80	80
5	2.5×10^{-3}	82	83
10	5×10^{-3}	83	83
30	1.5×10^{-2}	85	87

Table 2

P_g	T_1(s)	T_2(s)
1×10^{-4}	90	91
1×10^{-3}	375	141
5×10^{-3}	827	166
1×10^{-2}	∞	170

Table 3 T=1S

P_r	P_g	T_1(s)	T_2(s)
1×10^{-3}	1×10^{-4}	96	99
1×10^{-3}	1×10^{-3}	438	154
2.5×10^{-3}	1×10^{-3}	557	172
5×10^{-3}	1×10^{-2}	∞	188

References

(1) Sato T., Kawabe M., "Error-Free High Speed Data Transmission Protocol Simultaneously Applicable to Both Wire and Mobile Radio Channels", 38th IEEE Vehicular Technology Conference, pp 489 ~ 496, June, 1988.

(2) Fukasawa, "Adaptive Error Control Scheme for High Speed Data Transmission Through a Fading Channel", 36th IEEE Vehicular Technology Conference, pp 256 ~ 261, May, 1986.

Importance Sampling for TCM Scheme on Additive Non-Gaussian Noise Channel

Takakazu SAKAI and Haruo OGIWARA
Department of Electrical Engineering, Nagaoka University of Technology,
Nagaoka, 940-21, Japan

I. INTRODUCTION

Some error probability estimation methods of a trellis-coded modulation (TCM) scheme using importance sampling have been proposed [1]. However, these methods are not suitable for an additive non-Gaussian noise channel case. The main problem is how to design the probability density function in importance sampling. We propose a new design method of the probability density function related to the Bhattacharyya bound.

II. PROPOSED METHOD

Let s_1 and s_2 be transmitted signals, and r be the received signal. Now, we consider a decision system which decides that the transmitted signal is whether s_1 or s_2 from the received signal r. When the transmitted signal is s_1, the indicator function of the error region $\Phi(\cdot)$ is expressed as

$$\Phi(r) = \begin{cases} 1, & f(r|s_1) < f(r|s_2) \\ 0, & f(r|s_1) \geq f(r|s_2) \end{cases}, \qquad (1)$$

where $f(\cdot|\cdot)$ is the conditional probability density function. The ideal probability density function for importance sampling is propotional to $\Phi(r)f(r|s_1)$. The bound of the function $\Phi(\cdot)$ is very complex for most conditional probability density function cases. In Bhattacharyya bound, we evaluate the error probability from the upper bound of $\Phi(\cdot)$, that is, $\sqrt{f(r|s_2)/f(r|s_1)}$. The proposed probability density function $f^*(r|s_1)$ in importance sampling is designed almost the same idea with the Bhattacharyya bound and given by

$$f^*(r|s_1) \propto \sqrt{f(r|s_1)f(r|s_2)}. \qquad (2)$$

When the noise is an AWGN, the probability density function of the proposed method is reduced to that of mean translation method in [3]. The detail of the proposed method is in Ref. [5].

III. NUMERICAL EXAMPLE

A. NOISE MODEL

As an additive non-Gaussian noise model in the examle, an additive combination of an AWGN of variance σ_g^2 and an impulsive noise of Gaussian distribution of variance σ_i^2 which is observed with the probablity $\gamma(< 1)$ per symbol interval is used [4]. By taking the convolution of the two probability density functions, the probability density function of the additive non-Gaussian noise is rewritten as

$$f(x,y) = \frac{1-\gamma}{2\pi\sigma_g^2}\exp\left\{-\frac{x^2+y^2}{2\sigma_g^2}\right\} + \frac{\gamma}{2\pi(\sigma_g^2+\sigma_i^2)}\exp\left\{-\frac{x^2+y^2}{2(\sigma_g^2+\sigma_i^2)}\right\}. \qquad (3)$$

Since it is difficult to make random numbers following the probability density function designed by the proposed method, we approximate the probability density function $f^*(\cdot|\cdot)$ designed by the proposed method.

B. SIMULATION RESULTS

The encoder used in the example is (9, 2, 4) Ungerboeck code in [2]. As noise parameters, $\gamma = 0.01$ and $\sigma_i = 10\sigma_g$ were used. We selected 50 error events for the simulation based on the measure of the smaller Bhattacharyya distance. The number of simulation runs per error event were 1000. To compare with the proposed method, the ordinary Monte-Carlo simulation was tried. It was continued till 200 error bits were observed.

Figure 1 shows BER and variance performance. When BER $\leq 10^{-4}$, the proposed method approximates more than 95% of bit error rate of the ordinary Monte-Carlo method. The necessary CPU time of the proposed method is about 1/85 at BER of 10^{-6}. The variance of the simulation result of the proposed method is almost half of that of the ordinary Monte-Carlo method for all E_b/N_0. Under the condition of same variance, the reduction of simulation time of the proposed method is estimated about 1/170 at BER of 10^{-6}.

REFERENCES

[1] J.S. Sadowsky, "A New Method for Viterbi Decoder Simulation Using Importance Sampling," *IEEE Trans. Commun.*, pp. 1341-1351, Sep. 1990.

[2] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," *IEEE Trans. Inform. Theory*, pp. 55-67, Jan. 1982.

[3] D. Lu and K. Yao, "Improved Importance Sampling Technique for Efficient Simulation of Digital Communications Systems," *IEEE J. Select. Areas Commun.*, pp. 67-75, Jan. 1988.

[4] S.A. Kosmopoulos, P.T. Mathipoulos, and M.D. Gouta, "Fourier-Bessel Error Performance Analysis and Evaluation of M-ary QAM Schemes in an Impulsive Noise Environment," *IEEE Trans. Commun.*, pp.398-404, Mar. 1991.

[5] T. Sakai and H. Ogiwara, "Importance Sampling for TCM Scheme over Additive Non-Gaussian Noise Channel," to be appeared in *IEICE Trans. Fundamentals*, Sep. 1995.

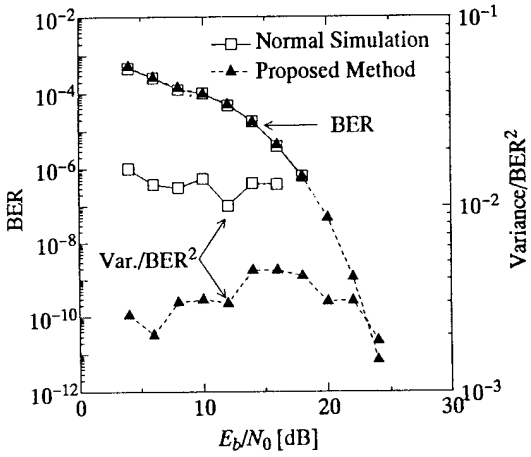


Fig. 1: BER and variance performance.

BRM sequence generators based on the field $GF(2^n)$ for DSP implementations

Sang-Jin Lee Seung-Cheol Goh[1] Kwang-Jo Kim Dai-Ki Lee

Electronics and Telecommunications Research Institute, 161 Kajong-Dong, Yusong-Gu, Taejon, 305-350, Korea

Abstract — This paper describes the extended LFSR(ELFSR) and the extended BRM(EBRM) based on the field $GF(2^n)$. We claim that those presented generators are efficient and suitable for S/W implementation. We also claim that the EBRM can be used as a good non-linear logic for stream cipher systems.

I. Introduction

A binary rate multiplier (BRM) sequence generator, consisting of two linear feedback shift registers(LFSRs) of length m and n respectively, has cryptographically good properties[1]. Under some constraints, it produces binary sequences of period $(2^m-1)(2^n-1)$ and linear complexity $m(2^n-1)$. The LFSR is well known to have good properties[2], however, it is not suitable for DSP implementation.

In this paper, we propose the extended LFSR(ELFSR) based on the field $GF(2^8)$, which can be efficiently and easily implemented by the general purposed DSPs. And then, we present the extended BRM(EBRM) sequence generator, which consists of two ELFSRs of length m and n respectively and based on the $GF(2^8)$. It produces byte sequences of period $(2^{8m}-1)(2^{8n}-1)$ and linear complexity $m(2^{8n}-1)$.

II. The Extended LFSRs

An ELFSR consists of m memory cells, which together form the state $(s_0, s_1, \cdots, s_{m-1})$ of the registers. The function $f(x)$ is mapping of $\{GF(2^n)\}^m$ to $GF(2^n)$.

$f(x) = c_0 \oplus (c_1 \otimes x) \oplus (c_2 \otimes x^2) \oplus \cdots \oplus (c_{m-1} \otimes s_{m-1}) \oplus x^m$

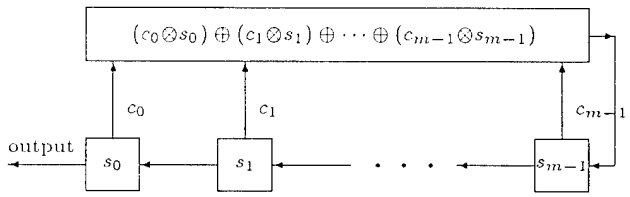


Fig 1. An ELFSR: $\oplus$ and $\otimes$ denote the operations of addition and multiplication, respectively, in the ground field $GF(2^n)$.

Property. The period of an ELFSR over $GF(2^n)$ with a primitive polynomial $f(x)$ of degree m is $2^{mn}-1$.

If we denote α and β in $GF(2^n)$ by $\alpha = (y_1, y_2, \cdots, y_n)$ and $\beta = (z_1, z_2, \cdots, z_n)$, then the addition of two elements is defined by $\alpha \oplus \beta = (y_1 \vee z_1, y_2 \vee z_2, \cdots, y_n \vee z_n)$, where $\vee$ means the XOR of two binary integers. Hence the operation $\oplus$ can be simply computed by the bitwise XOR of two binary blocks. However, in general, it is not easy to compute the multiplication of two elements in $GF(2^n)$. We adopt the method of multiplication introduced in [3].

Definition. A polynomial over $GF(2^n)$ is *simple* provided that all of its coefficients but the constant term are either 0 or 1.

Algorithm 1: The ELFSR

Input. A simple primitive polynomial $f(x)$ of degree m and two tables defined by the preprocessing. Let c[k] be the coefficients of $f(x)$ for all $0 \le k \le m-1$

Step 1. For $k = 0, \cdots, m-1$, initialize $s[k]$ by a random byte.

Step 2. Compute $\gamma = c[0] \otimes s[0]$.

Step 3. For $k = 1, \cdots, m-1$, if $c[k]$ is 1 then $t = t \oplus s[k]$.

Step 4. For $k = 1, \cdots, m-1$, set $s[k] = s[k-1]$. And then, set $s[0] = t$.

Step 5. Repeat Step 2 – Step 4 to produce sufficiently many random bytes.

III. The extended BRM

Now we present an extended BRM sequence generator, which consists of two extended LFSRs of length m and n respectively and based on the $GF(2^8)$. It produces byte sequences of period $(2^{8m}-1)(2^{8n}-1)$ and linear complexity $m(2^{8n}-1)$.

Algorithm 2: The extended BRM

Input. Two extended LFSRs SR1 and SR2 of length m, n respectively.

Step 1. Initialize all arrays of two ELFSRs by random bytes.

Step 2. At time $= t$, the two LFSRs are both clocked

Step 3. If the output of SR1 is odd, SR2 is then clocked one more time.

Step 4. Repeat Step 2 – Step 3 to produce sufficiently large number of random bytes.

IV. Concluding Remarks

In this paper, we proposed the ELFSR based on the field $GF(2^n)$, which can be efficiently and easily implemented by general purposed DSPs. And then, we presented the EBRM sequence generator, which consists of two ELFSRs of length m and n respectively and based on the $GF(2^8)$ so efficiently implemented by DSPs. We are now examining the security and efficiency of the proposed generators.

References

[1] W. G. Chambers and S. M. Jennings, *Linear Equivalence of Certain BRM Shift Register Sequences*, Electronics Letters, Vol. 20, No. 24, pp. 1018 – 1019, 1984.

[2] H. Beker and F. Piper, Cipher Systems: The Protection of Communications, New York: Wiley, 1982.

[3] Greg Harper, Alfred Menezes and Scott Vanstone, *Public-key Cryptosystems with Very Small Key Lengths*, Advances in Cryptology: Proc. Eurocrypt'92, Lecture notes in Computer Science 658, Springer-Verlag, pp. 163 – 173, 1993.

[1] e-mail: goh@dingo.etri.re.kr

Shift Register Synthesis For Multiplicative Inversion Over GF(2^m)

M. A. Hasan[1]

Elect. & Comp. Eng. Dept., Univ. of Waterloo, Waterloo, Ontario, Canada

I. Summary

Galois or finite fields have applications in cryptography and coding theory. For example, both encoding and decoding of Reed-Solomon codes require computations in the field over which the code is defined. Among the different arithmetic operations in finite fields, multiplicative inversion (hereafter called simply inversion) has been identified as the most complicated operation. Recently, several approaches have been made to compute the inverse efficiently. The approaches which have been given considerable attention in the literature are based on either Euclid's algorithm [1], or Fermat's theorem [2] or solution of a set of linear equations [3]. The latter approach is used in our present work to compute inverses.

Let $f(x) = \sum_{i=0}^{m} f_i x^i$ be a monic irreducible polynomial of degree m over GF(2) so that GF(2^m) = GF(2)[x]/$f(x)$. Let α be an element of GF(2^m) and satisfy $f(\alpha) = 0$. GF(2^m) can be viewed as a vector space of dimension m over GF(2) and the *canonical* basis $(1, \alpha, \cdots, \alpha^{m-1})$ is a vector A over GF(2^m). Let $M = [M_{i,j}]$ with

$$M_{i,j} = \begin{cases} f_{i+j+1} & 0 \le i+j \le m-1 \\ 0 & m \le i+j \le 2m-2. \end{cases} \quad (1)$$

Then $B = AM$ is the vector of the *triangular* basis corresponding to the canonical basis [4]. Any element $c \in$GF(2^m) can be written uniquely as $c = c_A A^T = c_B B^T$, where c_A and c_B being the vectors of *coordinates* of c with respect to the canonical and triangular bases, respectively.

Let a be any nonzero element of GF(2^m) and b be the inverse of a. Then it can be shown that

$$\sum_{i=0}^{m-1} s_{i+j} (b_A)_i = \delta_{j,m-1} \quad j = 0, 1, \cdots, m-1, \quad (2)$$

where $\delta_{i,j}$ is the Kronecker delta function which is equal to 1 when $i = j$ and 0 otherwise, and

$$s_i = \begin{cases} (a_B)_i & i = 0, 1, \cdots, m-1 \\ \sum_{j=0}^{m-1} s_{j+i-m} f_j & i = m, m+1, \cdots, 2m-2. \end{cases} \quad (3)$$

Let $h = b - \alpha^m$. Then it can also be shown that

$$\sum_{i=0}^{m-1} s_{i+j} (h_A)_i = \begin{cases} s_{j+m} & j = 0, 1, \cdots, m-2 \\ s^+_{j+m} & j = m-1, \end{cases} \quad (4)$$

where $s^+_{j+m} = s_{j+m} + 1$. Now the shift register synthesis algorithm of [5] can be used to solve (4) and hence to compute the inverse of a.

While the coordinates of a are taken with respect to the triangular basis, those of b are obtained with respect to the canonical basis. The use of these two bases has been exploited to realize efficient finite field arithmetic operations [4]. A basis change, if required, can however be performed using simple linear feed-back and feed-forward shift registers.

The area-time complexity for the inverter is $O(m^2 \log m)$. For an arbitrary field GF(2^m) the inverter has the least circuit complexity compared to the recently proposed ones, for example, [1], [2] and [3].

[1] This work was supported by an NSERC Research Grant

References

[1] K. Araki, I. Fujita, and M. Morisue. Fast Inverter over Finite Field Based on Euclid's Algorithm. *Trans. IEICE*, E 72(11):1230–1234, November 1989.

[2] C. C. Wang, T. K. Truong, H. M. Shao, L. J. Deutsch, J. K. Omura, and I. S. Reed. VLSI Architecture for Computing Multiplications and Inverses in GF(2^m). *IEEE Trans. Comput.*, C-34:709–717, August 1985.

[3] M. A. Hasan and V. K. Bhargava. Bit-Serial Systolic Divider and Multiplier for GF(2^m). *IEEE Trans. Comput.*, 41:972–980, August 1992.

[4] M. A. Hasan and V. K. Bhargava. Architecture for a Low Complexity Rate-Adaptive Reed-Solomon Encoder. To appear in IEEE Trans. Comput.

[5] J. L. Massey. Shift-Register Synthesis and BCH Decoding. *IEEE Trans. Inform. Theory*, IT-15:122–127, 1969.

On the Probability of Undetected Error and the Computational Complexity to Detect an Error for Iterated Codes

Toshihisa NISHIJIMA, and Shigeichi HIRASAWA

Abstract — **We discuss on practical and asymptotic capabilities of iterated codes used as error detecting codes. Throughout this paper, we assume that the codes are the binary linear block codes, and channel, the binary symmetric channel with cross-over probability ε.**

I. Iterated Codes

Let $\otimes$ be the direct product, then (N_O, K_O) iterated codes $C_I^{(s)}$ 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) code, and integer $s \geq 2$. The method for detecting any errors is the same method for correcting any errors of $C_I^{(s)}$. The decoding of the component code is only to detect any errors. If all syndrome of all component codes are zeros, the received sequence of length N_O is regarded as a transmitted codeword of $C_I^{(s)}$ and is accepted by the receiver. Under the below condition, $C_I^{(s)}$ are asymptotically bad codes.

Lemma 1 For $s \to \infty$, any $\epsilon > 0$, and some $J < i, j$, the necessary and sufficient condition to construct $C_I^{(s)}$ whose code rate R_O, $0 < R_O < 1$ is given by $|\frac{R_i}{R_j} - 1| < \epsilon$, where $R_i = \prod_{l=1}^{i} r_l$, $R_j = \prod_{l=1}^{j} r_l$, and $R_O = \frac{K_O}{N_O}$

II. Estimation Of Iterated Codes

Definition 1 We define the complexity of the operation required to detect an error by the product of the total number of shifts and the number of stages of the shift register to divide the polynomial of a received sequence.

[0]T. Nishijima is with Department of Industrial and Systems Engineering, College of Engineering, Hosei University, 3-7-2, Kajinocho, Koganei-shi, Tokyo, 184 JAPAN. E-mail nishi@nishi.is.hosei.ac.jp

[0]S. Hirasawa is with Department of Industrial Engineering and Management, School of Science and Engineering, Waseda University, 3-4-1, Ohkubo, Shinjuku-ku, Tokyo, 169 JAPAN.

Theorem 1 Let $\chi_I^{(s)}$ be the complexity of the operation required to detect an error for $C_I^{(s)}$. Then, $n_{\min}(N_O - K_O) < \chi_I^{(s)} < n_{\max}(N_O - K_O)$, where $n_{\max} = \max(n_1, n_2, \cdots, n_s)$, and $n_{\min} = \min(n_1, n_2, \cdots, n_s)$.

Corollary 1 For $C_I^{(s)}$ as $0 < R_O < 1$, and $s \to \infty$, we have $O(N_O) < \chi_I^{(s)} < O(N_O^2)$.

Let $P_I^{(s)}(\varepsilon)$ be the probability of undetected error for $C_I^{(s)}$. Then, by utilizing the structure of $C_I^{(s)}$ constructed by direct product of s codes c_i whose n_i is very small, comparing with N_O, we are able to calculate the exact value of $P_I^{(s)}(\varepsilon)$.

Theorem 2 By iterating the recurrent calculation until the stage $s-1$, finally we can have $P_I^{(s)}(\varepsilon) = [P_I^{(s-1)}(\varepsilon_{s-1})]^{L_s} - (1-\varepsilon_{s-1})^{N_s}$, where $P_I^{(s-1)}(\varepsilon_{s-1}) = \sum_{j=0}^{n_s} A_{sj} \varepsilon_{s-1}^{j} (1-\varepsilon_{s-1})^{n_s - j}$, A_{sj} is the number of codewords of Hamming weight j in code c_s, ε_{s-1} is the average error probability per bit at stage $s-1$, $N_s = n_s k_{s-1} \cdots k_1$, and $L_s = k_{s-1} k_{s-2} \cdots k_1$.

Corollary 2 For $0 < R_O < 1$, and $s \to \infty$, $P_I^{(s)}(\varepsilon_u) \to 0$.

III. Conclusion

The complexity of that for $C_I^{(s)}$ is more simple than that for the conventional single stage codes c under the same probability of undetected error, code length, and code rate. Also, the complexity of that for $C_I^{(s)}$ asymptotically is more simple than that for c.

The exact value of the probability of undetected error for $C_I^{(s)}$ can be always calculated. Furthermore, it is explicitly shown that the value of that for $C_I^{(s)}$ converges to zero for $s \to \infty$.

Wavefront Decoding of Trellis Codes

Torbjörn Larsson

National Semiconductor Corp., 2900 Semiconductor Dr., Mail Stop A1500, Santa Clara, CA 95052, USA

Abstract - **A novel reduced-complexity trellis decoding algorithm is described. The new algorithm, called Wavefront Decoding (WD), avoids the throughput bottleneck caused by metric and state-information feedback, which characterizes previously known breadth-first decoding algorithms. The error performance of WD for trellis-coded 8PSK on AWGN and Rayleigh fading channels is investigated by simulation. The results indicate that for a given number of survivor paths, the performance of WD is comparable, although necessarily inferior, to that of the M-algorithm. However, in contrast to the M-algorithm, WD exhibits a high degree of temporal parallelism, rendering it suitable for high speed applications.**

I. Introduction

The well-known M-algorithm [1] [2] is optimal in the sense that it minimizes, for any given number of survivor paths, the probability of rejecting the transmitted path [3]. However, the M-algorithm suffers from two structural deficiencies. First, the cost of survivor selection in terms of cycle and gate count will always be high. Second, due to the existence of a feedback loop in the decoder, in which the metrics and states of recursion n are fed back to be used in recursion $n+1$, the M-algorithm is incapable of simultaneously processing paths over several trellis stages. This excludes the use of the M-algorithm in high-speed applications, which require extensive pipelining. In this paper, we show that by generalizing the concept of breadth-first decoding, the feedback loop in the decoder may in fact be broken up to support pipelining over several trellis stages. Moreover, we find that for the decoding of short blocks, the survivor selection can be carried out at a cost significantly lower than in the M-algorithm. The price paid is a modest deterioration of error performance.

II. Wavefront Decoding

Consider first a breadth-first trellis decoder operating with C search paths selected from C state-classes. To proceed forward, the decoder first stores all successors of the old survivor paths in C lists associated with the C state-classes. Next, the best path from each list is extracted to become a new survivor. We shall refer to a group of C paths that propagate through the trellis in this fashion as a *wavefront*. Hence, in our notation the reduced-state sequence decoder (RSSD) considered in [3] and by several other authors is a single wavefront decoder. Consider next a decoder operating with $2C$ search paths divided in *two* wavefronts, each one consisting of C paths. The two wavefronts walk in file through the trellis, with the second one following immediately behind the first. To advance from time n to time $n+1$, the decoder first generates and stores all successors of the C paths in the first wavefront. C survivors are then extracted from the lists. These paths constitute the first wavefront at time $n+1$. Next, the successors of the C paths in the second wavefront are appended to the lists and C additional survivors are extracted to become the second wavefront at time $n+1$. Notice that the second wavefront selects its survivors both from its "own" successors and from those that were left over by the first wavefront. By introducing additional wavefronts in the same fashion, we obtain a decoder which, in the general case, operates with BC search paths divided into B wavefronts. We refer to this decoding principle as *Wavefront Decoding* (WD). Characteristic of WD is the fact that a wavefront, having arrived at stage n in the trellis, may directly select its survivors and then proceed forward to stage $n+1$ without waiting for the arrival of those paths that follow behind. Hence it can be seen that feedback of metrics and state-information only appears *internal* to each wavefront. The processing of the wavefronts may now be pipelined over several trellis stages to obtain a linear speedup.

Assuming that the correct path starts out in the first wavefront, it will eventually, as a result of channel noise, start to fall back in rank, from the first wavefront to the second, then to the third and so on, until it reaches the last wavefront where ultimate rejection awaits. The only way to escape from a certain loss of the correct path is to occasionally have the first wavefront stop and wait for the other wavefronts to arrive. Once the members of all waves have been accumulated in the C lists, B repeated selections are made from each list to produce B new wavefronts. The correct path now gets a chance to recapture its position in the first wave. Obviously, wavefront accumulation will reduce throughput, since the pipeline is broken up. Fortunately, it turns out that the time between accumulations L_A can be made fairly large without seriously degrading error performance. In particular, when data is encoded in short blocks (< 100 symbols), the accumulation of wavefronts need only be carried out at the end of the block. Notice that for the degenerate case $L_A = 1$, WD becomes the Generalized Viterbi Algorithm (GVA) [4].

The error performance of WD has been simulated for rate $2/3$ trellis-coded 8PSK on AWGN and Rayleigh fading channels. In all cases, $C = 4$ and $L_A = 64$ has been used. In general, it is observed that WD exhibits a certain performance degradation relative to GVA (with $C = 4$) and the M-algorithm with the same number of search paths. This is to be expected, since the selection of survivor combinations in WD (for $L_A > 1$ and $C > 1$) is more constrained than in the two other algorithms. However, in all cases considered here, the degradation is within a fraction of a dB.

References

[1] J. B. Anderson and S. Mohan, "Sequential Coding Algorithms: A Survey and Cost Analysis", *IEEE Trans. Comm.* vol. COM-32, pp. 169-176, Feb. 1984.

[2] T. Aulin, "A New Trellis Decoding Algorithm - Analysis and Applications", Tech. Report no. 2, Dept. of Information Theory, Chalmers University of Technology, Göteborg, Sweden, Dec. 1985.

[3] T. Larsson, "A State-Space Partitioning Approach to Trellis Decoding, Ph.D. Dissertation, Dept. of Computer Eng., Chalmers University of Technology, Sweden, Dec. 1991.

[4] T. Hashimoto, "A List-Type Reduced-Constraint Generalization of the Viterbi Algorithm", *IEEE Trans. Inf. Theory*, vol. IT-33, pp. 866-876, Nov. 1987.

† This work was supported by NUTEK, Sweden.

Potential-Decoding, Error Correction beyond the Half Minimum Distance for Linear Block Codes

Robert Löhnert
Daimler-Benz Aerospace, Sensorsysteme, Wörthstraße 85, 89070 Ulm, Germany

Abstract — **An error correction procedure for linear block codes is presented which corrects errors beyond the half minimum distance. The algorithm is based on minimizing a real valued function, called potential. Since the potential decreases monotonously with decreasing weight of the error vector, minimization of the potential can be done by local search.**

I. Introduction

Beneath the well known algebraic decoding for linear block codes there exist several non-algebraic approaches for error correction. In [1] a maximum-likelihood algorithm for linear block codes was shown which has exponential complexity. In [2] the minimum weight words are used as decoding vectors for binary codes.
The succeeding algorithm is applicable for all linear block codes and uses a statistical decoding approach based on the so called "potential".

II. Notation

The N-dimensional vector space over the q-element Galois field GF(q) will be denoted by $GF(q)^N$. For two vectors $\underline{a}$ and $\underline{b} \in GF(q)^N$ the inner product $S(\underline{a},\underline{b})$ is defined by $S(\underline{a},\underline{b}) := \sum_{i=0}^{N-1} a_i b_i$.

The code vectors of the Code **C** are denoted by $\underline{c}$, the error vector by $\underline{e}$ and the vectors of the dual code **C'** by $\underline{c}'$. Using $\underline{r} = \underline{c} + \underline{e}$ with $\underline{r}, \underline{c}, \underline{e} \in GF(q)^N$ and $S(\underline{c}, \underline{c}') = 0$, the $q^{N-K} - 1$ parity-check equations are defined by $A_j := S(\underline{r}, \underline{c}_j')$. Furthermore $wt(\underline{a})$ is the Hamming weight of $\underline{a} \in GF(q)^N$ and d_{min} the minimum distance.

III. Potential-Decoding

A model is presented which is capable of structuring the Galois field $GF(q)^N$. In this model a function is defined - called potential - which can be regarded as a measure for the distance of any vector to its nearest code vector. Various decoding algorithms can be derived from this model. The potential $U(\underline{r})$ of an arbitrary vector $\underline{r}$ is defined:

$$U(\underline{r}) := \sum_{j=1}^{q^{N-K}-1} \alpha_j \cdot I_j \qquad I_j := \begin{cases} 1, & \text{if } A_j \neq 0 \\ 0, & \text{if } A_j = 0 \end{cases} \tag{1}$$

I_j stands for an indicator variable for the parity-check equation A_j. α_j is a weighting factor which depends on the parity-check vector $\underline{c}_j'$. The characteristics of the potential $U(\underline{r})$ are:

$$U(\underline{c}) = 0 \tag{2}$$

$$U(\underline{r} \notin \mathbf{C}) > 0 \tag{3}$$

$$U(\underline{r}) = U(\underline{c} + \underline{e}) = U(\underline{e}) \tag{4}$$

$$U(\underline{e}_2) < U(\underline{e}_1), \quad \text{if} \quad wt(\underline{e}_2) < wt(\underline{e}_1) < d_{min}/2, \tag{5}$$

Although eq. (5) holds only up to $d_{min}/2$ it can be shown that statistically this property is valid up to considerably higher error numbers. Assuming that $\alpha_j \in \mathbf{R}$ is only dependent on the weight $L = wt(\underline{c}_j')$ of the vector $\underline{c}_j'$ gives:

$$\alpha_L := \alpha(\underline{c}_j' \mid wt(\underline{c}_j') = L). \tag{6}$$

With this assumption $U(\underline{r})$ is separable into subpotentials $U_L(\underline{r})$. Every subpotential $U_L(\underline{r})$ consists of the m_L parity-check vectors of weight L.

$$U_L(\underline{r}) = \alpha_L \sum_{j=1}^{m_L} I_j, \qquad U(\underline{r}) = \sum_{L=0}^{N} U_L(\underline{r}). \tag{7}$$

It can be shown [3], that the mean value of U_L is given by:

$$\overline{U_L(\underline{e} \mid wt(\underline{e}) = t)} = m_L \frac{q-1}{q}\left[1 - \left(1 - \frac{q}{q-1}\cdot\frac{L}{N}\right)^t\right] \tag{8}$$

For efficient decoding it is not necessary to use all $q^{N-K} - 1$ parity-check equations. Table 1 shows the decoding performance for several codes using only the two subpotentials U_L with the maximum and minimum weight vectors.

Error Number t	(31,11,11) BCH-code	(63,24,15) BCH-code	(113,57,15) QR-code
≤ 5	0 %	0 %	0 %
6	33.3 %	0 %	0 %
7	87.3 %	0 %	0 %
8	-	1.5 %	0.06 %
9	-	9.5 %	0.6 %
10	-	36.5 %	1.8 %
11	-	72.2 %	12.2 %
12	-	95.0 %	29.2 %

Table 1: Percentage of decoding errors of weight t.

Figure 1 shows the performance of decoding with the subpotential U_{98} for a (113,57,15) QR-code compared with Bounded Minimum Distance (BMD) decoding and with a rate 1/2 convolutional code (K=7) with Viterbi decoding. Potential-decoding is very well suited to implementations into VLSI. To reach the decoding performance of U_{98}, only 50 000 gates of an ASIC are necessary, up to data rates of approximately 10 Mbit/s.

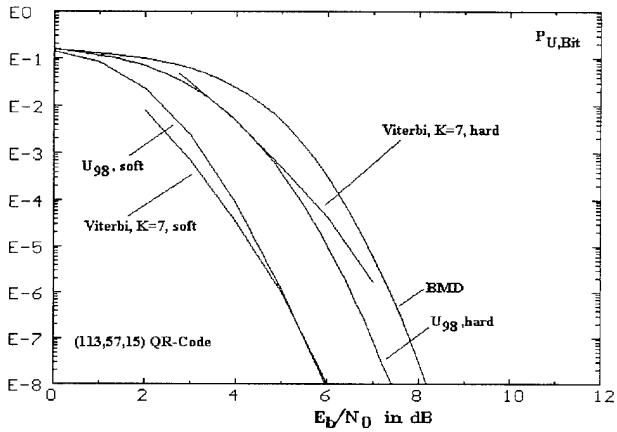


Figure 1 Bit error probability of U_{98} for the (113,57,15) QR-code on an AWGN-channel with BPSK-modulation.

References

[1] **Hartmann Carlos R. P., Rudolph Luther D.:** "An Optimum Symbol-by-Symbol Decoding Rule for Linear Codes", IEEE Trans. Inf. Theory Vol. 22, Sept. 1976.

[2] **Bossert M., Herget F.:** "Hard- and Soft-Decision Decoding beyond the Half Minimum Distance - An Algorithm for Linear Codes", IEEE Trans. Inf. Theory, No. 5, Sept., 1986.

[3] **Löhnert R.:** "A Model to Structure the N-dimensional Vector Space of Block Coding and Decoding Procedures for Error Correction beyond the Half Minimum Distance", Archiv für Elektronik und Übertragungstechnik (AEÜ), vol. 48, Heft 4, Stuttgart, Juli 1994.

Information Set Decoding Complexity for Linear Codes in Bursty Channels with Side Information[1]

Wonjin Sung and John T. Coffey
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109

Abstract — **General decoding algorithms for linear codes that have less complexity than exponential search have been studied by many researchers and exact complexities are known for the memoryless channel [1-4]. Among the various decoding strategies for linear codes, the information set decoding algorithm has complexity that is significantly lower than that for most other general algorithms over most code rates [3,4]. It is the purpose of this paper to derive the complexity for information set decoding used in channels where errors may occur in bursts, and to quantify the gain in complexity over the memoryless channel case.**

I. Introduction

Errors encountered in many communication channels are not independent but appear in bursts. One way to effectively model bursty channels is to assume that the channels have two states with different probabilities of channel error [5]. The channels we consider have probability π_g to be in the *good state* and probability $\pi_b(=1-\pi_g)$ to be in the *bad state*. The error probability for the good state is assumed to be r times the error probability for the bad state, where $0 \le r \le 1$. The Gilbert-Elliott channels are described by a two-state Markov chain model, and state transitions depend on the transition probabilities. We define the *complexity exponent* $F(R)$ of decoding algorithms for binary linear codes of rate R as

$$F(R) = \lim_{n\to\infty} \frac{1}{n} \log_2 M(n, R)$$

where $M(n, R)$ is the number of computations necessary.

II. Information Set Decoding on Bursty Channels

For the bursty channel with deterministic state transitions, the complexity exponent $F_D(R)$ of the information set decoding that gives error probability no greater than twice the error probability of maximum likelihood decoding is given by

$$F_D(R) = (1-R) - (1-R)H\left(\frac{\pi_b p + \pi_g r p}{1-R}\right)$$

when $R \ge \pi_g(1-r)$, and

$$F_D(R) = \pi_g H(rp) - (\pi_g - R)H\left(\frac{\pi_g r p}{\pi_g - R}\right)$$

otherwise, where $H(\cdot)$ is the binary entropy function and p is the value satisfying

$$1 - R = \pi_b H(p) + \pi_g H(rp).$$

The obtained complexity is shown to be strictly less than the complexity exponent for the memoryless channel for the entire range of code rates and channel parameters π_b, π_g, and r. When $r = 1$, $F_D(R)$ becomes identical to the complexity exponent for the memoryless channel. The gain in complexity gets larger as r gets closer to 0, i.e., when the channel error probabilities for two states differ by a larger amount. The optimal way to select information sets is to choose βnR bad state symbols and $(1-\beta)nR$ good state symbols, where β is given by

$$\beta = \frac{\pi_b(R - \pi_g(1-R))}{(\pi_b + \pi_g r)R}$$

when $R > \pi_g(1-r)$, and $\beta = 0$ otherwise. Bounds on the complexity exponent $F_{GE}(R)$ for Gilbert-Elliott channels can be achieved by modifying the result for the channel with deterministic state transitions. We obtain

$$F_D(R) \le F_{GE}(R) \le F_D(R) + \Delta(b, g)$$

where b is the transition probability from the good state to the bad state, g is the transition probability from the bad state to the good state, and $\Delta(b, g) = H(\frac{bg}{b+g}) + \frac{b}{b+g}H(g)$. The bounds become tight when the channel transitions take place slowly; we have $\Delta(b, g) \approx 0$ for small b and g.

It is possible to improve the bounds for $F_{GE}(R)$ when side information such as soft-decision information is available to the decoder. The extra complexity in the upper bound on $F_{GE}(R)$, when compared to $F_D(R)$, is due to the state estimation of the received sequences. By using soft-decision information, we can effectively estimate which symbols are transmitted through either the good state or the bad state. One scheme for state sequence estimation is to choose the $n\pi_g$ most reliable symbols out of a given sequence of n symbols, and assume that these are the symbols that have been transmitted through the good state. For the Gilbert-Elliott channel with soft-decision information available, we achieve the complexity exponents very close to $F_D(R)$ even when state transitions occur frequently.

[1]This work was supported in part by NSF Grant NCR-9115969

References

[1] G. S. Evseev, "Complexity of decoding for linear codes," *Probl. Peredach. Inform.*, vol. 19, pp. 3–8, 1983.

[2] I. I. Dumer, "Two decoding algorithms for linear codes," *Probl. Peredach. Inform.*, vol. 25, pp. 24–32, 1989.

[3] E. A. Kruk, "Decoding complexity bound for linear block codes," *Probl. Peredach. Inform.*, vol. 25, pp. 103–107, 1989.

[4] J. T. Coffey and R. M. Goodman, "The complexity of information set decoding," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 1031–1037, 1990.

[5] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, pp. 724–744, 1978.

When is Hard Decision Decoding Enough?

Peter F. Swaszek
Department of Electrical and Computer Engineering, University of Rhode Island
Kingston, Rhode Island, 02881

Abstract — **Many algorithms for soft and near-soft decision decoding of block codes start by implementing hard decision decoding. In several instances it has been noted that simple tests of the hard decision result may allow the algorithm to terminate at this point. This paper explores this notion in detail.**

I. Introduction

Consider the problem of decoding an (n, k, d_{min}) binary block code with codewords $\mathbf{c}_j$. Assume that antipodal signaling, $\mathbf{s}_j = \sqrt{E}(2\mathbf{c}_j - \mathbf{1})$, and additive Gaussian noise (zero mean, variance σ^2) produce the channel observation

$$\mathbf{x} = \mathbf{s}_j + \mathbf{n}.$$

To minimize the probability of error the two standard decoding techniques are *soft decision* and *hard decision* decoding (with resulting codewords $\mathbf{c}_s$ and $\mathbf{c}_h$ and performances P_e(soft) and P_e(hard), respectively). Soft decision decoding, while providing optimum performance, is computationally burdensome. Hard decision decoding has a significantly reduced implementation complexity at reduced performance. During the last 30 years many authors have searched the middle ground for high performance, low complexity approaches.

Many of these approaches start with hard decision decoding, searching the nearby codespace for a best choice of codeword. It has been noted that such algorithms can terminate early if the data $\mathbf{x}$ and the hard-decision result $\mathbf{c}_h$ together satisfy certain conditions. We envision, then, a decoder with operation:

1. Hard-decision decoding is implemented yielding $\mathbf{c}_h$.
2. A test is performed to see $\mathbf{c}_h$ matches $\mathbf{c}_s$ (without, of course, directly finding $\mathbf{c}_s$). If the answer is yes, the decoding algorithm terminates at this point.
3. If the test of step 2 fails, full soft decision decoding or some other strategy is implemented.

Without actually implementing soft decision decoding, the test in step 2 has three possible answers: *yes*, *no*, and *the data is inconclusive*. A "yes" response is called a *success* for the test; conversely, a "no" or "data inconclusive" response is a *failure* in that additional processing would be required before decoding is complete.

The motivation for such tests is that since hard decision decoding is correct a relatively high percentage of the time, it often matches the soft decision decoding result exactly. This idea can be made more mathematically formal. Specifically, it can be shown that

$$P_e(\text{hard}) - P_e(\text{soft}) \le \Pr(\mathbf{c}_h \ne \mathbf{c}_s) \le P_e(\text{hard}) + 2P_e(\text{soft}).$$

Since P_e(soft) is typically much smaller than P_e(hard), then $\Pr(\mathbf{c}_h \ne \mathbf{c}_s) \approx P_e(\text{hard})$. Thus, the failure probability for *any* test for step 2 is approximately lower bounded by P_e(hard). An efficient test should fail only about as frequently as hard decision decoding makes an error.

As an example of a test of $\mathbf{c}_h = \mathbf{c}_s$, consider the following well known condition:

The Codeword Test – *If the hard decision decoder's input is already a codeword, then* $\mathbf{c}_h = \mathbf{c}_s$.

Unfortunately, this result is far from the lower bound on the failure probability. Several tests for step 2 are described below with emphasis on the coherent Gaussian channel. Additional details, including tight upper and lower bounds to performance for these tests, are presented in [3].

II. Tests for the AWGN Channel

The first test has been mentioned previously [2]:

The Hypersphere Test – *If* $\mathbf{x}$ *is within* $\sqrt{d_{min}E}$ *units (in Euclidean distance) of the hard decision decoded signal then* $\mathbf{c}_h = \mathbf{c}_s$.

Realizing that the actual soft decision decoding region is a convex cone, the test region can be expanded from a hypersphere to the circumscribing right circular cone:

The Circular Cone Test – *If* $\mathbf{x}$ *satisfies*

$$\frac{\mathbf{x}(2\mathbf{c}_h - \mathbf{1})^T}{\sqrt{n\mathbf{x}\mathbf{x}^T}} \ge \sqrt{\frac{n - d_{min}}{n}} \qquad (1)$$

then $\mathbf{x}$ *falls within the aforementioned cone and* $\mathbf{c}_h = \mathbf{c}_s$.

While the cone test completely encloses the hypersphere test, the cone and codeword tests do have different support; hence, it seems reasonable to combine them:

The Combined Test – *If the hard decision decoder's input is already a codeword or if the received vector* $\mathbf{x}$ *satisfies the cone inequality in (1) then* $\mathbf{c}_h = \mathbf{c}_s$.

Algebraic analysis of the soft decision decoding operation [1] yields a further test:

The Polygonal Cone Test – *Define* z_i, $i = 1, 2, \ldots n$, *by*

$$z_i = \begin{cases} +x_i & \text{if } c_{h,i} = 0 \\ -x_i & \text{if } c_{h,i} = 1 \end{cases}$$

If the sum of the d_{min} *largest* z_i *does not exceed zero then* $\mathbf{c}_h = \mathbf{c}_s$.

This set subsumes all of the above tests with some increase in complexity. The resulting performance can be quite good.

References

[1] H. T. Moorthy, *Decoding of Linear Block Codes*, MS thesis, Dept. Elect. Eng., Univ. Rhode Island, 1992.

[2] O. O. Olaniyan, "Implementable soft decision decoding schemes," *Int'l. Jour. Electronics*, 66(3), pp. 321–332, March 1989.

[3] P. F. Swaszek, "When is hard decision decoding enough?" submitted to *IEEE Trans. Inform. Theory.*

First Order Approximation of the Ordered Binary Symmetric Channel

Marc P.C. Fossorier and Shu Lin[1]

Dept. of Electrical Engineering, University of Hawaii, Honolulu, HI 96822, USA.

Abstract — **In this paper, different results related to the ordering of a sequence of N received symbols with respect to their reliability measure are presented for BPSK transmission over the AWGN channel.**

I. Approximation of $\mathrm{Pe}(n_1, \cdots, n_j; N)$

For BPSK transmission over the AWGN channel, many maximum likelihood decoding (MLD) algorithms of binary linear block codes first reoder the received symbols within each block with respect to their reliability. In [1], the statistics of the noise after ordering are derived. These statistics allow to tightly bound the error performance of any suboptimum algorithm based on reordering.

After ordering a sequence of N symbols, the probability $\mathrm{Pe}(n_1, \cdots, n_j; N)$ that an error occurs at positions $n_1, \cdots, n_j$ can be computed exactly. However, no close form solution has been found for $N \geq 3$. This is mostly due to the fact the the noise for which the statistics are derived is not the ordered random variable. Based on the central limit theorem, we show in this paper that for N large enough, the distribution of $\tilde{W}_i$, the restriction of the i^{th} ordered noise value to the interval $[1, \infty)$, is well approximated by the distribution of a normal random variable that we specify. This approximation leads to

$$\mathrm{Pe}(i; N) \cong \mathrm{e}^{\frac{4(1-m_i)}{N_0}}, \tag{1}$$

where $m_i = \alpha^{-1}(1 - i/N)$, after defining, for $n \geq 1$, $\alpha(n) = \tilde{Q}(2-n) - \tilde{Q}(n)$, with the normalization $\tilde{Q}(x) = (\pi N_0)^{-1/2} \int_x^\infty \mathrm{e}^{-n^2/N_0} dn$. When N is large enough, Equation 1 provides a tight bound.

If W_i and W_j represent the i^{th} and j^{th} ordered noise values, it is possible to show that $W_j | W_i$ has the density function of the $(j-i)^{th}$ noise value after ordering a sample of size $N - i$ from a population with distribution truncated to the interval $[m(w_i), M(w_i)]$, where $m(w_i) = \min(2 - w_i, w_i)$ and $M(w_i) = \max(2 - w_i, w_i)$. Combining this result with Equation 1, we show that, for $i < N$,

$$\mathrm{Pe}(i, j; N) \cong \left(\frac{N}{N-i}\right) \mathrm{Pe}(i; N)\, \mathrm{Pe}(j; N). \tag{2}$$

Generalizing Equation 2 to any ordered set of indices $I_j = \{n_1, \cdots, n_j\}$ corresponding to positions in error after ordering, we compute, based on a chain argument,

$$\mathrm{Pe}(n_1, \cdots, n_j; N) \cong \prod_{l=1}^{j-1} \left(\frac{N}{N-n_l}\right) \mathrm{Pe}(n_l; N) \cdot \mathrm{Pe}(n_j; N). \tag{3}$$

Therefore, despite the fact that the random variables representing the noise after ordering are dependent, their associated error probabilities tends to behave as if they were independent, for $N >> n_{j-1}$ and large enough.

[1]This work was supported by NSF Grant NCR-91-15400

II. First Order Approximation of the Ordered Binary Symmetric Channel

The value $\mathrm{Pe}(n_1, \cdots, n_j; N)$ represents the probability that at least the bits in position $n_1, \cdots, n_j$ are in error after ordering a sequence of length N. We now also define $\mathrm{Pe}_N(n_1, \cdots, n_j)$ as the probability that only the bits at position $n_1, \cdots, n_j$ are in error after ordering a sequence of length N. While $\mathrm{Pe}(n_1, \cdots, n_j; N)$ is computed by integrating the joint distribution of the n_j ordered random variables $W_{n_1}, \cdots, W_{n_j}$, the computation of $\mathrm{Pe}_N(n_1, \cdots, n_j)$ requires to integrate the joint distribution of the N ordered random variables $W_1, \cdots, W_N$. It follows that the discrete time channel model after ordering is a 2^N-state BSC with transition probabilities $\mathrm{Pe}_N(n_1, \cdots, n_j)$'s. We refer this channel as the Ordered BSC (OBSC). Based on Equation 3, we approximate

$$\mathrm{Pe}(n_1, \cdots, n_j; N) \cong \prod_{l=1}^{j} \mathrm{Pe}(n_l; N), \tag{4}$$

which expresses that after ordering, the events of having errors at positions $n_1, \cdots, n_j$ remain independent. Therefore, the 2^N-state fully connected OBSC is equivalent to N time-shared BSC's corresponding to each ordered position. We name this approximation **the first order approximation of the OBSC**.

The capacity of the OBSC $C_{N,ave}$ requires the computation of 2^N N-order integrals and rapidly becomes too complex to evaluate as N increases. In contrast, the capacity of the first order approximation of the OBSC

$$\tilde{C}_{N,ave} = 1 - \frac{1}{N} \sum_{i=1}^{N} h(\mathrm{Pe}(i; N)) \quad \text{bit} \tag{5}$$

is easily derived. For $N = 1$, $\tilde{C}_{1,ave}$ is simply the capacity of the BSC with crossover probability $\tilde{Q}(1)$, while $\lim_{N\to\infty} \tilde{C}_{N,ave}$ should provide the capacity C_{bpsk} of the continuous Gaussian channel for BPSK transmission. We observe that $C_{N,ave} \approx \tilde{C}_{N,ave}$ and that the convergence to this limit is very fast as N increases, so that

$$C_{N,ave} \approx \tilde{C}_{N,ave} \approx C_{bpsk}, \tag{6}$$

for N large enough. Equation 6 indicates that when considering an ordered sequence of sufficiently long N, the first order approximation of the OBSC should provide a good approximation of the continuous Gaussian channel, for BPSK transmission. Therefore, for a given SNR, knowing the position in the ordering instead of the exact received value should be sufficient from a performance point of view.

References

[1] M. P. C. Fossorier and S. Lin, "Soft-Decision Decoding of Linear Block Codes based on Ordered Statistics," *IEEE Transactions on Information Theory*, to appear.

An Asymptotic Evaluation on the Number of Computation Steps Required for the Nearest Point Search Over a Binary Tree

Hisashi Suzuki
Dept. Info. & Syst. Eng., Central University
1-13-27 Kasuga, Tokyo 112, Japan

Suguru Arimoto
Dept. Math. Eng. & Info. Physics, University of Tokyo
7-3-1 Hongo, Tokyo 113, Japan

Abstract — **This paper analyzes the number of computation steps on a binary tree searching fast for one in some beforehand-given points that is the nearest to a query point in a Hamming space.**

I. Introduction

$\{0,1\}^l$ denotes the whole set of binary sequences (called points) of a length $l \geq 2$. We measure the distance between points by the Hamming distance normalized by l.

Suppose that $n \geq 2$ arbitrary points $x_1, \cdots, x_n$ (called samples) in $\{0,1\}^l$ are given, where duplications are allowed. We consider arranging the samples into a binary tree and, over it, searching fast for some $\hat{x} \in \{x_1, \cdots, x_n\}$ that is the nearest to any querried point $x \in \{0,1\}^l$.

The authors' last paper [2] mentioned a KM tree [1] that could search for the nearest point fast but the search time was neither clear in theoretical nor in experimental. The present paper evaluates theoretically the number of computation steps required for the nearest point search over an alternative tree.

II. Tree Construction

Fix a real constant $\gamma > 0$ called a stopping threshold. Given an arbitrary sequence $\mathbf{x} = x_1 \cdots x_n$ of n samples, the following procedure constructs a binary tree $\boldsymbol{T}_\gamma(\mathbf{x})$ each leaf of which stores at most γn samples.

Procedure 1 (tree construction procedure):
Step 1: Construct a tree comprising only a root that stores $\mathbf{x}$. (Regard this root also as a leaf to start (a)-(b).)
Step 2: While the present tree has at least one leaf N storing a sequence $\mathbf{z} = z_1 \cdots z_{|\mathbf{z}|}$ of points such that $|\mathbf{z}| \geq \gamma n$ and all of $z_1, \cdots, z_{|\mathbf{z}|}$ are not the same, do (a)-(b). Otherwise, answer the present tree.
(a) Let $c = z_1$. Discover one of r-values that make $|\mathbf{z}_\mathrm{L}|$ and $|\mathbf{z}_\mathrm{R}|$ as equal as possible, where $\mathbf{z}_\mathrm{L}$ and $\mathbf{z}_\mathrm{R}$ denote the sequences composed of ξ_is respectively for which $d(c,\xi_i) \leq r$ and $d(c,\xi_i) > r$.
(b) Store (c,r) on N. Store $\mathbf{z}_\mathrm{L}$ and $\mathbf{z}_\mathrm{R}$ respectively on the left and right child nodes of N. Next, remove $\mathbf{z}$ from N. □

III. The Nearest Point Search

Let an arbitrary subtree $\boldsymbol{T}^*$ of $\boldsymbol{T}_\gamma(\mathbf{x})$ whose vertex is a nonleaf or leaf node of $\boldsymbol{T}_\gamma(\mathbf{x})$ be given with a supposition that the total length of $\mathbf{z}$s stored on all leaves of $\boldsymbol{T}_\gamma(\mathbf{x})$ is $\leq 1/\gamma$. Let $\mathcal{S}$ denote the set of all points stored on leaves of $\boldsymbol{T}^*$. Fixed a real constant $\Delta \geq 0$ called pre-bounding parameter, the following recursive procedure $f_\Delta(\boldsymbol{T}^*, x)$ for an arbitrary query point $x \in \{0,1\}^l$ tries to answer one of points $\hat{y}$s in $\mathcal{S}$ that achieve $d(x,\hat{y}) = \min_{y \in \mathcal{S}} d(x,y) \leq \Delta$.

Procedure 2 (search procedure $f_\Delta(\boldsymbol{T}^*, x)$):
Step 1: If $\boldsymbol{T}^*$ is a minimal tree, then do Step 3, else do 2.
Step 2: For the pair (c,r) of point and nonnegative real stored on the root N of $\boldsymbol{T}^*$, execute one of (a)-(c).
(a) In case of $d(c,x) < r-\Delta$, compute $\hat{y}_\mathrm{L} = f_\Delta(\boldsymbol{T}^*_\mathrm{L}, x)$ and answer $\hat{y}_\mathrm{L}$ as the output of $f_\Delta(\boldsymbol{T}^*, x)$, where $\boldsymbol{T}^*_\mathrm{L}$ denotes the subtree of $\boldsymbol{T}^*$ whose vertex is the left child node of N.
(b) In case of $d(c,x) > r+\Delta$, compute $\hat{y}_\mathrm{R} = f_\Delta(\boldsymbol{T}^*_\mathrm{R}, x)$ and answer $\hat{y}_\mathrm{R}$.
(c) Otherwise, compute both of $\hat{y}_\mathrm{L}$ and $\hat{y}_\mathrm{R}$. If $d(x,\hat{y}_\mathrm{L}) \leq d(x,\hat{y}_\mathrm{R})$, then answer $\hat{y}_\mathrm{L}$, else answer $\hat{y}_\mathrm{R}$.
Step 3: Now $\boldsymbol{T}^*$ coincides with a leaf of $\boldsymbol{T}_\gamma(\mathbf{x})$ that stores a finite sequence $\mathbf{z} = z_1 \cdots z_{|\mathbf{z}|}$ of points (Note that $z_1 = \cdots = z_{|\mathbf{z}|}$ provided that $\gamma \leq 1/n$). Answer z_1. □

The branching into (a)-(c) based on the triangle inequality cuts off wasteful traversal over $\boldsymbol{T}_\gamma(\mathbf{x})$ efficiently. The nearest point is searchable by initializing $\boldsymbol{T}^*$ as $\boldsymbol{T}_\gamma(\mathbf{x})$ provided that $\gamma \leq 1/n$.

IV. Computation Steps for a Point Search

Lemma 1: Selected n i.i.d. samples $x_1, \cdots, x_n$ in $\{0,1\}^l$, the depth of $\boldsymbol{T}_\gamma(\mathbf{x})$ is almost surely $\leq \log_2(1/\gamma)$ if l and n are sufficiently large. □

For each nonleaf node N of $\boldsymbol{T}_\gamma(\mathbf{x})$, let $\mathcal{G}_\mathrm{N}$ (called a gray zone) denote the set of all query points that activate Step 2(c) in Proc. 2, i.e., $\mathcal{G}_\mathrm{N} = \{x |\ x \in \{0,1\}^l,\ r-\Delta \leq d(c,x) \leq r+\Delta\}$, where (c,r) is one stored on N.

Lemma 2: Fix an arbitrary real constant $\eta > 0$. Selected a query point x uniformly in $\{0,1\}^l$, *the probability that x may belong to $\mathcal{G}_\mathrm{N}$ on condition that a node pointer latches a nonleaf node* N *of* $\boldsymbol{T}_\gamma(\mathbf{x})$ at Step 2 in Proc. 2 is $\leq \eta$ if l is sufficiently large. □

In applying Proc. 2 on $\boldsymbol{T}_\gamma(\mathbf{x})$, let $\mu_\gamma(\mathbf{x})$ denote

$$\sum_{x \in \{0,1\}^l} \left(\begin{array}{l} \text{the number of the latched nodes} \\ \text{for a query point } x \end{array} \right) \cdot P(x), \quad (1)$$

where $P(x) = 1/|\{0,1\}^l| = 2^{-l}\ \forall x \in \{0,1\}^l$. We can regard this $\mu_\gamma(\mathbf{x})$ as the mean number of the latched nodes in once application of Proc. 2.

Lemma 3: Selected n i.i.d. samples $x_1, \cdots, x_n$ in $\{0,1\}^l$, $\mu_\gamma(\mathbf{x})$ is almost surely $\leq \log_2(1/\gamma) + 2 + 1/\eta$ if l is sufficiently large. □

Corollary 3: $\mu_\gamma(\mathbf{x})$ with $\gamma \leq 1/n$ is almost surely of $O(\log n)$ if l is sufficiently large. □

Thus, the mean number of computation steps of Proc. 2 is almost surely of $O(\log n)$ if l is sufficiently large.

References

[1] I. Kalantari and G. McDonald, "A data structure and an algorithm for the nearest point problem," *IEEE Trans. Software Eng.*, vol. SE-9, no. 5, pp. 631–634, 1983.

[2] H. Suzuki and S. Arimoto, "A method of managing perfectly-balanced trees for solving quickly the nearest point problems," *IEICE Trans. Fundamentals*, vol. E76-A, no. 9, pp. 1373–1382, 1993.

New estimation of the probability of undetected error

Volodia Blinovsky*
Inst. for Information Transmission Problems
Russian Acad. of Sci., Moscow 101447,GSP-4,Ermolovoy st.,19,Russia,
e-mail blinov@ippi.msk.su

Abstract— **We obtain the upper bound on the probability of undetected error which is valid uniformly on choosing the probability of the symbol invertion. This bound is better than previous known bounds**

Let F_2^n – Hamming space of binary sequences of length n with metric $d(x,y) = \sum_{i=1}^{n} \mid x_i - y_i \mid; x, y \in F_2^n$. Let $C_n(y,r) \triangleq \sum_{x \in F_2^n, d(x,y)=r} x-$ sphere of radius r with center in $y \in F_2^n$. For arbitrary linear code $A_{nk} \subset F_2^n$ of dimension $k(\mid A_{nk} \mid = 2^k)$ define the set $A_x \triangleq \{A_x^r; r = 0,1,\ldots n\}, x \in A_{nk}$ where $A_x^r \triangleq \mid Ank \bigcap C_n(x,r) \mid -$ numder of vectors from A_{nk} which distance from $x \in A_{nk}$ is equel to r. It is easy to see that A_x, A_x^r does not depends on $x \in A_{nk}$ so we ommit the index x in the notations A_x, A_x^r. The set A is called the spectrum of the code A_{nk}. For arbitrary $p \in [0,1]$ the probability of undetected error $P_{ue}(p, A_{nk})$ is defined by the equality

$$P_{ue}(p, A_{nk}) = \sum_{r=1}^{n} A^r p^r (1-p)^{n-r}.$$

We are interesting in the value

$$P(n,k) \triangleq \min_{A_{nk} \subset F_2^n} \max_{p \in [0,1]} P_{ue}(p, A_{nk}).$$

It is easy to show that

$$P_{ue}\left(\frac{1}{2}, A_{nk}\right) = \frac{2^k - 1}{2^n}$$

so

$$P(n,k) \geq 2^{k-n} - 2^{-n}.$$

The best known upper bound on $P(n,k)$ which is valid for all n,k was obtained in [1] and is the follows

$$P(n,k) \leq C_1 \sqrt{n} 2^{k-n}$$

where C_1 is constant $(C_1 \leq \sqrt{\pi/2}(1 + o(1)), n \to \infty)$. This bound was obtained by the estimation of the **RHS** of the following inequality offered earlier in [2]

$$P(n,k) \leq 2^{k-n} \sum_{r=1}^{n} C_n^r \left(\frac{r}{n}\right) \left(1 - \frac{r}{n}\right)^{r-n}.$$

Here we present the result which is the statement of the following theorem.

Theorem 1 *For some constant C_2 and for all n,k the following estimation is valid*

$$P(n,k) \leq (C_2 \sqrt{\ln n} + 1) 2^{k-n}.$$

During the proof of this theorem we show that at least for sufficiently large n the estimation $C_2 \leq 2/\sqrt{\pi}$ is valid, but it can be improved by the more precise calculations.

To prove this theorem we divide the spectrum A into six parts and prove the existence of the code A_{nk} which spectrum satisfying the following relations

$$\sum_{r=1}^{s_1-1} A^r p^r (1-p)^{n-r} = 0;$$

$$\sum_{r=s_1}^{s_2-1} A^r p^r (1-p)^{n-r} \leq \sqrt{\frac{\ln n}{\pi}} 2^{k-n};$$

$$\sum_{r=s_2}^{n-s_2} A^r p^r (1-p)^{n-r} \leq \left(1 + \frac{1}{\ln n}\right) 2^{k-n},$$

$$\sum_{r=n-s_2+1}^{n-s_1} A^r p^r (1-p)^{n-r} \leq \sqrt{\frac{\ln n}{\pi}} 2^{k-n};$$

$$\sum_{r=n-s_1+1}^{n} A^r p^r (1-p)^{n-r} = 0$$

for some $1 \leq s_1 \leq s_2 \leq n/2$. Note that if $k > C_3 \ln^2 n$ for some constant $C_4 > 0$ then using the same arguments as in the proof of the theorem it is easy to prove the more strong upper bound for $P(n,k)$:

$$P(n,k) \leq C_4 2^{k-n}$$

where $C_5 > 0$ come constant. We conjecture that in order to prove the last estimation for all values of k and n it is necessary to use some additional nonprobabilistic arguments.

*Supported by Russian Foundation of Fundamental Research Under Grant 93-012-458

REFERENCES

[1] T.Kløve"On Massey's Bound on the Worst-Case Probability of Undetected Error",*Proc.IEEE Int.Symp. on Inf.Theory,Trondheim,Norway*, pp. 242,1994.

[2] J.Massey "Coding techniques for digital data networks", *Proc. Int. Conf.Inform. Theory Syst.,NTG-Fachbeirichte 65,Berlin,Germany*, pp.307-315,1978.

Some Remarks on Efficient Inversion in Finite Fields

Christof Paar[1]

Worcester Polytechnic Institute, ECE Department, Worcester, MA 01609, email: christof@ece.wpi.edu

Abstract — **This contribution is concerned with bit parallel inverters over finite fields. Two alternative approaches for inversion with low complexity will be reviewed. Both methods are based on multiple field extension of $GF(2)$. It will be shown that one architecture is a generalization of the other's architecture core algorithm. As an impressive example, the complexity of an inverter in the field $GF(2^8)$ will be computed.**

I. Inversion in Extension Fields of Degree Two

The first architecture was proposed in [2] in 1989 and reintroduced in [3] in 1991. The core part of the architecture is the following. Let us consider an element $A = a_0 + a_1 x$ from $GF((2^{k/2})^2)$, where $a_0, a_1 \in GF(2^{k/2})$. There exists always a field polynomial of the form $P(x) = x^2 + x + p_0$, where $p_0 \in GF(2^{k/2})$. If the inverse is denoted as $B = A^{-1} = b_0 + b_1 x$, the equation $A \cdot B = [a_0 b_0 + p_0 a_1 b_1] + [a_0 b_1 + a_1 b_0 + a_1 b_1]x = 1$ must be satisfied, which is equivalent to a set of two linear equations in b_0, b_1 over $GF(2^{k/2})$ whose solution is:

$$\left.\begin{array}{rcl} b_0 & = & \frac{a_0+a_1}{\Delta} \\ b_1 & = & \frac{a_1}{\Delta} \end{array}\right\}, \text{ where } \Delta = a_0(a_0+a_1) + p_0 a_1^2. \quad (1)$$

The advantage of this algorithm is that all operations are performed in $GF(2^{k/2})$. The algorithm can be applied recursively.

II. Inversion in Composite Fields

The second architecture was proposed in the last section of Itoh-Tsujii's paper from 1988 [1, Section 6]. It is based on so-called composite fields which are finite fields with two extensions $GF((2^n)^m)$. We start with the trivial notation $A^{-1} = (A^r)^{-1} A^{r-1}$. If the auxiliary parameter r is defined as $r := \frac{2^{nm}-1}{2^n-1} = 1 + 2^n + \cdots + 2^{(m-1)n}$, we obtain the important property: $A^r \in GF(2^n)$, $\forall A \in GF((2^n)^m)$. We are now able to state a four step algorithm for computing the inverse of A:

Step 1 Compute A^{r-1}

Step 2 Compute $A^{r-1}A = A^r$

Step 3 Compute $(A^r)^{-1} = A^{-r}$ (Inversion in $GF(2^n)$)

Step 4 Compute $A^{-r}A^{r-1} = A^{-1}$

III. A Relation Between the Architectures

For the development of a relation between the two architectures, we consider [1] with composite fields $GF((2^n)^2)$ and $P(x) = x^2 + x + p_0$. An arbitrary field element is represented by $A(x) = a_1 x + a_0$, its inverse by $B := A^{-1} = b_1 x + b_0$. The parameter r is now $r = 2^n + 1$. By denoting $x^{r-1} = s_1 x + s_0$, Step 1 of the algorithm is: $A^{r-1} = [a_1 s_1]x + [a_1 s_0 + a_0]$. The computation in Step 2 is: $A^r = [a_0 s_1 + a_1 s_0 + a_0 + a_1 s_1]a_1 x + [a_0 a_1 s_0 + a_0^2 + a_1^2 s_1 p_0]$. Since A^r is an element of the subfield its coefficient at x is zero, and thus $a_1 s_0 + a_0 = (a_0 + a_1)s_1$. Inserting this relation in the expressions for A^{r-1} and A^r yields:

$$B(x) = A^{r-1}(A^r)^{-1} = \frac{a_1 x + (a_1 + a_0)}{a_0(a_1 + a_0) + a_1^2 p_0}. \quad (2)$$

Equation (2) is the same as the Equations (1). [1] can thus be viewed as a generalization of the core algorithm of [2]. [1] is, however, not a generalization of the *architecture* of [2], since the latter allows *multiple* field extensions of degree two.

IV. Efficient Bit Parallel Inversion in $GF(2^8)$

For the application of the architecture [2] the decomposition of $GF(2^8)$ into $GF((2^4)^2)$ is considered. Let $Q(y) = y^4 + y + 1$ be the primitive polynomial generating $GF(2^4)$ with $Q(\omega) = 0$ and $P(x) = x^2 + x + \omega^{14}$ the primitive polynomial generating the composite field. For computing Equations (1) in hardware, the following $GF(2^4)$ arithmetic modules must be provided:

- A direct approach allows inversion with not more than 15 XOR/10 AND gates [4, Appendix A].
- Three multiplications require 45 XOR/48 AND [5].
- The two additions require $2 \cdot 4 = 8$ XOR gates.
- Constant multiplication with ω^{14} requires 1 XOR gate.
- Squaring of an element requires 2 XOR gates.

The resulting over-all gate count of 71 XOR/58 AND is remarkably low. It is interesting to compare this complexity with bit parallel *multiplication*. For instance, the multiplier [5] has a gate count of 84 XOR/64 AND.

V. Conclusions and Further Research

Decomposition of Galois fields $GF(2^k)$ can lead to area-efficient inverters. In general, this approach seems promising since multipliers over composite fields can also be realized efficiently [3] [6]. For certain fields, in particular for $GF(2^8)$, and inverter can be realized with a gate count smaller than that of a multiplier. This result is contrary to common belief.

For technical applications it will be helpful to provide generators $x^2 + x + p_0$ for tower fields with *multiple* field extensions of degree two. Lists with irreducible polynomials over non-prime fields are very rare in literature. The zero coefficients p_0 of these polynomials should be optimized.

References

[1] T. Itoh and S. Tsujii, "A fast algorithm for computing multiplicative inverses in $GF(2^m)$ using normal bases," *Information and Computation*, vol. 78, pp. 171–177, 1988.

[2] M. Morii and M. Kasahara, "Efficient construction of gate circuit for computing multiplicative inverses over $GF(2^m)$," *Trans. of the IEICE*, vol. E 72, pp. 37–42, January 1989.

[3] V. Afanasyev, "On the complexity of finite field arithmetic," in *5th Joint Soviet-Swedish Intern. Workshop on Information Theory*, (Moscow, USSR), pp. 9–12, January 1991.

[4] C. Paar, *Efficient VLSI Architectures for Bit-Parallel Computation in Galois Fields*. PhD thesis, (Engl. transl.), Institute for Experimental Mathematics, University of Essen, Essen, Germany, June 1994. ISBN 3-18-332810-0.

[5] E. Mastrovito, "VLSI design for multiplication over finite fields $GF(2^m)$," in *Lecture Notes in Computer Science 357*, pp. 297–309, Springer-Verlag, Berlin, March 1989.

[6] C. Paar, "A parallel galois field multiplier with low complexity based on composite fields," in *6th Joint Swedish-Russian Workshop on Information Theory*, (Mölle, Sweden), pp. 320–324, August 22-27 1993.

[1] The research was done while the author was with the Institute for Experimental Mathematics, University of Essen, Germany.

Multilevel Coding with the 8-PSK Signal Set

Joakim Persson
Department of Information Theory, Lund University, Box 118, S-221 00 LUND, Sweden
email: Joakim.Persson@dit.lth.se

Abstract - **Simulation results for concatenated outer Reed-Solomon and inner convolutional codes used in multilevel schemes are presented. Different high-rate inner convolutional codes are considered, viz., punctured codes and partial unit memory (PUM) codes. Best results are obtained for PUM codes, since they have a better extended row distance profile. The effect of channel and block interleaving at the different levels is also studied, and iterative decoding is tried[1].**

I. Introduction

A multilevel code uses some signal set S_0 which is a finite subset of a lattice or a set of points with some group structure. This set is partitioned into a k-level partitioning chain, $S_0/S_1/\ldots/S_k$, which can be described as a rooted tree with $k+1$ levels (the root is level zero). Every node at level i is partitioned into disjoint subsets which are cosets. Each partition at level i, S_{i-1}/S_i, is determined by a component code C_i. In general these component codes may be of any type, but for this work we have only considered convolutional component codes and concatenated component codes with inner convolutional and outer Reed-Solomon codes. Using multilevel codes one can achieve arbitrarily large squared Euclidean free distance.

The structural properties of multilevel codes make them attractive for code constructions. Unfortunately, the decoding will be carried out in a way which is not maximum likelihood, otherwise the computational efforts become far too large even for small systems (i.e., systems with not very complex component codes). The computational complexity of the preferred multistaged decoding procedure from [1] is proportional to the sum of the complexities of each component code, but it suffers from error propagation. In order to minimize the errors at each level, a concatenated scheme with outer Reed-Solomon and inner convolutional codes was considered. The errors of the inner convolutional decoders occur in bursts, and the idea is that the inherent burst error correcting capability of the outer RS code will correct these errors.

Our system transmits signals over the AWGN channel. The used signal constellation is 8-PSK. This implies three levels in the system. Since the partition chain is 8-PSK/4-PSK/2-PSK/1-PSK, the minimum squared Euclidean distance among the signal points in the subsets at the different levels increases for each partition. Therefore the encoder of level 1 must be protected by a more powerful code than that of level 2, *et cetera*.

II. Simulation Results

The simulations show that there is no need for a concatenated code at level 3. In order to retain as high overall rate as possible, the rate of the inner code at level 2 must be quite large. Due to its simple decoder implementation, a punctured convolutional (PC) code was tested. Simulations then show that the bit error rate (BER) performance of level 2 bounds the overall code BER. This is caused by the bad extended row distance profile of punctured codes, i.e., error vectors $\boldsymbol{e}$ of small weight are enough to result in quite long bursts. As an alternative, a PUM code was tested. There exist decoding procedures for these codes [3] that are not more complex than decoding of PC codes. The simulations show that a PUM code with overall constraint length one less than the previously used PC code, performed only negligible worse (< 0.05 dB). From a practical point of view the reduced decoding complexity is far more important.

We need to minimize error propagation *at each* level (between the inner and outer code) as well as the error propagation *between* levels. The first is accomplished by reducing the length of error events from the inner decoder by applying block interleaving. The simulations confirmed a theoretical result from [2] on how many rows the interleaver matrix need in order to maximize the free distance of the concatenated system. One idea how to decrease the error propagation between levels is to interleave the coset labels of different levels in time (channel interleaving). However, this seems to be of little help. Comparing simulations of our system without channel interleaving with simulations of a theoretical system without any error propagation at all (a genie between every level), shows a difference of less than 0.05 dB already at a BER of 10^{-4}.

Finally we studied iterative decoding and its influence at the different levels. There is no immediate way of extracting the error probability of individual decoded bits because of the hard decoding of the RS codes. It turns out that only the first level benefits from 'hard' iterative decoding. The improvement on the whole system is only marginal (the asymptotic error performance follows level 2 instead of level 1). The total BER is not changed more than a few tenths of a dB.

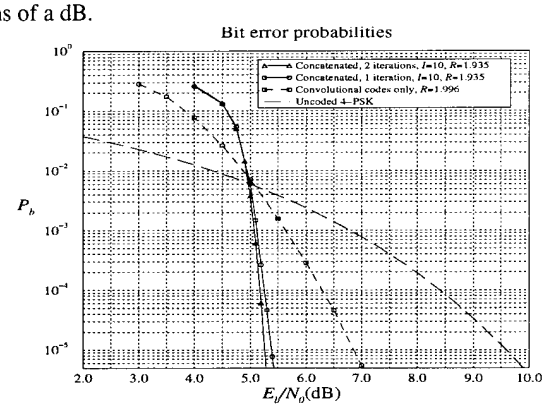


Fig. 1. Simulation results for the concatenated system vs. uncoded 4-PSK.

1. This work was supported in part by the Swedish Research Council for Engineering Sciences under Grants 92-661 and 94-83.

References

[1] H. Imai, S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. on Inform. Theory*, vol. IT-23, pp. 371-376, May 1977.

[2] J. Justesen, C. Thommesen, V. Zyablov, "Concatenated Codes with Convolutional Inner Codes," *IEEE Trans. on Inform. Theory*, vol. IT-34, pp. 1217-1225, Sept. 1988.

[3] V. V. Zyablov, V. R. Sidorenko, "Soft-decision Maximum-Likelihood Decoding of Partial-Unit-Memory Codes", *Prob. Peredachi Inform.* (English transl.), Vol. 28, No. 1, pp. 22-27, Jan.-March 1992.

Soft-Decision Decoding for Trellis Coding and Phase-Difference Modulation

Michael B. Pursley and John M. Shea
Department of Electrical & Computer Engineering
Clemson University, Clemson, SC

Abstract -**A simple method is presented for performing soft-decision demodulation and decoding of trellis-coded phase-difference modulations. Results are given for trellis-coded M-ary differential phase-shift keying and M-ary double-differential phase-shift keying, each with soft-decision demodulation and decoding. The performances of these combinations of coding, modulation, demodulation, and decoding are presented for channels which may introduce a phase ramp in the modulated signal.**

SUMMARY

Phase-difference modulation, such as M-ary differential phase-shift keying (M-DPSK) and M-ary double-differential phase-shift keying (M-D^{2}PSK) [1], is desirable for some mobile radio systems and channels in which it is difficult to obtain an accurate phase reference. Either M-DPSK or M-D^{2}PSK may be coupled with a trellis code to decrease the probability of bit error for a given signal-to noise ratio (SNR). As the rate of the code is decreased, the number M of points in the M-ary PSK (M-PSK) signal constellation must be increased in order to transmit the same rate of information in the same bandwidth. As M is increased, the probability of symbol error increases, even for a channel with perfect phase stability. However, even greater degradation results if there is Doppler shift in the channel or phase drift in the system's oscillators. It is therefore of interest to investigate modulation and coding systems that can tolerate such a phase variation in the carrier signal. To avoid trivialities, it is assumed in all that follows that M>4.

Trellis coding provides coding gain to offset the increase in symbol error probability that results from increasing M. Optimal trellis demodulation and decoding may be too complex to implement in a mobile radio system. An alternative method which performs nearly as well and is much less complex is to perform the demodulation and decoding separately. For example, it has been suggested that the pragmatic trellis code be demodulated in this way, with hard or soft bit decisions at the output of the demodulator being input to a convolutional decoder modified to correct parallel branch errors [2].

The decision regions for standard hard-decision demodulation of M-PSK signals correspond to equal-length intervals for the phase of the received signal. As a consequence, standard hard-decision demodulation is easy to implement, but it does not provide information on the relative reliabilities of the bit decisions that result from a symbol decision. Because some bit decisions are more reliable than others, soft-decision demodulation and decoding should be employed.

The natural generalization of the standard method for soft-decision demodulation and decoding of binary signals (e.g., binary PSK) is not effective in M-PSK demodulation, in part because the reliabilities of the bit decisions do not depend only on the received signal strength. The optimum method for soft-decision decoding for a channel with perfect phase stability is too complex for most applications; in particular, it requires an accurate measurement of the SNR in the front end of the receiver. In addition, this method may perform very poorly if there is any phase drift in the carrier.

We propose a suboptimal method to generate quantized soft information for each bit associated with an M-PSK symbol. This method exploits the way bits are assigned to symbols in the M-PSK constellation, and it is simple to implement in the last stage of the demodulator. Simulation results show that the proposed method provides a significant performance improvement over hard-decision demodulation and decoding. The method is based on dividing each hard-decision phase interval into subintervals, using phase as the only criterion. The weights for the individual bits are constant throughout each subinterval, but they vary among the subintervals, even within the same hard-decision interval. The length of the subintervals can be adjusted to optimize performance.

A simulation was employed to obtain numerical values for the additional coding gain for soft-decision decoding over hard-decision decoding. The bit error probability is shown in Figure 1 as a function of E_b/N_0, the energy per information bit divided by the one-sided spectral density of the white Gaussian noise. The dashed curves illustrate that the simple two-bit quantized soft-decision decoding scheme for 8-DPSK with the rate 2/3 pragmatic trellis code provides up to 1.5 dB additional coding gain over the hard-decision system on the additive white Gaussian noise channel with a stable phase. The solid curves show that the simple two-bit quantized soft-decision decoding scheme with 8-DPSK performs up to 3.5 dB better than the hard-decision system for a system with a 10 degree phase rotation. The phase rotation is defined as the phase change over the duration of one M-ary symbol due to a linear phase drift in the carrier. The two-bit soft-decision decoding scheme used with M-D^{2}PSK provides up to 2.2 dB coding gain over hard-decision decoding for channels with stable phase and channels with phase ramps.

REFERENCES

[1] M. K. Simon and D. Divsalar, "On the implementation and performance of single and double differential detection techniques," *IEEE Trans. Commun.*, vol. 40, no. 2, pp. 279-291, February 1992.

[2] Viterbi, Wolf, Zehavi, and Padovani, "A pragmatic approach to trellis-coded modulation", *IEEE Commun. Mag.*, pp. 11-19, July 1989.

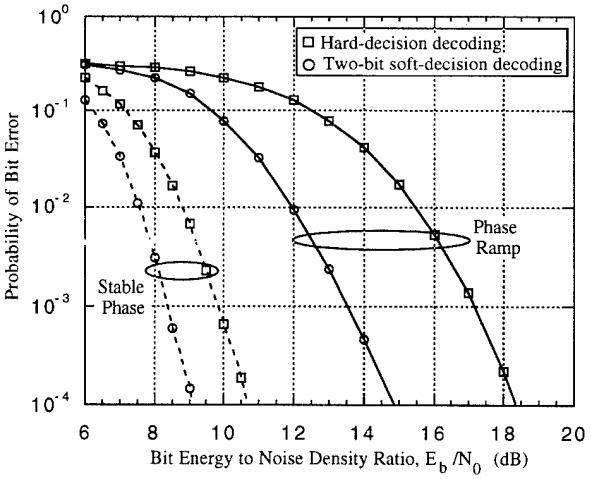


Figure 1. Comparison of hard- and soft-decision decoding for two channels

This research was funded in part by the Army Research Office under grants DAAH04-93-G-0253 and DAAH04-94-G-0154 and in part by a grant from ITT Aerospace and Communications Division. John M. Shea is the recipient of a National Science Foundation Graduate Research Fellowship.

Coding and Decoding of Punctured QAM Trellis Codes*

François Chan and David Haccoun
Department of Electrical and Computer Engineering
Ecole Polytechnique de Montreal
P.O. Box 6079, station "Centre-Ville", Montreal, Canada, H3C 3A7

Abstract — Punctured convolutional codes allow an easy implementation of variable-rate encoders/decoders. In this paper, the puncturing technique is used to generate new QAM trellis codes from a rate-1/2 code. These codes are true high-rate codes, without parallel branches in the trellis. A simplified decoding technique is also presented. It is shown that the advantages the puncturing technique provides with binary convolutional codes are essentially maintained with Trellis-Coded Modulation.

Summary

Trellis-Coded Modulation (TCM) can yield significant coding gains of 3 to 6 dB over uncoded modulation without bandwidth expansion [1]. Unfortunately with Ungerboeck's usual TCM, each signal constellation requires a different code. For example, a code for 8-PSK is different from a 16-PSK code. As a consequence, implementing a system with various spectral efficiencies (e.g., 2, 3 and 4 bits/s/Hz) would necessitate several distinct encoders/decoders. In addition, since there are 2^m branches converging onto each trellis state for a rate R=m/(m+1) TCM code, decoding such a code with the Viterbi algorithm requires (2^m-1) binary comparisons per state. Hence, Viterbi decoding in the usual manner becomes quickly impractical as the number of states and the coding rate increase. A pragmatic approach to this problem has been proposed by using a rate-1/2, 64-state convolutional code and adding (m-1) uncoded bits to the output to produce a rate R=m/(m+1) code [2, 3]. The disadvantage of this approach is that the trellis exhibits parallel branches. For some codes, limiting the free distance to the distance between parallel branches leads to suboptimality.

It has been shown that the puncturing technique can be applied to TCM [4]. Using extensive computer searches, 8-PSK and 16-PSK punctured codes have been found with free squared Euclidean distances that are either equal to or almost as large as the distances of the best known codes discovered by Ungerboeck. The puncturing technique can also provide codes with uncoded input bits and parallel branches in the trellis. Furthermore, variable-rate punctured TCM codes have also been found using computer search. Families of QPSK, 8-PSK and 16-PSK codes, which are quite good in the sense of Euclidean distance as compared to the best known codes, have been obtained from a single rate-1/2 convolutional code and a varying puncturing pattern [5]. The advantage of using a single rate-1/2 code is that variable bandwidth efficiencies and hence, variable throughputs can be achieved with a single encoder/decoder.

The puncturing technique presented here is quite flexible, allowing either a true high-rate code or a code with parallel branches. In this paper, new 8-QAM, 16-QAM and 32-QAM punctured trellis codes are presented. These codes are true high-rate codes without parallel branches. The free Euclidean distance is not limited by the distance between parallel branches and hence, when the number of states is large, these codes can provide a larger free distance than codes with parallel branches. Furthermore, over Rayleigh fading channels, the absence of parallel branches in the trellis is beneficial since codes without parallel branches yield a better error performance than codes having parallel branches.

By using the fact that these QAM codes are generated from a rate-1/2 code, decoding can be performed on the low-rate trellis. Hence, the reduction in the number of binary comparisons the puncturing technique provides with convolutional codes is essentially maintained with TCM at the cost of a slight degradation in the error performance. These decoding techniques and simulations results will be presented.

* This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada

References

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55-67, Jan. 1982.

[2] A. J. Viterbi, J. Wolf, E. Zehavi, and R. Padovani, "A pragmatic approach to trellis-coded modulation," *IEEE Commun. Mag.*, pp. 11-19, Jul. 1989.

[3] J. Wolf and E. Zehavi, "P^2 codes: Pragmatic trellis codes utilizing punctured convolutional codes," in *Proc. IEEE Int. Symp. on Inform. Theory*, p. 448, Trondheim, Norway, 1994.

[4] F. Chan and D. Haccoun, "High-rate punctured convolutional codes for trellis-coded modulation," in *Proc. IEEE Int. Symp. on Inform. Theory*, p. 414, San Antonio, Texas, 1993.

[5] F. Chan and D. Haccoun, "Variable-rate punctured trellis-coded modulation and applications," in *Abstracts of papers, 1993 Canadian Workshop on Information Theory*, p. 5, Rockland, Ontario, 1993.

Power Efficient Rate Design for Multilevel Codes with Finite Blocklength

Johannes Huber Udo Wachsmann[1]

Lehrstuhl fuer Nachrichtentechnik, Universitaet Erlangen-Nuernberg, Cauerstrasse 7, D-91058 Erlangen, Germany

Abstract — **New design rules for multilevel codes with finite codeword length are derived from information theory leading to digital transmission schemes with high power and bandwidth efficiency.**

I. Introduction

Multilevel coding (MLC) is a well known approach to create power and bandwidth efficient communication schemes. Usually, the component codes are designed for balanced Euclidean distance for all levels, see e.g. [2]. But this rule does not take into account the tremendously increasing number of nearest neighbour error events for low levels due to the multiple representation of code symbols by signal points, cf. [3]. Thus, in multistage decoding a predomination of errors in low levels can be observed which leads to a serious degradation in power efficiency. Therefore, we propose to design the component codes using parameters from information theory of the equivalent channels at the individual levels.

II. Multilevel Coding

MLC for a $M = 2^\ell$-ary digital modulation scheme is based on a binary set partitioning of the signal constellation $\mathbf{A} = \{a_m | m \in \{0, 1, \ldots, M-1\}\}$ defining a mapping $m \leftrightarrow \mathbf{c}$ of binary labels $\mathbf{c} = (c^0, c^1, \ldots, c^{\ell-1})$ to the signal points a_m. The subsets of signal points at level i are denoted by the path to the subsets in the set partitioning tree, i.e.

$$\mathbf{A}_{c^0 \ldots c^i} = \{a_m | m \leftrightarrow (c^0, \ldots, c^i, x^{i+1}, \ldots, x^{\ell-1}), x^j \in \{0,1\}\}.$$

At each level i equivalent channels can be considered for the transmission of binary symbols c^i. The sum of capacities C^i of these equivalent channels yields the capacity C of the communication scheme ([4], [3]). Consequently, we proposed to choose the rates R^i of long codes at levels i equal to the capacities C^i [3].

III. Rate Design for Finite Blocklength

The blocklength of MLC schemes is limited due to restrictions like delay or decoder complexity. Therefore, a design rule for MLC with finite and uniform length n of the component codes at each level is presented in this paper.

The tool to consider codes with finite length n is the random coding bound

$$p_e \leq 2^{-n \cdot E_r(R)}, \tag{1}$$

where p_e denotes the probability of block errors and $E_r(R)$ the random coding exponent.

For transmission of a symbol c^i at level i in a MLC scheme a point of the subset $\mathbf{A}_{c^0 \ldots c^i}$ is selected equiprobably. Thus, the probability density function (pdf) of the continuous channel output y for given c^i reads

$$f_{\mathbf{y}}(y|c^i) = \frac{1}{|\mathbf{A}_{c^0 \ldots c^i}|} \sum_{a_m \in \mathbf{A}_{c^0 \ldots c^i}} f_{\mathbf{y}}(y|a_m), \tag{2}$$

where the conditional pdf's $f_{\mathbf{y}}(y|a_m)$ characterize the discrete memoryless channel. From this equation, the random coding exponents $E_r^i(R^i)$ for the equivalent channels at levels i of a MLC scheme can be calculated in a straightforward way.

A suitable representation of the random coding bound for the rate design are isoquants

$$E_r^i(R^i) = -\frac{\log_2 p_e}{n} = \text{const.} \; \forall \; \sigma^2, \tag{3}$$

where σ^2 denotes the noise variance per dimension. We propose the design rule:

> *For a maximum tolerable block error rate p_e and given codeword length n at all levels, choose the rates R^i of a MLC scheme from the corresponding isoquants of the random coding exponents $E_r^i(R^i)$ for given noise variance σ^2 or given total rate $R = \sum_i R^i$.*

IV. Simulation Results

Simulation results for digital PAM transmission with MLC over the AWGN channel are presented. Turbo codes [1] with rates designed from random coding bound are employed as component codes. For 16QAM with total rate $R = 3$ and blocklength $n = 2000$ a bit error rate (BER) $\leq 10^{-5}$ is achieved only 1.4 dB above capacity limit. For $n = 20000$, BER $\leq 10^{-5}$ only 0.8 dB above capacity has been observed. For 8PSK with total rate $R = 2$, simulation results are similar. The results for 16QAM can be extended to $M > 16$-ary QAM schemes by imposing further uncoded levels. Furthermore, these uncoded levels can be employed to achieve an additional shaping gain.

V. Conclusion

The benefits of powerful binary codes can be transferred to any digital transmission scheme via the multilevel coding approach, if the individual rates are well chosen, e.g. according to the random coding bound criterion for the individual levels. Application of Turbo codes to MLC schemes offers digital communication close to capacity limit for a wide range of trading power for bandwidth efficiency.

[1] The work was supported by Deutsche Forschungsgemeinschaft under contract Hu 634/1-1

References

[1] C. Berrou, A. Glavieux, P. Thitimajshima. Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes (1). In *Proc. of the Int. Conf. on Comm. (ICC'93)*, pp. 1064-1070, Geneva, May 1993.

[2] E. Biglieri e.a. *Introduction to Trellis-Coded Modulation with Applications*. Macmillan, New York, 1991.

[3] J. Huber, U. Wachsmann. Capacities of the Equivalent Channels in Multilevel Coding Schemes. *Electronics Letters*, vol. 30: pp.557-558, March 1994.

[4] Y. Kofman, E. Zehavi, S. Shamai (Shitz). A Multilevel Coded Modulation Scheme for Fading Channels. *AEÜ (Int. Journal of Electronics and Comm.)*, No. 6: pp.420-428, 1992.

Trellis Coding of Gaussian Filtered MSK#

Piotr Tyczka* and Witold Hołubowicz**

*Poznań University of Technology, Institute of Electronics and Telecommunications, ul.Piotrowo 3a, 60-965 Poznań, Poland
E-mail: tyczka@et.put.poznan.pl

**Franco-Polish School of New Information and Communication Technologies - EFP, ul.P.Mansfelda 4, 60-854 Poznań, Poland
E-mail: holub@cfp.poznan.pl

Abstract - **Trellis coding of Gaussian minimum shift keying (GMSK) is considered. The structure of combinations of rate 1/2 and 2/3 binary convolutional encoders and GMSK modulation with several values of the parameter *BT* is studied by means of the so called "matched coding approach" [4,5]. It is shown that in such connections up to 3 distinct classes of codes can be identified each with different receiver complexity. The results of the optimization procedure for codes combined with GMSK are given. The results show that significant coding gains (over 6.5 dB) are obtained. Power-bandwidth performance of the best coded schemes is presented where it is demonstrated that variation of *BT* offers another degree of freedom in the design of communication systems.**

I. Introduction

Demand for spectrally-efficient modulation techniques for use in various communication systems and the inherent properties make Gaussian minimum shift keying (GMSK) [1] an attractive scheme for prospective applications. In recent years, trellis coding of modulations with memory has gained much attention since it usually offers significant coding gains and hence, improved power efficiency what is especially important in power-limited systems [2, 3]. In this paper, we study application of trellis coding technique to GMSK schemes with selected values of the normalized bandwidth of the premodulation filter *BT*. The first objective is to analyze how convolutional codes interact with the memory of the GMSK modulator and how it influences the trellis of the combined receiver for the coded scheme. We also give quantitative results of coding gains over the uncoded signals that can be achieved due to trellis coding. Finally, we present the performance of the best coded GMSK schemes in terms of power-bandwidth tradeoffs and compare them to other binary systems.

The considered system consists of a convolutional encoder followed by a GMSK modulator, AWGN channel and the optimum Viterbi receiver which uses a combined encoder-modulator trellis for joint demodulation and decoding. The GMSK signal is a constant envelope RF phase-modulated signal where the information carrying phase is given by:

$$\phi(t,\underline{\beta}) = \pi \int_{-\infty}^{t} \sum_{i=-\infty}^{\infty} \beta_i g(\tau - iT) d\tau + \phi_0 \qquad (1)$$

where β_i is the transmitted symbol, and $g(t)$ is the frequency impulse of the form:

$$g(t+\frac{LT}{2}) = \frac{1}{2T}\left[Q\left(\frac{2\pi B}{\sqrt{\ln 2}}(t-\frac{T}{2})\right) - Q\left(\frac{2\pi B}{\sqrt{\ln 2}}(t+\frac{T}{2})\right)\right] \qquad (2)$$

The values of L which determine the duration of the impulse $g(t)$ depend on the particular GMSK scheme. For a finite length LT of $g(t)$ a modulator can be represented as a finite-state sequential machine. Following the approach of [4], a precoder $T(D)=1+D$ was used in our system which precodes the input to the modulator making it a feedback-free one.

II. Convolutional codes combined with GMSK

We consider combinations of noncatastrophic convolutional codes of rates 1/2 and 2/3 and precoded GMSK modulators. We assume that when concatenating convolutional encoders with modulators the initial state of both circuits is a zero state. Let S_G denote the number of an encoder states and S_V the number of states in the combined Viterbi receiver. The following lemmas can be formulated for these schemes.

Lemma 1: For the GMSK, BT=0.5 and BT=0.4 modulators combined with the rate 1/2 and rate 2/3 convolutional codes and for every $S_V \geq 4$, there are exactly two distinct classes of codes (*A* and *B*) producing the required value of S_V, namely:

$$A: \quad S_G = 1/4\, S_V \qquad (3)$$
$$B: \quad S_G = 1/2\, S_V \qquad (4)$$

Lemma 2: For the GMSK, BT=0.3 and BT=0.25 modulators combined with the rate 1/2 convolutional codes and for every $S_V \geq 4$, there is exactly one class of codes (*A*) producing the required value of S_V, namely:

$$A: \quad S_G = 1/4\, S_V \qquad (5)$$

Lemma 3: For the GMSK, BT=0.3 and BT=0.25 modulators combined with the rate 2/3 convolutional codes and for every $S_V \geq 8$, there are exactly three distinct classes of codes (*A, B* and *C*) producing the required value of S_V, namely:

$$A: \quad S_G = 1/8\, S_V \qquad (6)$$
$$B: \quad S_G = 1/4\, S_V \qquad (7)$$
$$C: \quad S_G = 1/2\, S_V \qquad (8)$$

Codes of (4), (5) and (8) are called matched codes (encoders) [5] for the respective GMSK modulators. The remaining codes are mismatched ones.

III. Numerical results

A systematic search for best matched and mismatched short convolutional codes maximizing minimum squared Euclidean distance of the coded GMSK schemes has been performed. Table 1 contains the distances of the best connections of GMSK signals and rate 1/2 codes. All schemes presented in the table were obtained using matched codes. The results show that matched codes usually outperform mismatched codes by 0.5 to 1 dB. Coding gains over uncoded signals range from 1.3 to 6.6 dB for all considered GMSK signals and code rates, increasing with the receiver complexity.

The comparison of the coded GMSK with other binary systems has been done in terms of the power-bandwidth performance. In particular, it turned out that best rate-2/3 coded GMSK with BT=0.5 found by us perform nearly the same as rate-1/2 coded TFM schemes of [4] for Viterbi receivers with more than 16 states.

Table 1
Normalized minimum squared Euclidean distances of the best rate-1/2 coded GMSK schemes with optimum receivers of up to 128 states.

BT\S_V	4	8	16	32	64	128
0.5	3.00	4.00	5.91	5.97	7.91	8.87
0.4	3.00	4.00	5.83	5.95	7.83	8.77
0.3	1.12	3.00	4.88	5.77	6.77	7.67
0.25	1.19	3.00	4.82	5.64	6.64	7.52
0.2	---	3.00	3.92	5.19	5.68	7.04
0.15	---	1.56	3.02	4.56	5.07	6.10

References

[1] K. Murota and K. Hirade, "GMSK modulation for digital mobile radio telephony," *IEEE Trans. Commun.*, vol. COM-29, pp. 1044-1050, July 1981.

[2] J. B. Anderson, T. Aulin, and C.-E. Sundberg, *Digital Phase Modulation*. Plenum Press, New York, 1986.

[3] J. B. Anderson and C.-E. Sundberg, "Advances in constant envelope coded modulation," *IEEE Commun. Mag.*, vol. 29, pp. 36-45, Dec. 1991.

[4] F. Morales-Moreno, W. Hołubowicz, and S. Pasupathy, "Optimization of trellis coded TFM via matched codes," *IEEE Trans. Commun.*, vol. 42, pp. 1586-1594, Feb./Mar./Apr. 1994.

[5] F. Morales-Moreno and S. Pasupathy, "Structure, optimization and realization of FFSK trellis codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 730-751, July 1988.

This work was supported by Grant KBN-8S50401905

Bit Error Rate Reduction of TCM Systems Using Linear Scramblers

Paul K. Gray and Lars K. Rasmussen

Institute for Telecommunications Research, University of South Australia, The Levels, SA, Australia 5095

Abstract — **It is shown that the use of Gray scramblers and Gray mapped signal sets are equivalent. A search is performed for better scramblers, including a search for scramblers with memory. Memoryless scramblers are found to give best performance and an explanation for this is given.**

I. Introduction

Recent authors have suggested ways in which the BER of trellis codes can be reduced. In [2] the scrambling of the information bits with a Gray coder prior to encoding is discussed, while in [3] and [4] the use of a Gray coded signal set mapper is examined. We will show that these two methods are equivalent. We also present a systematic technique based on bounds for P_e and P_b for finding the best scrambler to be used with a given trellis code. This search is not limited to combinatoric circuits, we also search for scramblers with memory.

II. Algebraic Relation Between Gray Coded Scrambler and Signal Mapper

The Gray coded 8-PSK signal set mapper used in [3] can be represented as a naturally mapped 8-PSK signal set mapper preceded by an $n \times n$ matrix transformation C. The Gray coded scrambler considered in [2] precedes the generator matrix and is represented by the $k \times k$ matrix transformation S. In general, the algebraic relation between an 8-PSK trellis code with a Gray scrambler and a natural signal mapper, and an equivalent 8-PSK trellis code based on a Gray coded signal mapper is

$$SG_n = G_g C \tag{1}$$

where G_n and G_g are the generator matrices for the code with the naturally mapped signal set and the code with the gray coded signal set, respectively. This relationship does not hold between all the 8-PSK codes in [1] and [3], because in [3] the authors have found codes with a better P_e than those in [1]. However, it is possible to use (1) to transform the codes of [3] to equivalent naturally mapped codes which will have a better P_e than the Ungerboeck codes. Preceded by a Gray scrambler the BER performance of the new code will be identical to the Gray mapped code.

III. Search Method

The union bound on P_b is used as a cost function to choose the best scrambler, so that the effect of the scrambler on an error path is weighted by its probability. Consider the effect of some scrambler $s(\cdot)$ on a sequence of correct data c; the input to the encoder will be $s(c)$, and if an error e occurs the output of the decoder will be $s(c)+e$, and the output of the descrambler will be $s^{-1}(s(c)+e)$. If the scrambler is linear we have

$$s^{-1}(s(c)+e) = s^{-1}(s(c)) + s^{-1}(e) = c + s^{-1}(e) \tag{2}$$

so the scrambling does not affect the correct path. Thus we wish to find a scrambler s which minimises

$$\hat{P}_b = \sum_{\epsilon_i \in \epsilon} W(s^{-1}(\epsilon_i))\,\Pr(\epsilon_i) \tag{3}$$

where ϵ is a subset of the set e of all error paths, consisting of only the error paths which have a significant effect on the cost function $\hat{P}_b$. $W(\cdot)$ is the Hamming weight of the error path.

For any code G there exists an equivalent *systematic* encoder matrix G_{sys} such that $G_{sys} = TG$. G_{sys} has a trivial right-inverse of degree 0 whereas G generally does not. This means that the error paths produced by G_{sys}^{-1} will have lower degree than those produced by G^{-1}, hence, while scramblers S_1 and S_2 give identical performance with generator matrices G_{sys} and G, respectively, scrambler S_1 will have lower degree than S_2. Thus the search for the best scrambler for the code generated by G should involve first finding the error paths for G_{sys}. The best scrambler S for G_{sys} can then be found, and the best scrambler for G will then be ST.

IV. Search Results

A search was performed for the best scrambler for $\nu = 3$ systematic Ungerboeck codes with k varying from 2 to 5. In all cases a memoryless scrambler was found to give best performance. The reason for this can be seen if we look at a list of error vectors ordered according to probability. It is clear that the best memoryless scrambler found will reduce the Hamming weight of *all* vectors, producing an almost ideal list, *i.e.*, vectors with high probability have low Hamming weight and *vice versa*. To get further improvement we must permute a small number of vectors, leaving most fixed. However for a k-dimensional vector space there are at most k invariant subspaces, so it is clear that we cannot change a small number of vectors. If we were to use a nonlinear scrambler we could do this, but then (2) would not hold.

The best scrambler in all cases was found to reduce the BER by approximately 1/3. This gain is only significant in applications where the gradient of the BER curve is small, such as low E_b/N_0 operating points or on fading channels. For example, the E_b/N_0 required to achieve a BER of 10^{-2} with the $\nu = 3$ 8-PSK Ungerboeck trellis code is reduced by 0.25 dB when a scrambler is used.

Acknowledgements

The authors would like to acknowledge the help and assistance provided by Mr. Weimin Zhang.

References

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[2] W. Zhang, C. Schlegel, and P. Alexander, "The BER reduction for systematic 8-PSK trellis codes by a Gray scrambler," in *IEEE Intern. Conf. Universal Wireless Access*, (Melbourne), pp. 35–38, Apr. 1994.

[3] J. Du and M. Kasahara, "Improvements of the information-bit error rate of trellis coded modulation systems," *Trans. of the IEICE*, vol. E 72, pp. 609–614, May 1989.

[4] J. Du and B. Vucetic, "New M-PSK trellis codes for fading channels," *Electron. Letters*, vol. 26, Aug. 1989.

Multidimensional Signaling for Bandlimited Channels

Fred Daneshgaran[a] and Marina Mondin[b]

[a] Elec. & Comp. Eng. Dept., California State University, Los Angeles, CA (USA)
[b] Dip. di Elettronica, Politecnico, Torino (ITALY)[1]

Abstract — **In this paper we focus on the issue of *distribution* of dimensions in time, describing a method, suited to different types of envelope functions used for modulation, that can generate an almost arbitrary distribution of dimensions in time with spectral efficiencies near the Nyquist limit. Subsequently, we propose a modulation scheme whereby the distribution of dimensions in time is used to carry additional information in the same BandWidth (BW).**

I. Introduction

The dimensionality theorem states that using shift orthogonal functions for modulation with shift period Δ and bandwidth B, in a T seconds interval we can generate at most $2BT$ dimensions [1]. We are interested in how these dimensions are distributed in time. Classically we have had two options: (1) if $\Delta = T$, the best basis functions to use are prolate spheroidal wave functions; (2) if $\Delta << T$, it is natural to use shift orthogonal functions such as the raised cosine shaping pulses, or the recently proposed scaling functions and wavelets [2]. The purpose of this paper is to present systematic methods based on the theory of wavelets and filter banks to generate almost arbitrary distributions of dimensions in time, achieving the highest spectral efficiency in a given BW.

II. Generation of Distribution of Dimensions

We describe here the basic steps of a procedure for the generation of distribution of dimensions. In the proposed method we use two shift orthogonal frequency overlapping functions, $q(t)$ and $w(t)$, where $q(t)$ is a lowpass function while $w(t)$ is a bandpass function. Both $q(t)$ and $w(t)$ are shift orthogonal with period Δ. $q(t)$ can be either a scaling function [2] or an even or odd shift orthogonal function [3]. $w(t)$ will be, respectively, the function $w(t) = \sqrt{2}q(t)sin(2\pi t/\Delta)$ or the wavelet associated with the scaling function $q(t)$.

Step 1: the overlap space between $q(t)$ and $w(t)$ is isolated by filtering the portion of $w(t)$ that falls on the BW of $q(t)$. For this purpose, either wavelet packets or nearly ideal low pass filters can be used, depending on the characteristics of the modulation waveforms. This operation generates a function $o(t)$ which is shift orthogonal with shift period $L\Delta$ (L is an integer). spanning a space occupying the same BW as $q(t)$ yet completely orthogonal to it. This function can be used to generate additional dimensions in the same BW as $q(t)$.

Step 2: the space spanned by $q(t)$ can be split into orthogonal frequency channels using the combination of wavelet packets and multiplicity-M wavelets. The overlap space spanned by $o(t)$ can be similarly partitioned. This orthogonal frequency channelization can be extremely flexible [2]. These results are subsequently used to introduce a novel coded modulation scheme based on the concept that *the way the time-frequency plane is partitioned into orthogonal frequency channels can carry information.*

[1]This work was partially supported by M.U.R.S.T.

III. Application to Coded Modulation

Suppose we have a two-state modulator which can choose between the shift orthogonal function $\phi(t)$ with shift period Δ (state σ_0) and two shift orthogonal functions $\phi_1(t)$ and $\phi_2(t)$ with shift period 2Δ (state σ_1). Then the dimensional rate in a given BW is fixed, but how the dimensions are distributed in the time-frequency plane differs for states σ_0 and σ_1. Consider parsing the source symbols a_n into non overlapping blocks. The state of the modulator can be controlled by an extra binary data stream, whose rate matches the symbol *block* rate.

The switching of the basis for two adjacent blocks could lead to ISI at the boundary of the adjacent blocks. However, given the state of the modulator, this ISI is deterministic and can be remedied.

The coherent demodulator at the receiver can either operate following a Maximum Likelihood (ML) detection rule, or performing hierarchical (suboptimal) demodulation.

The ML detection rule can be formulated to determine the state of the modulator from the observation of the received signal associated with the InterSymbol Interference (ISI) free portion of the blocks. Efficient search for the ML estimate of the a_n can be performed using the Viterbi algorithm with state complexity of $A^{0.5(L+1)}$ (assuming that L is odd), where A is the alphabet size of the sequence a_n and $L+1$ is the number of samples of the scaling and wavelet vectors [2]. Once the sequence a_n is detected, assuming that the receiver operates with very low error probability, we can use the ML estimated data vector $\tilde{a}$ to estimate the modulator state.

A practical alternative may be to use the correlation properties of the sampled outputs of the Matched Filters (MFs) at the receiver. Suppose the receiver employs one set of MFs for each state of the modulator. Then only the outputs of the correct MFs will be uncorrelated. Hence, time-averaged auto-correlation of the output samples of the MFs can be used to determine the modulator state. Once the modulator state has been estimated for a given block, the output of the correct MF is sampled to demodulate the received sequence for the portion of the block that is not corrupted by ISI. The portions of the block that may experience ISI are demodulated from the knowledge of the modulator state in the previous and the present blocks.

All the concepts presented above can be generalized to the case where there are other orthogonal channelizations of the available spectrum and can further be combined with channel coding.

References

[1] D. Slepian, "On Bandwidth," *Proc. of IEEE*, Vol.64, No.3, pp.292-300, Mar. 1976.

[2] F. Daneshgaran and M. Mondin, "Bandwidth Efficient Modulation With Wavelets," *Electronics Letters*, Vol.30, No. 15, July 1994.

[3] F. Daneshgaran and M. Mondin, "Multidimensional Signaling with Wavelets," *Proc. of the 29th CISS*, Philadelphia, MA (USA), March 1995.

EFFECT OF ENCODER PHASE ON SEQUENTIAL DECODING OF LINEAR CODED MODULATION

Krishna Balachandran[1] and John B. Anderson
Dept. of Electrical, Computer and Systems Eng., Rensselaer Polytechnic Institute,
Troy, NY 12180-3590, USA

Abstract — **We show that the performance of an M-Algorithm detector for linear partial response coded modulation depends critically on phase and is characterized by the partial energy function of the encoder.**

I. Introduction

Many practical communication channels may be adequately described by an equivalent discrete time model

$$r_k = h_0 a_k + \sum_{i=1}^{m} h_i a_{k-i} + n_k \qquad (1)$$

where a_k represents the data symbol, h_k represents an impulse response, n_k an additive white Gaussian noise (AWGN) component and m represents the channel memory. The above discrete time model can be used to construct a trellis. Maximum likelihood sequence estimation may be performed by searching this trellis with the Viterbi Algorithm (VA), but its complexity grows exponentially with the length of the channel impulse response.

A number of reduced search techniques like the M-Algorithm (MA) have been developed to achieve near optimum performance at a fraction of the optimum receiver complexity. In applications like mobile communication, the physical channel must often be characterized as a non-minimum phase channel. The purpose of this work is to characterize the effect of non-minimum phase channels on reduced search decoding complexity.

One feature that distinguishes channels having identical spectra and free distance but different phase is the partial energy given by $E(n) = \sum_{k=0}^{n} |h(k)|^2$. If E(n) represents the partial energy of any finite duration channel $h(n)$, then

$$E_{max}(n) \leq E(n) \leq E_{min}(n) \qquad (2)$$

where $E_{min}(n)$ and $E_{max}(n)$ represent the partial energies of the minimum and maximum phase channels having the same magnitude frequency response as $h(n)$.

II. Decoder Simulation Results

Channel phase effects were determined by performing MA decoder tests on different channels with the same autocorrelation. The results for one representative 10 tap channel class[2] having one real zero and 4 pairs of complex conjugate zeros, are described here. The class is specified by the normalized 99% bandwidth (NBW) and minimum distance loss (MDL) measured by $MDL = 10 \log_{10} \frac{d_{free}^2}{2}$.

The minimum phase channel, maximum phase channel and 4 mixed phase channels belonging to this class were chosen for performing MA tests. The partial energy curves and column distance profiles of these channels are plotted in Figures 1(a) and 1(b) respectively. MA simulations were carried out on these channels and the complexity was measured in terms of the minimum number of paths (M) needed by the decoder at each tree level in order to achieve near-MLSE performance. The complexity required by each of the channels is summarized in Table 1 . The minimum phase channel (a) needs the lowest value of M (4 paths) while the maximum phase channel (f) needs the highest complexity (M=128 paths). Channels that have similar partial energy curves turn out to require the same complexity. The partial energy curves of any one channel class show groups of channels having similar curves and the complexity required by the MA decoder increases as we move from one group to another one lower in the partial energy picture.

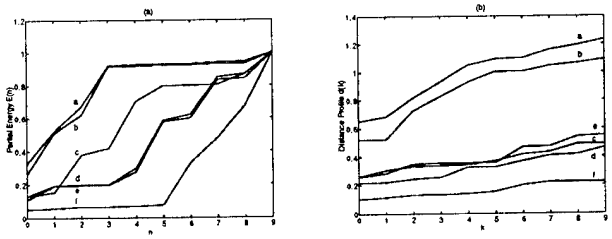


Figure 1: (a) Partial energy curves (b) Distance profile for selected channels of 10 tap class (NBW=0.36, MDL=0.19 dB).

Channel	a	b	c	d	e	f
Number of paths (M)	4	5	18	32	32	128

Table 1: MA decoder results for 10 tap equivalence class.

We have analyzed many channels (i.e., sets of $\{h_k\}$) and all show a behaviour[1] similar to Figure 1, except that the number of partial energy groups varies from 1 to 8. A superior partial energy curve and distance profile guarantees lower MA decoder complexity, but we see that partial energy serves as a better indicator of performance than the distance profile.

References

[1] K. Balachandran, "Effect of encoder phase on sequential decoding of partial response coded modulation," M.S. Thesis, Rensselaer Polytechnic Institute, Troy, NY, Feb. 1994.

[2] A. Said, "Design of optimal signals for bandwidth-efficient linear coded modulation," Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, NY, Feb. 1994.

[1]This work was partly supported by General Electric Corporate Research and Development Center, Schenectady, New York.

Reduced complexity algorithms in multistage decoding of multilevel codes[#]

Ryszard Bobrowski, Witold Hołubowicz

The Franco-Polish School of New Information and Communication Technologies - EFP, ul. P. Mansfelda 4, 60-854 Poznań, Poland,
E-mail: bobrow@efp.poznan.pl, holub@efp.poznan.pl

Abstract -**In this paper we investigate applicability of simplified decoders of convolutional codes to the case of multilevel coding [1]. System behaviour is examined by means of minimum distance analysis and simulation.**

I. PROBLEM STATEMENT

The objective of our research is to investigate applicability of reduced complexity algorithms for the decoding of multilevel codes combined with multi-resolution QAM, as proposed for the terrestrial transmission of HDTV signals in Europe.
Multistage decoder shown in the Fig. 1 will be examined in the paper. In this figure only the inner level of coding and modulation is shown. Other elements of the system are omitted [3].
Simplifications of the receiver are based on two different approaches: on the M-algorithm [4], which is the optimum solution for searching a limited part of trellis, and RSSE algorithm which is not optimum but is easier to build in hardware than M-algorithm. Both of these solutions consist of using a smaller number of states than that of the Viterbi algorithm.
The following benefits can be potentially achieved via the simplified algorithms in the receiver:

a) reduction of complexity of the decoder (reduction of the total cost of the system)
b) additional performance gain, for the fixed receiver complexity by the proper choice of the code structure in the transmitter.

The main purpose of the paper is to see if there is an additional coding gain achievable via the use of the multistage decoders based on the M-algorithm and RSSE [4] approach and if so, how large it is. Analysis is done on the basis of asymptotic coding gain (minimum distances). Numerical results of computer simulation are also provided.

II. NUMERICAL RESULTS

Firstly, we examine the degradation of performance due to simplifications of decoding. It has been done by simulations. An example of numerical results is shown in the Fig. 2. These are simulated bit error rates for convolutional codes of rate $r_0=1/3$ and $r_1=2/3$ decoded by RSSE algorithm, for the system of Fig. 1. Losses in this case are about 1 dB for reduction from 64 states to 32 states for Gaussian channel. Results for Rayleigh channel are also provided. Typically, it turns out that for Rayleigh channel and complexity reduction greater than 2, losses are significant (greater than 3 dB).
For concatenated coding systems very important to investigate are the properties of error bursts at the output of the decoder. We have analyzed the distribution of the average value of burst length. Numerical results for different rates and complexity reduction are provided for Gaussian and for Rayleigh channels. Typically, the length of bursts at the output of decoders with reduced complexity increases with decreasing number of the decoder states. Additionally, for multistage decoding average value of burst length are up to 3 times greater than for the case of single stage coding.

REFERENCES

[1] H. Imai and S. Hirakawa "A new multilevel coding method using error-correction codes", IEEE Trans. on IT, IT-23, May-1977, pp. 371-37
[2] J. Hagenauer "Rate compatible punctured convolutional codes and their applications" IEEE Trans. on Comm., COM-36, Apr. 1889, pp. 389-400.
[3] K. Fazel, M. J. Ruf "Combined multilevel coding and Multiresolution modulation", Proc. International Conference on Communication ICC-93, Geneva, Switzerland, May 1993, pp. 1081-1085
[4] J. B. Anderson, S. Mohan, "Source and channel coding - algorithmic approach" Kluwer Academic Publishers 1991

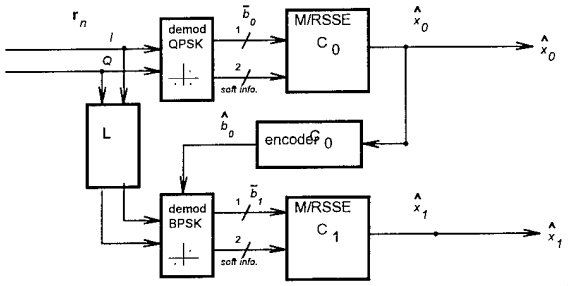


Fig. 1. Block diagram of multistage decoder of multilevel coding of 4 QAM.

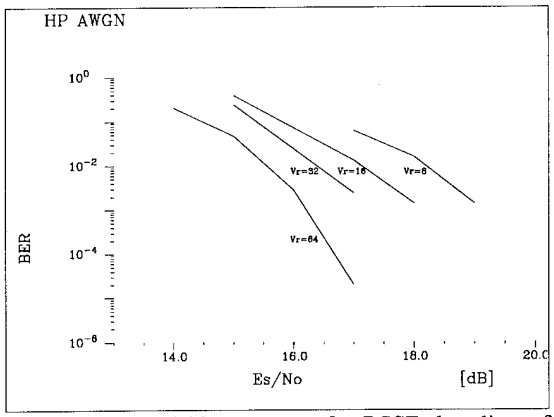


Fig. 2. Simulated bit error rate for RSSE decoding of multilevel convolutional codes with rates 1/3 and 2/3 combined with 4 QAM.

Table 1. Comparison of the values of average burst length for simplified decoding of single level convolutional codes (r=1/3,r=2/3) with multilevel coding. Results for constant value of bit error rate (BER=10^{-3}) and different number of states of the decoder Vrec.

	Vrec [states]	*64*	*32*	*16*	*8*
B_{av} [bits]	r=1/3 RSSE AWGN	*10*	*20*	*50*	*150*
	r=2/3 RSSE AWGN	*10*	*60*	*150*	*300*
	(1/3.2/3) RSSE AWGN	*30*	*100*	*300*	*750*

[#] This work was partially sponsored by the following grants: RACE R2082, KBN-8S50401905.

A Novel General Approach to the Optimal Synthesis of Trellis-Codes for Arbitrary Noisy Discrete Memoryless Channels

E. Baccarelli, R. Cusani, L. Piazzo

INFOCOM Dpt., University of Rome 'La Sapienza', Via Eudossiana 18, 00184 Roma, Italy

Abstract - **A new general criterion for the optimal design of (possibly) time-varying and nonlinear trellis-codes for reliable transmission of digital information sequences over arbitrary (possibly) time-varying Discrete Memoryless Channels (DMCs) is presented. The criterion is derived on the basis of new tight generally time-varying analytical upper bounds developed for the performance evaluation of MAP decoders with finite decoding constraint-lengths which minimise the symbol-error probability. New procedures related to the proposed criterion are also presented, allowing a direct construction of good trellis-codes for any arbitrary DMC and for any assigned value of the decoding constraint-length.**

SUMMARY

The common design criterion for trellis-codes requires the maximisation of the minimum Hamming distance (the so-called "free-distance") between codewords. Although this criterion is largely used in practical applications, its validity is not quite general. In fact it is well known that, almost in principle, it is optimal only in the case when the employed trellis-code is linear (i.e., it is a convolutional code), the assigned DMC is time-invarying, binary, symmetric and with a very small cross-over probability and a sequence Maximum Likelihood (ML) decoder with infinite decoding constraint-length Δ is present at the receiver site [1]. Barring for this case, the general issue of "good" trellis-code design for arbitrary noisy DMC channels seems not yet well explored in the literature.

In this contribution a novel general criterion is presented for the optimal design of trellis-codes (in general, nonlinear and time-varying) for arbitrary noisy DMCs (in general, non-binary, non-symmetric, time-varying and characterised by an arbitrary error-rate) when a decoder which minimises the symbol-decoding-error probability (i.e., a symbol-by-symbol MAP decoder) with an assigned and limited value Δ of the decoding constraint-length is employed.

The application environments of the proposed criterion are larger than that pertaining to the other criteria known in literature. In particular, the validity of the mentioned criterion is <u>not</u> restricted to the class of linear trellis-codes (i.e., of the convolutional codes) and of symmetric DMCs; moreover, it allows to take into account explicitly the value Δ assigned to the decoding constraint-length. The presented criterion is based on the following (generally) time-varying upper-bound derived in [3] as an application of the Chebyshev inequality to the performance evaluation of symbol-by-symbol MAP decoders:

$$P(\xi(k) \neq \hat{\xi}_{MAP}(k \mid k+\Delta)) \leq 2\, Tr\{\bar{S}(k \mid k+\Delta)\}, \quad k \geq 1. \qquad (1)$$

In (1) the Markov chain $\{\xi(k),\ k \geq 1\}$ is the so-called "state-transition sequence" of the trellis-encoder (defined as in [4,Sect.II]) and $\{\hat{\xi}_{MAP}(k|k+\Delta),\ k \geq 1\}$ is the corresponding (optimal) MAP estimate sequence (computed recursively as in [2]) when the decoding constraint-length takes on the value Δ. Moreover, $Tr\{\bar{S}(k|k+\Delta)\}$ is the trace of the average covariance error matrix $\bar{S}(k|k+\Delta)$ of the so-called "fixe-lag basic smoother" [2] and the sequence $\{\bar{S}(k|k+\Delta),\ k \geq 1\}$ can be recursively computed with respect to (wrt) the index k on the basis of a Riccati-type equation (formally similar to the well-known equation employed for the computation of the mean square error performance of a conventional Kalman filter), as shown in [2]. It must be remarked [3] that the sequence $\{Tr\{\bar{S}(k|k+\Delta)\}\}$ jointly depends on the sequence of the probability transition matrices of the assigned noisy DMC and on the set of the codewords of the employed trellis-code; therefore, the minimisation of the upper-bound sequence of (1) wrt the admissible sets of codewords gives a fully general criterion for the synthesis of good trellis-codes for any assigned value of Δ and for any arbitrary noisy DMC.

Procedures based on the described criterion for the construction of good trellis-codes with assigned rate R=b/n and encoding constraint-length L (defined as in [1]) have been implemented via computer [3]; for illustrative purposes, the trellis diagrams of the best trellis-codes with rate R=1/2 and L=2 obtained by means of an application of the mentioned construction procedure are reported in the Figures for some simple cases of stationary binary DMCs with transition probabilities p=P(1|1) and q=P(0|0). The two cases Δ=0 and Δ=2 have been considered in (a),(b),(c) and (d),(e),(f) respectively. In the Table the steady-state value of the sequence $\{Tr\{\bar{S}(k|k+\Delta)\}\}$ (denoted by $Tr\{\bar{S}(\infty|\infty+\Delta)\}$) is reported, together with the corresponding average bit-error-rate (BER) (evaluated by Montecarlo simulations) of the encoders generating the presented codes (the bold numbers denotes that the code has been optimized for Δ=0 or Δ=2).

On the basis of our analysis [3], some conclusions can be drawn:

- for an assigned DMC, the best trellis code for the case Δ=0 not always agrees with the best for Δ=2; in fact, in general, the topology and/or the labelling of the optimal trellis-code <u>change</u> with the value assumed by the decoding constraint-length; moreover, a value of Δ nearly equal to the encoding constraint-length L results in a negligible degradation wrt the optimum performance (ideally obtained for $\Delta \to \infty$);
- the topology and/or the labelling of the optimal trellis-code strongly depend on the statistical properties of the assigned DMC.

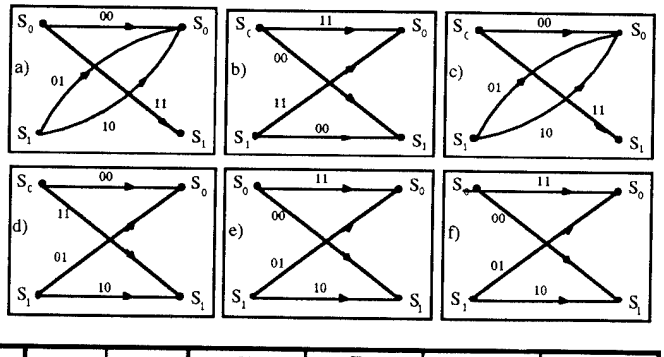
REFERENCES

[1] A.M. Michelson, A.H. Levesque, Error control techniques for digital communication, Chaps.8-9, John Wiley & Sons, Inc., 1985.

[2] E. Baccarelli, R. Cusani, "Universal Optimal Estimation and Detection of Markov Chains over Noisy Discrete Channels", INFOCOM Dpt Tech. Rep. n.005-02-94, Univ. Rome La Sapienza, Italy, Feb. 1994.

[3] E. Baccarelli, R. Cusani, G. Di Blasio, "A General Trellis-Code Design Criterion for Reliable Data Transmission over Noisy Discrete Memoryless Channels", submitted to *IEEE Trans. on Comm.*, 1995.

[4] G.D. Forney, Jr., "The Viterbi algorithm", *Proc. of the IEEE*, vol.61, no.3, pp.268-278, March 1973.



	p	q	Tr $\{\bar{S}(\infty\|\infty)\}$	Tr$\{\bar{S}(\infty\|\infty+2)\}$	BER $(\infty\|\infty)$	BER $(\infty\|\infty+2)$
a)	0.999	0.999	**1.01 -3**	3.86 -4	**9.25 -4**	3.60 -4
b)	0.995	0.9999	**4.42 -4**	4.42 -4	**1.00 -4**	1.00 -4
c)	0.95	0.98	**4.01 -2**	2.59 -2	**2.71 -2**	1.67 -2
d)	0.999	0.999	1.02 -3	**1.84 -5**	9.30 -4	**1.05 -5**
e)	0.995	0.9999	1.44 -3	**8.34 -5**	1.30 -3	**3.50 -5**
f)	0.95	0.98	4.35 -2	**1.95 -2**	3.12 -2	**7.60 -3**

Identification via Compressed Data*

Rudolf Ahlswede[1], En-hui Yang[2], and Zhen Zhang[3]

I. Introduction

In this paper, a combined problem of source coding and identification is considered. To put our problem in perspective, let us first review the traditional problem in source coding theory. Consider the following diagram, where $\{X_n\}_{n=1}^{\infty}$ is an

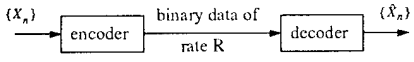


Figure 1: Model for source coding

i.i.d source taking values on a finite alphabet $\mathcal{X}$. The encoder output is a binary sequence which appears at a rate R bits per symbol. The decoder output is a sequence $\{\hat{X}_n\}_1^{\infty}$ which take values on a finite reproduction alphabet $\mathcal{Y}$. In traditional source coding theory, the decoder is required to be able to recover $\{X^n\}_1^{\infty}$ completely or with some allowable distortion. That is, the output $\{\hat{X}_n\}_1^{\infty}$ must satisfy

$$n^{-1}\sum_{i=1}^{n}\rho(X_i,\hat{X}_i)\le d \tag{1}$$

for sufficiently large n, where $\rho:\mathcal{X}\times\mathcal{Y}\to[0,+\infty)$ is a distortion measure and $d\ge 0$ is the allowable distortion. The problem is then to determine the infimum of rate R such that the system shown in Fig.1 can operate in such a way that (1) is satisfied. From rate distortion theory, this infimum is given by the rate distortion function of the source $\{X_n\}_1^{\infty}$.

Let us now consider the system shown in Fig. 2. The se-

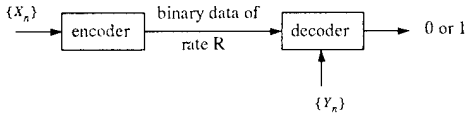


Figure 2: Model for joint source coding and identification.

quence $\{Y_n\}_1^{\infty}$ is a sequence of i.i.d random variables taking values on $\mathcal{Y}$. Known $\{Y_n\}$, the decoder is now required to be able to identify whether or not the distortion between $\{X_n\}$ and $\{Y_n\}$ is less than or equal to d in such a way that two kinds of error probabilities satisfy some prescribed conditions. The problem we are now interested in is still to determine the infimum of rate R such that the system shown in Fig.2 can operate in this way.

II. Formal Formulation of Problem

Let $\{(X_n,Y_n)\}_1^{\infty}$ be a sequence of independent drawings of a pair (X,Y) of random variables taking values on $\mathcal{X}\times\mathcal{Y}$ with joint distribution P_{XY}. Fix $0\le d<E\rho(X,Y)$. An nth-order identification (ID) code $\mathcal{C}_n$ is defined to be a triple $\mathcal{C}_n=(f_n,B_n,g_n)$, where $B_n\subset\{0,1\}^*$ is a prefix set, f_n(called an "encoder") is a mapping from $\mathcal{X}^n$ to B_n, and g_n(called a "decoder") is a mapping from $\mathcal{Y}^n\times B_n\to\{0,1\}$. When $\mathcal{C}_n$ is used in the system shown in Fig.2, its performance can be measured by the following three quantities: the resulting average rate defined by $r_n(\mathcal{C}_n)=En^{-1}$(the length of $f_n(X^n)$), the first kind of error probability defined by $p_{e1}(\mathcal{C}_n)=\Pr\{g_n(Y^n,f_n(X^n))=0|\rho_n(X^n,Y^n)\le d\}$, and the second of error probability defined by $p_{e2}=\Pr\{g_n(Y^n,f_n(X^n))=1|\rho_n(X^n,Y^n)>d\}$.

Let $R\in[0,+\infty)$, $\alpha\in(0,+\infty]$ and $\beta\in(0,+\infty]$. A triple (R,α,β) is said to be achievable if for any $\epsilon>0$, there exists a sequence $\{\mathcal{C}_n\}$ of ID codes, where $\mathcal{C}_n=(f_n,B_n,g_n)$ is an nth-order ID code, such that for sufficiently large n,

$$r_n(\mathcal{C}_n)\le R+\epsilon\ ,\ \ p_{e1}\le 2^{-n(\alpha-\epsilon)}\ \text{ and }\ p_{e2}\le 2^{-n(\beta-\epsilon)}\ ,$$

where as a convention, $\alpha=+\infty(\beta=+\infty$, resp.) means that the first(second, resp.) kind of error probability of $\mathcal{C}_n$ is equal to 0. Let $\mathcal{R}$ denote the set of all achievable triples. In this paper, we are interested in determining the closure $\bar{\mathcal{R}}$ of $\mathcal{R}$. Specifically, we define for each pair (α,β), where $\alpha,\ \beta\in[0,+\infty]$,

$$R^*_{XY}(\alpha,\beta,d)=\inf\{R|(R,\alpha,\beta)\in\bar{\mathcal{R}}\}\ .$$

Our main problem is then the determination of the function $R^*_{XY}(\alpha,\beta,d)$.

III. Main Results

Assume that X and Y are independent. For any $0<d<E\rho(X,Y)$, define $\beta(d)$ by $\beta(d)=\inf D(P||P_{XY})$, where the infimum is taken over all distributions P on $\mathcal{X}\times\mathcal{Y}$ such that $\sum_{x,y}P(x,y)\rho(x,y)\le d$. Let U be a random variable taking values on some finite set $\mathcal{U}$. Let P_{XU} denote the joint distribution of X and U. For any $\alpha\ge 0$, define

$$\mathcal{E}(P_{XU},\alpha,d)=\inf\{D(P_{\tilde{Y}}||P_Y)+I(U;\tilde{Y})\}\ ,$$

where the infimum is taken over all random variables $\tilde{Y}$ taking values on $\mathcal{Y}$ such that $E\rho(X,\tilde{Y})\le d$ and $D(P_{\tilde{Y}}||P_Y)+I(XU;\tilde{Y})\le\beta(d)+\alpha$. Here we make use of the convention that the infimum taken over an empty set is $+\infty$. For any $\beta>0$, let $R(P_X,P_Y,\alpha,\beta,d)$ be the infimum of $I(X;U)$ over all random variables U such that $\mathcal{E}(P_{XU},\alpha,d)\ge\beta$, and let

$$R(P_X,P_Y,\alpha,0,d)=\lim_{\beta\to 0^+}R(P_X,P_Y,\alpha,\beta,d)\ .$$

The following theorem gives a general formula for $R^*_{XY}(\alpha,\beta,d)$.

Theorem 1 *For any $0<d<E\rho(X,Y)$, $0\le\beta<\beta(d)$, and $\alpha\in(0,+\infty]$, the following holds*

$$R^*_{XY}(\alpha,\beta,d)=\bar{R}(P_X,P_Y,\alpha,\beta,d)\ ,$$

where

$$\bar{R}(P_X,P_Y,\alpha,\beta,d)=\lim_{\beta'\to\beta^-}R(P_X,P_Y,\alpha,\beta',d)\ .$$

The converse part of Theorem 1 is related to the general isoperimetric problem. During the process of proving the converse part, we develop a new powerful method for converse-proving in multi-user information theory. For more details, please refer to [1].

*This work was supported in part by NSF Grant NCR-9205265.

[1]Fakultaet fuer Mathematik, Universitaet Bielefeld, 4800 Bielefeld 1, Germany

[2]Dept. of Math., Nankai University, Tianjin 300071, P.R. China.

[3]Commun. Science Institute, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565.

References

[1] R. Ahlswede, E.-H. Yang and Z. Zhang, "Identification via compressed data," Preprint, 1994.

Asymptotics of Fisher Information under Weak Perturbation

Vyacheslav V. Prelov
Institute for Problems of Information Transmission of the RAS,
19 Bol'shoi Karetnyi, Moscow 101447, Russia

Edward C. van der Meulen
Department of Mathematics, Katholieke Universiteit Leuven,
Celestijnenlaan 200B, 3001 Heverlee, Belgium

Abstract — **An asymptotic expression is derived for the Fisher information of the sum of two independent random variables X and Z_ϵ when Z_ϵ is small, under some regularity conditions on the density of X and conditions on the moments of Z_ϵ. Using this result for the case $Z_\epsilon = \epsilon Z$, some asymptotic generalization of De Bruijn's identity is obtained.**

I. Introduction

The Fisher information of a random variable Y with absolutely continuous density f_Y is given by

$$J(Y) = \int_{-\infty}^{\infty} \left[\frac{f_Y'(y)}{f_Y(y)}\right]^2 f_Y(y)dy. \qquad (1)$$

It plays an important role in information theory and statistics. Under certain regularity assumptions, the Fisher information of an additive noise random variable characterizes the main term in the asymptotic expansion of the Shannon mutual information between the input and output signal of an additive noise channel when the input signal is weak [1,2,3]. Fisher information also appears in the well-known Cramér-Rao inequality.

II. Problem Formulation

If $Y = X+Z$, with X and Z independent random variables, an explicit calculation of the integral (1) is impossible in general. Therefore, it is of interest to investigate the asymptotic behavior of $J(Y)$, when the perturbation Z of X is weak in the sense that $Z = Z_\epsilon$ and $E(Z_\epsilon^2) = \epsilon^2 \to 0$. In this paper we derive an asymptotic expansion for the Fisher information $J(X + Z_\epsilon)$ in terms of the probability density function (pdf) of X and higher moments of Z_ϵ, if certain conditions are satisfied. The similar problem of deriving an asymptotic expression for the differential entropy $h(X+Z_\epsilon)$ of the sum of two independent random variables X and Z_ϵ when Z_ϵ is small has been investigated in [4].

III. Main Result

Without loss of generality we assume $E(X) = E(Z_\epsilon) = 0$. Suppose $E(Z_\epsilon^2) = \epsilon^2$, and $E|Z_\epsilon/\epsilon|^{n+\gamma} \le c < \infty$ for some integer $n \ge 2$, some constant c and $0 < \gamma \le 1$. Let X have a bounded pdf $f_X(x) = f(x)$, which has bounded continuous derivatives $f^{(k)}(x)$ for $k = 1, ..., n+2$. Then, under some additional conditions on $f(x)$ (which hold for a large class of smooth densities), and if X and Z_ϵ are independent, the following asymptotic expansion holds as $\epsilon \to 0$:

$$J(X + Z_\epsilon) = J(X) + A_m(X, \{E(Z_\epsilon^k)\}) + o(\epsilon^m) \qquad (2)$$

for some integer $m \ge 1$. $A_m(X, \{E(Z_\epsilon^k)\})$ depends on $f(x)$ and $E(Z_\epsilon^k), k = 2, ..., m$. For $m = 2$, (2) becomes

$$J(X + Z_\epsilon) = J(X) + L(X)\epsilon^2 + o(\epsilon^2), \quad \epsilon \to 0, \qquad (3)$$

where $L(X)$ is an integral expression involving $f(x)$ and its first three derivatives. For example, if X is Gaussian with variance σ^2, the above expansion yields :

$$J(X + Z_\epsilon) = \frac{1}{\sigma^2} - \frac{\epsilon^2}{\sigma^4} + o(\epsilon^2). \qquad (4)$$

IV. Some asymptotic generalization of De Bruijn's identity

For the special case $Z_\epsilon = \epsilon Z$ the asymptotic expansion (2) can be written as

$$J(X + \epsilon Z) = J(X) + B_m(X, \{E(Z^k)\}, \{\epsilon^k\}) + o(\epsilon^m) \qquad (5)$$

where B_m depends on $f(x)$, the moments $E(Z^k)$ and the powers $\epsilon^k, 2 \le k \le m$. Also, when Z has a Gaussian distribution with unit variance, X has a pdf with finite variance, and X and Z are independent, De Bruijn's identity holds :

$$\frac{dh(X + \epsilon Z)}{d(\epsilon^2)} = \frac{1}{2} J(X + \epsilon Z). \qquad (6)$$

Thus, by using the expansion (5) if Z is Gaussian and substituting it in the integral version of (6), we obtain an asymptotic expansion for $h(X + \epsilon Z)$. This expansion coincides with the expansion for $h(X + \epsilon Z)$ obtained in [4] if Z is Gaussian. Moreover, by comparing both the asymptotic expansion for $h(X + \epsilon Z)$ in [4] and the one derived here for $J(X + \epsilon Z)$ for non-Gaussian Z, we obtain some asymptotic generalization of De Bruijn's identity to the case where Z is non-Gaussian.

References

[1] V.V. Prelov, "Asymptotic behavior of the capacity of a continuous channel with a large amount of noise", *Probl. Peredachi Inform.*, Vol. 6, No. 2, pp. 40-57, 1970.

[2] I.A. Ibragimov and R.Z. Khas'minsky, "Weak signal transmission in a memoryless channel", *Probl. Peredachi Inform.*, Vol. 8, No. 4, pp. 28-39, 1972.

[3] V.V. Prelov and E.C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals", *IEEE Trans. Inform. Theory*, Vol. 39, No. 5, pp. 1728-1735, 1993.

[4] V.V. Prelov, "Asymptotic expansion for the mutual information and for the capacity of continuous memoryless channels with weak input signal", *Probl. Control and Inform. Theory*, Vol. 18, No. 2, pp. 91-106, 1989.

A Matrix Form of the Brunn-Minkowski Inequality

Ram Zamir and Meir Feder

330 E&TC Building, Cornell University, Ithaca, NY 14853 . e-mail: zamir@ee.cornell.edu
Dept. of Electrical Eng. - Systems, Tel-Aviv University, Tel-Aviv 69978 Israel . e-mail: meir@eng.tau.ac.il

I. Introduction

The well known *Brunn-Minkowski Inequality* (BMI), is one of the basic inequalities in geometry. Its formal statement is the following. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two sets in $\mathcal{R}^d$. Then,

$$\mu(\mathcal{A}_1+\mathcal{A}_2)^{1/d} \geq \mu(\mathcal{A}_1)^{1/d}+\mu(\mathcal{A}_2)^{1/d} \ . \qquad (1)$$

where $\mu(\mathcal{A})=\int_{x\in\mathcal{A}} dx$ is the (d-dimensional) volume of $\mathcal{A}$, and $\mathcal{A}_1+\mathcal{A}_2=\{x+y : x\in\mathcal{A}_X, y\in\mathcal{A}_N\}$ is the *Minkowski sum* of $\mathcal{A}_1$ and $\mathcal{A}_2$. This sum may be interpreted as the geometric convolution of the two regions. Equality in (1) holds if the two regions are convex and proportional, e.g., if they are balls or cubes (with parallel edges). For $d=1$, this condition is reduced to the simple case where $\mathcal{A}_1$ and $\mathcal{A}_2$ are intervals (and not, e.g., a union of intervals).

The BMI is dual in some sense to the Entropy-Power Inequality (EPI) [1], which lower bounds the entropy-power of the sum of independent random variables. In [2] a matrix form for the EPI was derived, and some of its applications have been pointed out. Analogously, we derive in this work a matrix form for the BMI, and discuss its applications.

II. Linear Transformation of Sets and the Matrix BMI

We first introduce the matrix form of the Minkowski sum. Let $\underline{\mathcal{A}}^t=(\mathcal{A}_1\dots\mathcal{A}_n)$ be a vector, whose n components are d-dimensional sets. We define a linear transformation of $\mathcal{A}_1\dots\mathcal{A}_n$ as

$$T\underline{\mathcal{A}}=\{T\underline{x} : x_i\in\mathcal{A}_i \text{ for } i=1\dots n\} \ , \qquad (2)$$

where T is an $m\times n$ matrix. In particular, $t\mathcal{A}$ means scaling the coordinates of $\mathcal{A}$ by the scalar t. Note that $T\underline{\mathcal{A}}$ is an md-dimensional shape. Denote the volumes of the shapes by $\mu(\mathcal{A}_i)=\mu_i, i=1\dots n$. Following simple laws of integration, the md-dimensional volume of $T\underline{\mathcal{A}}$, in the particular case $m=n$, is $\mu(T\underline{\mathcal{A}})=|T|^d\cdot\mu(\underline{\mathcal{A}})=|T|^d\cdot\prod_{i=1}^n\mu_i$, where $|\cdot|$ denotes the absolute value of the determinant. For the general case, we suggest the following matrix generalization of the BMI:

Theorem 1 (Matrix-BMI): *Let $\underline{\widetilde{\mathcal{A}}}^t=(\widetilde{\mathcal{A}}_1\dots\widetilde{\mathcal{A}}_n)$ be a vector of d-dimensional* cubes *whose edges parallel the axes, and whose volumes are the same as of $\mathcal{A}_1\dots\mathcal{A}_n$, i.e., $\mu(\widetilde{\mathcal{A}}_i)=\mu_i, i=1\dots n$. Then*

$$\mu(T\underline{\mathcal{A}})^{1/d}\geq\mu\left(T\underline{\widetilde{\mathcal{A}}}\right)^{1/d}=\sum_{i=1}^{\binom{n}{m}}|\widetilde{T}_i| \qquad (3)$$

where $\widetilde{T}=T\cdot L$, L is an $n\times n$ diagonal matrix whose diagonal elements are $\mu_1^{1/d}\dots\mu_n^{1/d}$ (the edges' lengths of the cubes $\widetilde{\mathcal{A}}_1\dots\widetilde{\mathcal{A}}_n$), and $\left\{\widetilde{T}_i, i=1\dots\binom{n}{m}\right\}$ is the set of all possible $m\times m$ sub-matrices of $\widetilde{T}$, obtained by choosing m out of the n columns of $\widetilde{T}$.

For $m=1$, (3) reduces to $\mu\left(\sum_{i=1}^n t_i\mathcal{A}_i\right)^{1/d}\geq\sum_{i=1}^n|t_i|\mu_i^{1/d}$, i.e., to the regular BMI (1). Equality in (3) holds in each one (or in a mixture) of the following cases: if $\mathcal{A}_1\dots\mathcal{A}_n$ are cubes whose faces parallel each other; if (after removing the all zero columns of $\widetilde{T}$, if any) $m=n$; or if $\widetilde{T}$ does not have a full row rank, where then $\mu(T\underline{\mathcal{A}})=0$. Theorem 1 is proved via a double induction over the dimensions of T, using a conditional form of the BMI, analogously to the proof of the matrix-EPI in [2].

In order to appreciate the usefulness of Theorem 1, consider the following example. Let $\underline{\mathcal{A}}=(\mathcal{A}_1\dots\mathcal{A}_n)^t$ and $\underline{\mathcal{B}}=(\mathcal{B}_1\dots\mathcal{B}_n)^t$, where $\mathcal{A}_1\dots\mathcal{A}_n,\mathcal{B}_1\dots\mathcal{B}_n$ are d-dimensional shapes of unit volume, and let T_1 and T_2 be $n\times n$ matrices. Consider the volume of the sum $T_1\underline{\mathcal{A}}+T_2\underline{\mathcal{B}}$. A direct application of the regular BMI (1) gives

$$\begin{aligned}\mu(T_1\underline{\mathcal{A}}+T_2\underline{\mathcal{B}})^{1/nd} &\geq \mu(T_1\underline{\mathcal{A}})^{1/nd}+\mu(T_2\underline{\mathcal{B}})^{1/nd}\\ &= |T_1|^{1/n}+|T_2|^{1/n} \ . \qquad (4)\end{aligned}$$

On the other hand, we may view the sum $T_1\underline{\mathcal{A}}+T_2\underline{\mathcal{B}}$ as a transformation of the $2n$-dimensional vector $(\mathcal{A}_1\dots\mathcal{A}_n,\mathcal{B}_1\dots\mathcal{B}_n)$ by the $n\times 2n$ matrix $T=(T_1;T_2)$. Theorem 1 may then be used to obtain

$$\mu(T_1\underline{\mathcal{A}}+T_2\underline{\mathcal{B}})^{1/d}\geq\mu\left(T_1\underline{\widetilde{\mathcal{A}}}+T_2\underline{\widetilde{\mathcal{B}}}\right)^{1/d}=\sum_{i=1}^{\binom{2n}{n}}|(T)_i| \ . \qquad (5)$$

But, by Theorem 1, for $\mathcal{A}_1\dots\mathcal{B}_n$ cubes of unit volume, (5) becomes an equality, while (4) remains an inequality (unless T_1 and T_2 are proportional). We conclude, then, that (5) is a tighter lower bound than (4).

As in the case of other information-theoretic inequalities [1], the new matrix BMI can be used to derive interesting inequalities for determinants. One such example is the inequality just discussed between the right hand sides of (4) and (5). To obtain another inequality, we apply the matrix BMI to a linear transformation of rectangular parallelepipeds while substituting the expression for its exact volume (which is computable in this case). Finally, we note that the matrix BMI can be used to lower bound the volume of a projection of a lattice cell, and so it can find applications in calculating the effective number of codewords of lattice constellations or lattice quantizers satisfying spectral constraints.

[0]This research was supported in part by the Wolfson Research Awards, administered by the Israel Academy of Science and Humanities.

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] R. Zamir and M. Feder. A generalization of the Entropy Power Inequality with applications. *IEEE Trans. Information Theory*, IT-39:1723–1727, Sept. 1993.

The Influence of the Memory for a Special Permutation Channel

Ulrich Tamm
Dept. of Mathematics, University of Bielefeld, P. O. Box 100131, 33501 Bielefeld, Germany

Consider the following model of a permutation channel. In each time unit two sources produce one bit each (0 or 1 with probability P(0) = P(1) = 0.5). These two bits arrive at an organizer, who in the same time unit has to output one bit. The other bit he may store in some memory device. If it is possible the output bit must be a 1. So if the arriving bits are 11, 10 or 01, then the organizer will send a 1 for sure. If both sources produce a 0, then the organizer may send a 1, which is stored in the memory device (and the size of the memory will be reduced by one bit in this case). If this is not possible, then the organizer must send a 0.

A natural question is: How much influence does the size of the memory have on the behaviour of the sequence of bits transmitted by the organizer? As a simple measure for the influence of the memory we consider the expected value of the first occurence of a 0 in this sequence.

If there is **no memory** at all, then this expected value is 4, since in this case we have a geometric distribution with parameter 0.25 as probability that a 0 is transmitted in each time unit.

If the memory device can store every incoming bit (i.e. , **the size of the memory is linear in time**), it turns out that this expected value does not exist. To see this, observe that the bits produced by the two sources yield a sequence $(b(j))_{j=1}^{\infty}$ of 1's and -1's, if we represent a 0 by a -1 and let the bits produced by the first source take the odd positions and the bits produced by the second source take the even positions in the sequence. Two necessary conditions for the occurence of the first 0 at time t are i)$\sum_{j=1}^{2(t-1)} b(i) = 0$ (i. e., the memory is exhausted at time $t-1$) and ii) all partial sums $\sum_{j=1}^{2(i-1)} b(j), i = 1, \ldots, t-2$ are nonnegative (i. e., no 0 has been transmitted before). By the Ballot Theorem the number of $\{1,-1\}$–sequences fulfilling i) and ii) is just the number $\frac{1}{t+1} \cdot \binom{2t}{t}$. Since there are 4^t possible sequences $(b(j))_{j=1}^{2t}$ until time t, the probability that the first 0 occurs at time t is $\frac{\binom{2t}{t}}{(t+1)4^t}$. By Stirling's formula $\binom{2t}{t} \sim \frac{4^t}{\sqrt{t}}$ and the expected value for the first occurence of a 0, $\sum_{t=1}^{\infty} \frac{\binom{2t}{t}}{(t+1)4^t} \cdot t$ does not exist, since the single summands are about $\sqrt{t}$.

If **the size of the memory is limited by some constant** k, the probability for the occurence of the first 0 at time t is $\frac{a(0,t-1)}{4^t}$, where $a(0,t-1)$ is the number of all sequences produced by the two sources leading to the all-one sequence of bits transmitted by the organizer with memory size 0 at time $t-1$.

Analogously, $a(m,t)$ is defined for every time $t = 1, 2, \ldots$ and memory size $m = 0, \ldots, k$. In each time unit the source outputs 01 and 10 do not change the size of the memory, 00 decreases the memory by one bit (and is forbidden for $m = 0$), and 11 increases the memory size by one bit if $m < k$ (and does not change the memory if $m = k$). So we obtain recursion formulae for the numbers $a(m,t)$ which can be written in matrix form as

$$\begin{pmatrix} a(0,t) \\ \vdots \\ a(k,t) \end{pmatrix} = A_k \cdot \begin{pmatrix} a(0,t-1) \\ \vdots \\ a(k,t-1) \end{pmatrix}$$

where $A_k = \begin{pmatrix} 2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 1 & 2 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 2 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 & 3 \end{pmatrix}$.

The behaviour of $a(0,t)$ is essentially determined by the largest eigenvalue of the matrix A_k which can be calculated to be $4 \cdot cos^2(\frac{1}{4k+6}\pi)$.

So for size of memory bounded by $k = 0, 1, \ldots$, we obtain a sequence of expected values for the occurence of the first 0 $(E_k)_{k=0}^{\infty}$ with

$$E_k = \sum_{t=1}^{\infty} \frac{a(0,t-1)}{4^t} \cdot t \sim \sum_{t=1}^{\infty} (cos^2(\frac{1}{4k+6}\pi))^{t-1} \cdot t \cdot \frac{1}{4} < \infty$$

In the special case $k = 1$ it turns out that $a(0,t) = 5^t \cdot F_{2t}$, where F_{2t} denotes the $2t$-th Fibonacci number.

Since the sequence $(E_k)_{k=0}^{\infty}$ is divergent, it is immediate that the expected value for the occurence of the first 0 in the sequence of bits transmitted by the organizer does not exist, **if the size of the memory is bounded by a function** $f(t)$ **which exceeds every** $k > 0$ **from some** t_0 **on.**

One might also consider the general case in which in each time unit s letters from a finite alphabet arrive at the channel and $t \leq s$ letters have to be transmitted. For $s = t = 1$ and constant memory size this model has been discussed (under different aspects) in [1] (see also [2]).

References

[1] *Ahlswede,R., Ye, J. P., and Zhang, Z.*, Creating order in sequence spaces with simple machines, Information and Computation, Vol. 89 (1), pp. 47–94, 1990.

[2] *Ahlswede, R. and Zhang, Z.*, Contributions to a theory of ordering for sequence spaces, Problems of Control and Information Theory, Vol. 18 (4), pp. 197–221, 1989

The relation of description rate and investment growth rate

Elza Erkip and Thomas M. Cover[1]
Stanford University, Information Systems Lab, Stanford, CA 94305-4055
elza@isl.stanford.edu, cover@isl.stanford.edu

Abstract — **We have shown that if one invests in the outcome of a random variable X, where investment consists of gambling at any odds, then every bit of description of X increases the doubling rate by one bit. However, if the provider of the information has access only to V, a random variable jointly distributed with X, then this maximal efficiency is not generally possible. We find the increase $\Delta(R)$ in doubling rate for a description of V at rate R for the jointly Gaussian and jointly binary cases. We investigate the extension to multivariate Gaussian random variables. We prove a general result for the derivative of $\Delta(R)$ at $R = 0$.**

We then consider the problem in which there are k separate encoders and each observes a random variable V_i correlated with X. We find how efficiently these encoders, without cooperation, help the investor who is interested in X.

SUMMARY

Suppose one gambles on the outcome of a random variable X. The investor distributes his wealth according to $b(x)$ and the investment pays odds of $o(x)$ for one. Also suppose that the description of another random variable V, which has a known joint distribution with X, at the rate of R bits is allowed. Let $\Delta(R)$ be the maximum increase in the doubling rate from no description to a description of rate R. It can be seen that $\Delta(R)$ is a concave and nondecreasing function of R. We can show [2] that

$$\Delta(R) = \max_{p(\tilde{v}|v,x):\ I(V;\tilde{V})\le R,\ \tilde{V}\to V\to X} I(\tilde{V};X).$$

We define *initial efficiency* as the derivative of $\Delta(R)$ at the origin. Initial efficiency is the maximum possible increase in $\Delta(R)$ per bit of description. For $V = X$, $\Delta(R) = R$; hence the efficiency is 1. However, for a general V, the efficiency is generally less than 1. We find $\Delta(R)$ and examine the efficiency of the jointly binary and Gaussian cases.

Theorem 1 *Suppose V and X are both Bernoulli($\frac{1}{2}$) random variables associated by a binary symmetric channel with crossover probability p. The $\Delta(R)$ curve is given by*

$$(R, \Delta(R)) = (1 - h(\alpha), 1 - h(\alpha * p))$$

where $0 \le \alpha \le 1$, h is the binary entropy function and $*$ is the cascade operation.

We use a lemma by Wyner and Ziv, known as 'Mrs. Gerber's Lemma' [4] to prove the optimality of the descriptions in the above theorem. The initial efficiency can be calculated as $(1 - 2p)^2$.

Theorem 2 *Suppose X and V are jointly Gaussian with correlation ρ. Then*

$$\Delta(R) = \frac{1}{2}\log\left(\frac{1}{1 - \rho^2(1 - 2^{-2R})}\right).$$

The proof of optimality in the Gaussian problem requires a lemma by Bergmans, which is a conditional version of the entropy power inequality [1]. We note that the initial efficiency is ρ^2.

A natural generalization of this theorem is to multivariate Gaussian. Suppose $V^n \sim N(0, K_V)$, $Z^n \sim N(0, K_Z)$, V^n and Z^n are independent and $X^n = V^n + Z^n$. By changing the coordinate system , we can obtain diagonal covariance matrices and hence transform the problem to one on parallel subchannels with a total rate constraint. The solution is given by water-filling in the entropy domain. We distribute the total rate so that the derivative of $\Delta(R)$ with respect to R at the operating point is the same for all the subchannels used.

We note that in all the problems examined, the initial efficiency is related to the correlation between V and X. We define the *maximal correlation* between V and X as the supremum of $Ef(X)g(V)$, where the supremum is over all functions f and g such that $Ef(X) = Eg(V) = 0$ and $Ef^2(X) = Eg^2(V) = 1$. Maximal correlation depends only on the joint distribution of V and X and is independent of the actual labeling. Conditions under which the maximal correlation can be attained have been investigated by Renyi [3]. Our next theorem examines the relationship between the initial efficiency and maximal correlation.

Theorem 3 *Initial efficiency is equal to the square of the maximal correlation between V and X.*

Next we consider k separate senders. We are interested in the increase in the doubling rate, Δ, for gambling on X when sender i observes V_i correlated with X and the senders operate at respective rates $R_1, \ldots, R_k$. We prove an achievable region for $(R_1, \ldots, R_k, \Delta)$, and show that a Slepian-Wolf type of rate region is achievable for this investment problem.

REFERENCES

[1] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Transactions on Information Theory*, vol. 20, pp. 279-280, 1974.

[2] T. M. Cover and E. Erkip, "Information efficiency in investment," *Proceedings of 1995 IEEE International Symposium on Information Theory.*

[3] A. Rényi, "On measures of dependence," *Acta Mathematica*, vol. 10, pp. 441-451, 1959.

[4] A. A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications I," *IEEE Transactions on Information Theory*, vol. 19, pp. 769-771, 1973.

[1]This work was supported by NSF Grant NCR-9205663, ARPA Contract J-FBI-94-218 and JSEP Contract DAAH04-94-G-0058.

Multi-way Alternating Minimization

Raymond W. Yeung[1] and Toby Berger[2]

Abstract — **In a K-way minimization problem, we are interested in finding**

$$\min_{x_1 \in S_1} \cdots \min_{x_K \in S_K} f(x_1, \ldots, x_K),$$

where f is continuous and bounded from below, and S_i is a compact convex set in $\mathbb{R}^{n_i}$, $1 \le i \le K$. In a paper by Csiszar and Tusnady [2], a similar problem with somewhat less stringent conditions was studied for $K = 2$, where it was shown that an alternating minimization algorithm converges to the infimum provided a certain geometric condition is satisfied. In this paper, we take an approach (also with strong geometric flavor) different from theirs, which enables us to obtain a sufficient condition for an alternating minimization algorithm to converge to the minima. In particular, we show that it is sufficient for f to be convex. The Arimoto-Blahut algorithm for computing channel capacity is discussed as an example of application of our results.

I. An Alternating Minimization Algorithm

In a K-way minimization problem, we are interested in

$$f^* = \min_{x_1 \in S_1} \cdots \min_{x_K \in S_K} f(x_1, \ldots, x_K),$$

where S_i is a subset of $\mathbb{R}^{n_i}$, $1 \le i \le K$. Here x_i is an n_i-tuple. We assume that S_i is compact and convex, and f is continuous and bounded from below. Let $\mathbf{x} = (x_1, \ldots, x_K)$ be a generic point in $\prod_{j=1}^K S_j$. For each $\mathbf{x}$, define for $1 \le i \le K$

$$x_i^*(\mathbf{x}) = x_i^*(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_K) \in S_i$$

such that $x_i^*(\mathbf{x})$ achieves

$$\min_{y \in S_i} f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_K)$$

when $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_K$ are fixed, and let

$$g_i(\mathbf{x}) = (x_1, \ldots, x_{i-1}, x_i^*(\mathbf{x}), x_{i+1}, \ldots, x_K).$$

Let $g(\mathbf{x}) = g_{i^*}(\mathbf{x})$, where $1 \le i^* \le K$ and

$$f(g_{i^*}(\mathbf{x})) = \min_{1 \le i \le K} f(g_i(\mathbf{x})),$$

and define

$$\Delta f(\mathbf{x}) = f(\mathbf{x}) - f(g(\mathbf{x})).$$

Since $f(\mathbf{x}) \ge f(g_i(\mathbf{x}))$ for $1 \le i \le K$, $\Delta f(\mathbf{x}) \ge 0$.

Let $\mathbf{x}_0$ be any point in $\prod_{j=1}^K S_j$, and $\mathbf{x}_k = g(\mathbf{x}_{k-1})$ for $k \ge 1$. This paper is devoted to study of this "greedy" alternating minimization algorithm. We show that, under suitable conditions, $f(\mathbf{x}_k) \to f^*$ as $k \to \infty$. Henceforth, we will abbreviate $f(\mathbf{x}_k)$ to f_k.

[1] Department of Information Engineering, the Chinese University of Hong Kong, N.T., Hong Kong; whyeung@ie.cuhk.hk

[2] School of Electrical Engineering, Cornell University, Ithaca, NY 14853, USA; email:berger@ee.cornell.edu

II. Sufficient Conditions for Convergence

Since f_k is non-increasing and f is bounded from below, f_k must converge to some value. We now state a condition that is sufficient for $f_k \to f^*$.

(SC-1) Let $\mathbf{x}^* = (x_1^*, \ldots, x_K^*) \in \prod_{j=1}^K S_j$ achieves f^*. For any $\mathbf{x} = (x_1, \ldots, x_K) \in \prod_{j=1}^K S_j$ such that $f(\mathbf{x}) > f^*$, there exists y which is a convex combination of x_i^* and x_i for some $1 \le i \le K$ ($y \in S_i$ since $x_i^*, x_i \in S_i$ and S_i is convex) such that

$$f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_K) < f(x_1, \ldots, x_K).$$

It is not difficult to show that if (SC-1) is satisfied, then $\Delta f(\mathbf{x}) > 0$ whenever $f(\mathbf{x}) > f^*$. Therefore, the algorithm cannot be trapped at a local minimum. Using the assumption that f is continuous and that $S_j, 1 \le j \le K$ is compact, it can be shown that f_k always converges to f^*.

We have further proved that (SC-1) is satisfied if f is convex in $x_1, \ldots, x_K$. This condition is stronger than (SC-1), but it has the advantage that it is easy to check. In the next section, we will show how this condition can be used to prove the convergence of the Arimoto-Blahut algorithm for computing channel capacity.

III. An Example of Application

Let $\{Q(k|j)\}$ be the set of transition probabilities of a channel. Then the channel capacity is given by

$$\max_{\mathbf{p}} \max_{\mathbf{q}} \sum_j \sum_k p(j) Q(k|j) \log \frac{q(j|k)}{p(j)},$$

(see Blahut [1]), which is equivalent to the negative of

$$\min_{\mathbf{p}} \min_{\mathbf{q}} \sum_j \sum_k p(j) Q(k|j) \log \frac{p(j)}{q(j|k)}.$$

Let

$$f(\mathbf{p}, \mathbf{q}) = \sum_j \sum_k p(j) Q(k|j) \log \frac{p(j)}{q(j|k)}.$$

The Arimoto-Blahut algorithm is a special case of the algorithm described in Section 1 with $K = 2$; it is easy to check that all the required conditions are satisfied. Using the results in Section II, in order to show that the algorithm converges to the channel capacity, we only have to show that f is convex in both $\mathbf{p}$ and $\mathbf{q}$. This can be done by invoking the *log-sum inequality* on p. 29 of Cover and Thomas' textbook [3]. So, the algorithm does converge to the channel capacity.

References

[1] R. E. Blahut, *Theory and Practice of Error Control Codes.* Addison-Wesley, Reading, MA, 1983.

[2] I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures," *Statistics and Decisions*, Supplementary Issue 1: 205-237, 1984.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

Generating non-Markov random sources with high Shannon entropy

A. G. D'yachkov, V. M. Sidelnikov

Moscow State University, Faculty of Mechanics and Mathematics; Vorob'evy Gory, 119899, Moscow, Russia,
E-mail: sid@inria.phys.msu.su

I. Introduction

We study the properties of a sequence of dependent random variables generated with the following scheme. A sequence of independent identically distributed random variables α arrives at the input of a k-register; the random variables take values 0 and 1 with probabilities not equal to 1/2. The sequence η of random variables generated by the k-register for $k \geq 2$ is a stationary random sequence with dependent components. The sequence η is taken as the input to a memoryless binary symmetrical channel with input-independent noise, i.e. η is added coordinatewise modulo 2 to a sequence β of independent identically distributed random variables that also take the values 0 and 1 with probabilities not equal to 1/2 and are independent with the sequence α.

In this paper we derive upper and lower estimates for the entropy of the stationary non-Markov random source identified with the channel output. The upper estimate is based on the well-known subadditivity property [1] of the entropy of a finite-dimensional distribution. The main result is the proof of nontrivial lower estimate of the entropy for two particular k-registers: $k = 2$ and $k = 3$. If the probabilities of 0 and 1 in the sequences α and β are close to 1/2, this estimate shows that the entropy of the source increases when k grows from 1 to 3. Pre-transformation of α by the k-register, $k > 1$, yields the increase of the entropy of the additive source $\alpha + \beta$ over the case $k = 1$. This property of increase of the entropy is significant for constructing a strong random source from several "weak" ones.

II. Results

Let a and b, $0 < a, b < 1$, be real numbers. By α_i, β_i for $i = 1, 2, \ldots$, we denote independent random variables that take the values 0 and 1 with probabilities

$$\begin{aligned} \mathbf{P}\{\alpha_i = x\} &= \frac{1 + a(-1)^x}{2}, \\ \mathbf{P}\{\beta_i = x\} &= \frac{1 + b(-1)^x}{2}, \quad x = 0, 1. \end{aligned}$$

Take an integer $k \geq 1$ and consider a stationary discrete random source $\eta^{(k)} = (\eta_1^{(k)}, \eta_2^{(k)}, \ldots)$, where $\eta_i^{(k)}$, $i \geq 1$, take the values 0 or 1 and are defined as

$$\eta_i^{(k)} = \beta_i + \sum_{j=0}^{k-1} \alpha_{i+j} \bmod 2 \qquad (1)$$

The symbol

$$Q_k^{(a,b)}(x(n)) = \mathbf{P}\{\eta^{(k)}(n) = x(n)\}$$

denotes the finite-dimensional distribution of source (1). The entropy of source (1) is defined as

$$H_k(a,b) = \lim_{n \to \infty} \frac{1}{n} H(\eta^{(k)}(n)),$$

where

$$H(\eta^{(k)}(n)) = -\sum_{x^n} Q_k^{(a,b)}(x(n)) \ln Q_k^{(a,b)}(x(n))$$

is the entropy of finite-dimensional distribution.

We define a binary entropy as

$$h(\delta) = -\delta \ln \delta - (1 - \delta) \ln(1 - \delta).$$

Theorem 1. *For any $k \geq 3$*

$$H_k(a,b) \leq \min\left(h\left(\frac{1 - ba^k}{2}\right), h\left(\frac{1 - b^2a^2}{2}\right)\right).$$

Theorem 2. *For $k = 2$,*

$$H_2(a,b) \geq \ln 2 - \ln\left(1 + \frac{b^2a^4}{1 - b^2}\right).$$

Theorem 3. *For $k = 3$,*

$$H_3(a,b) \geq \ln 2 - \ln\left(1 + \frac{a^4b^4 + a^6b^2(1 - a^2b^2)}{(1 - a^2b^2)^2 - b^4}\right).$$

Theorem 4. *Assume that the probabilities of 0 and 1 in sequences α and β are close to 1/2, i.e., the parameters a and b are close to zero. Then for $k = 1, 2, 3$ the entropy $H_k(a,b)$ increases with the growth of k.*

References

[1] R. Gallager, "Information Theory and Reliable Communication", Willey, New York (1968).

[2] A. G. D'yachkov, V. M. "Sidelnikov, Entropy of Some Binary Sources", *Probl. Inform. Transmission,* vol. 28, no. 4 (1992).

[0]This research is supported by RFBR Grant 93-01-00492, and by ISF Grant MEF300.

On the Equivalence of Some Different Definitions of Capacity of the Multiple-Access Collision Channel with Multiplicity Feedback

Miklós Ruszinkó [1]
Comp. & Aut. Res. Inst., Hung. Ac. of Sciences
Budapest, POB 63, Hungary-1518

Abstract — **Various different definitions were investigated in Random Multiple Access theory for capacity of the multiple-access collision channel. However, as it was pointed out by Tsybakov [4], almost nothing about the relations between the various definitions is known. In this paper we try to fullfil this gap showing about some widely used capacity definitions that they are equivalent.**

I. Introduction

The study of collision channels, also called random-access channels, started with Abramson's ALOHA system [1] which uses only binary feedback (collision/no collision). Later on this channel model (and its modifications) became a special interest: it was investigated in numerous research articles. The goal of all such papers is to present good conflict resolution algorithms and to get bounds on the efficiency of the best possible ones. For this reason one have somehow to measure how efficient an algorithm is. But different authors measured this quite often in different ways, getting by this different definitions for the *throughput* of an RMA algorithm which is nothing else as one of these measures. This led to the study of different *capacity* notions, since it is, raughly speaking, the best possible throughput which might be achieved. On the other hand it is not obvious at all, that an algorithm beeing efficient (i.e. having a high throughput) in one sense, is also efficient from another point of view as well. In [4] Tsybakov gave an excellent survey about the Random Multiple-Access communication, where he wrote about this problem that "... we know almost nothing about the relations between the various definitions of delay, throughput and capacity". In this paper we will show, that some of the most widely used definitions for the throughput and capacity of the multiple access collision channel are equivalent in the case when the feedback is the multiplicity of the collisions.

II. Summary of Results

Assume that $x_1 \le x_2 \le x_3 \le \dots$ is a random process where x_i is the generating time of the i^{th} packet. We will suppose that the instants of new-packet generations form a Poisson process, i. e. the differences $(x_{i+1} - x_i)$ are independent random variables with the identical distribution

$$Pr\{x_{i+1} - x_i > x\} = e^{-\lambda x}.$$

A *conflict resolution protocol* (or *random multiple access algorithm*) is a retransmission algorithm f for the packets in a collision. The delay δ of a packet is the time from its moment of generation until the moment of its successful transmission. Let δ_i denote the delay of the i^{th} packet. The *delay* of a random multiple access algorithm f is

$$D_f = \limsup_{i\to\infty} \mathbf{E}(\delta_i),$$

where $\mathbf{E}()$ denotes expectation, and its throughput is

$$R_f^1 = \sup\{\lambda : D_f < \infty\}.$$

In the case of *blocked access* the number of active users in subsequent epochs forms a Markov chain $\mathcal{M}$. This implies the following definition for the throughput of a blocked access algorithm:

$$R_f^2 = \sup\{\lambda : \ \mathcal{M} \textit{ is stable}\}.$$

It is very natural to define the throughput as fraction of the number of generated messages and the time is needed to transmit them. More precisely, denote by $X(t)$ the number of generated messages in the time interval $(0,t)$ and by $\gamma(X(t))$ the number of steps which are needed to transmit these messages. Thus the throughput might be also defined [2] as

$$R_f^3 = \liminf_{t\to\infty} \frac{\mathbf{E}X(t)}{\mathbf{E}(\gamma(X(t)))}.$$

The above listed three different throughput notions imply three different capacity definitions in the following way. Let $\mathcal{A}$ denote the set of random multiple access algorithms. The *capacity* of the random multiple access collision channel is defined as

$$C = \sup\{R_f : \ f \in \mathcal{A}\},$$

which supremum can be taken over the different throughputs defined before. Thus we get C^1, C^2, and C^3, resp.

In 1981 Pippenger proved in probabilistic way [2], that there exist an algorithm f, for which $R_f^3 = 1$. Ruszinkó and Vanroose [3] constructed such an algorithm. Let us denote this algorithm by RV. We claim that the following statement holds.

Theorem.

$$R^1(RV) = R^2(RV) = R^3(RV) = 1,$$

thus

$$C^1 = C^2 = C^3 = 1.$$

Consequently these throughput and capacity definitions are *equivalent.*

[1] This work was supported by OTKA Grant T016414

References

[1] N. Abramson, "The ALOHA system - another alternative for computer communications," *AFIPS Conf. Proc., Fall Joint Computer Conf.*, vol. 37, pp. 281-285, 1970.

[2] N. Pippenger, "Bounds on the performance of protocols for a multiple access broadcast channel," *IEEE Tr. Inf. Th.*, vol. IT-27(2), pp. 145-151, 1981.

[3] M. Ruszinkó and P. Vanroose, "A constructive code reaching capacity 1 for the multiple-access collision channel with multiplicity feedback," *Proc. 1995 IEEE ISIT*, Trondheim, Norway, p. 287.

[4] B. S. Tsybakov, "Survey of USSR contributions to random multiple-access communications," *IEEE Tr. Inf. Th.*, vol. IT-31(2), 143-164, 1985.

Matrix approach to the problem of matrix partitioning

S.I. Stasevich, V.N.Koshelev [1]
RAS, Council for Cybernetics,
Moscow, Russia

Abstract — **We derive upper and lower bounds on the number of all variants a rectangular $M \times N$ matrix can be partitioned into fragments. Next the problem of matrix partitioning is considered as a particular example of a more general problem of constructing two-dimensional Markov processes (fields) on discrete rectangular latices. We discuss a matrix- theoretical approach to the problem to explore the structure of discrete fields defined by a given matrix of local interaction.**

In this paper we continue to study the problem formulated in [1]. Let $\phi(M;N)$ be a number of all variants an $M \times N$ rectangular matrix (with empty cells) can be splitted into fragments. Immediate calculations show that

$$\phi(2;2) = 12, \phi(2;3) = 74, \phi(3;3) = 1442, \phi(4;4) \geq 1.7 \times 10^6$$

and so on. Each individual partition of the matrix is considered as an output of a block source with block size $2MN - M - N$ and information rate

$$R(M;N) = \frac{\log_2 \phi(M;N)}{2MN - M - N}.$$

The rate is measured in "bits per edge", because the denominator of $R(M;N)$ is the total number of all internal edges between the cells of the matrix. So defined source is called form source, where "form" means the set of contours resulting from an individual matrix partition [2,3,4]. The exponential behavior of $\phi(M;N)$ may present an interest in image processing [2], statistical mechanics [3] and other applications exploiting different models based on the conception of random fields. In [4] a special technique founded on the theory of Fibonacci numbers was suggested and some upper and lower bounds on $\phi(M;N)$ were obtained.

Now we develop a formal matrix approach to the problem. This approach is based on Perron-Frobenius theory for nonnegative matrices [5]. We introduce special (0,1)-matrices describing a physical process of breaking down of an $M \times N$ matrix into fragments. This four "splitting matrices" are

$$A_{00} = \begin{pmatrix} 1\,0 \\ 0\,1 \end{pmatrix}, A_{01} = A_{10} = \begin{pmatrix} 0\,1 \\ 1\,1 \end{pmatrix}, A_{11} = \begin{pmatrix} 1\,1 \\ 1\,1 \end{pmatrix}.$$

They define the splits allowed to run through a solid state matrix with unbreakable cells. In these terms we prove

Theorem 1. For any integers $M, N = 1, 2, \ldots$

$$\phi(M+1;N+1) = \left\| [\|A_{i^N j^N}\|]^M \right\|,$$

where

$$A_{i^N j^N} = \prod_{n=1}^{N} A_{i_n j_n}, i^N, j^N \in \{0,1\}^N,$$

and $\|.\|$ denotes the sum of all matrix elements.

Then we prove the main result of the paper establishing an exact exponential behavior of $\phi(M;N)$. Let

$$R(\infty;\infty) = \lim_{M,N\to\infty} R(M;N)$$

be asymptotical information rate of the form source.

Theorem 2.

$$R(\infty;\infty) = \frac{log_2\lambda}{2} = 0.8322611,$$

where $\lambda = 3.1700865$ is maximal eigenvalue of the matrix

$$\begin{pmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{pmatrix} = \begin{pmatrix} 1\,0\,0\,1 \\ 0\,1\,1\,1 \\ 0\,1\,1\,1 \\ 1\,1\,1\,1 \end{pmatrix}.$$

The second part of the paper is devoted to a probabilistic modification of the problem. We consider an 4×4 stochastic "splitting matrix" with the same zero elements as in 4×4 matrix shown in Theorem 2 and with arbitrary positive probabilities substituted instead of ones. We show that the maximal entropy rate of the so defined probabilistic form sou asymptotically coincides with the information rate of the deterministic form source shown in Theorem 2.

The 4×4 matrix shown in Theorem 2 reflects the demand of continuity of a contour. We present a more general form of lower and upper bounds on $\phi(M;N)$ defined by an arbitrarily given matrix of the local interaction [6] in the field and show some results of computational experiments.

Finally some possible schemes of the form source coding are discussed.

[1] This work was supported by RFFI Grant 93-012-467.

References

[1] S. I. Stasevich, V. N. Koshelev, "Entropy rate of the Form Sources," *Proceedings ISIT-94*, p. 190.

[2] Ju. P. Pytiev, "A Problem of Morphological Image Analysis," *Technical Report*, Moscow, 1984.

[3] Ya. G. Sinai, "Theory of Phase Transitions;Strong Results," Nauka, Moscow.

[4] S. I. Stasevich, "Number of Forms for Two-Dimentional Images," *PPI*, 1994, vol. 30, is. 4, pp. 90-94.

[5] F. R. Gantmacher, "Theory of Matrices," Nauka, Moscow, 1967.

[6] E. E. Beckenbach, "Applied Combinatorial Mathematics," John Wiley, New York, 1964.

Using the Ideas of the Information Theory to Study Communication Systems of Social Animals

Zhanna I. Reznikova, Boris Ya. Ryabko
Inst. of Anim. Syst. and Ecol., Frunze str.,11, Novosibirsk, 630091, Russia;
Novosibirsk Telecommunication Inst.,Kirov str.,86, Novosibirsk 630102, Russia

I. Introduction

During the last decades, quite new experimental approaches to the study of communication systems of animals have been developed including those based on a direct dialog with animals taught artificial intermediary languages. The use of simple grammatical rules as well as number - related skills at a level of pre - school children have been demonstrated in chimpanzee [1] and in grey parrot [2]. However, the question of existence of a developed natural language in social animals is still open for discussion. We have been suggested quite a different approach to the study of communication systems based on the ideas of the Information Theory.Our recent experiments allowed to evidence the presence of potentially unlimited number of messages in ant " language", and to show ants as being able to use the "text" regularities for information compression [3,4]. In this report we consider plasticity of ant language as well as their numerical competence.

II. Methods

Ants were kept in transparent nests in the laboratory arenas. Each worker was labelled with an individual colour mark. As soon as discovering ants found food, they informed the relatively constant teams of 5 - 8 foragers about it. During experiments ants were fed in setups, consisted of a long " trunk " with equally spaced 25- 40 branches, made of thin plastic sticks.Each branch ended in an empty trough, except for one filled with syrup.To start the experiment, an ant scout was placed at the trough containing food. When it returned to the nest,the duration of the contact between foragers and the scout was measured. As soon as foragers began following the scout, the scout was removed from the arena with tweezers. To avoid odour tracks,the original maze was replaced by an identical one.

III. Ant Numerical Competence and Plasticity of Ant "Language"

It turned out that ants can count within several tens, and that in their " language" there are means of transmitting messages about the number of objects. In all experiments the teams abandoned the nest after they were contacted and moved towards troughs 130 times. In 90 cases the team immediately found the correct way. The probability of finding the food-containing trough randomly is less than 10^{-10}. The relation between the number i of the branch and the duration t of the contact between scout and foragers is linear and described by the equation $t = ai + b$.Note that in modern human languages with decimal numeration the length of the written form of a number i and the time to pronounce the number i are proportional to $logi$, but not to i. Archaic human languages are known to have used another system of numeration. The number "one" was encoded as the word "finger", "two" as "finger, finger" , etc. In this case , the time required to pronounce i is also proportional to i, as in ant "language". Such a large difference between modern human and ant languages does not necessarily show that the latter is primitive; as it is known that in a "reasonable" language the length of a word should correspond to its frequency of occurrence in communications. We then consider to which extent may ant "language" transform to keep this equation valid. In special experiments a horizontal trunk with 40 branches was used, however,the trough was placed on different branches with different frequencies: on two "special" branches (N 10 and N 20) it appeared in about 2 cases of 3. At first the time required to transmit information on the number i of the branch i was proportional to i. But about halfway through the series, the time of transmission of information about the fact that the trough was on a "special" branch became much shorter than in the cases when the trough was on other, seldom used branches. It should be emphasized that the time ceased to be proportional to i, perhaps as a result of a transformation in the communication system of these ants, caused by a change in "numerical" frequency.

References

[1] S.T.Boysen, "Counting and number — related skills in chimpanzees (Pan troglodytes)", *XXIII Internat. Ethological Conference* Torremolinos, Spain, p. 325, 1993.

[2] I.M.Pepperberg, "Acquisition of the same — different concept by an African Grey parrot (Psitacus eritacus): learning with respect to categories of colour, shape and material", *Animal Learn and Behav.*, vol. 15, N 4, pp. 423 - 432 , 1987.

[3] Zh.I.Reznikova, B.Ya. Ryabko, "Information Theory approach to communication in ants", *Sensory Systems and Communication in Arthropods*, Birkhäuser Verlag, Basel, pp. 305 - 308 , 1990.

[4] Zh.I.Reznikova, B.Ya.Ryabko, "Using Shannon Enthropy and Kolmogorov Complexity to Study the Language and Intelligence of Ants", *Proc. 1994 - IEEE . Intern. Symp. on Information Theory.* Norway , p.195.

Fixed-Slope Universal Algorithms for Lossy Source Coding Via Lossless Codeword Length Functions*

En-hui Yang[1], Zhen Zhang[2], and Toby Berger[3]

I. Introduction and Algorithms

Let A be an abstract source alphabet and $\hat{A}$ a finite reproduction alphabet. If $x = (x_i)$ is a finite or infinite sequence of symbols from A or $\hat{A}$ (or of random variables taking their values in these sets), let $x_m^n = (x_m, \cdots, x_n)$ and, for simplicity, write x_1^n as x^n. We denote the set of all n-tuples drawn from $A(\hat{A})$ by $A^n(\hat{A}^n)$. A lossless codeword length function(LCLF) l is a map from $\hat{A}^*$, the set of all finite sequences from $\hat{A}$, to $\{1, 2, \cdots, \}$ satisfying $\sum_{y \in \hat{A}^n} 2^{-l(y)} \leq 1$ for any $n > 0$. Clearly, there exists a one to one correspondence between lossless codeword length functions and prefix codes: for any LCLF l, there exists a prefix code $\phi : \hat{A}^* \to \{0,1\}^*$ such that for any $y \in \hat{A}^*$, $l(y) =$ the length of $\phi(y)$, and vice versa. The well-known examples are the Lempel-Ziv codeword length function $LZ(y^n)$ and the k-th order arithmetic codeword length function $L_{A,k}(y^n)$. Let $\rho : A \times \hat{A} \to [0, +\infty)$ be a single-letter distortion measure. For any stationary, ergodic source μ, let $R(D, \mu)$ and $D(R, \mu)$ denote its rate distortion function and distortion rate function with respect to the fidelity criterion $\{\rho_n\}$ generated by ρ, respectively, where $\rho_n(x^n, y^n) = n^{-1} \sum_{i=1}^n \rho(x_i, y_i)$ for any $x^n \in A^n$ and $y^n \in \hat{A}^n$. For simplicity, we shall assume that a reference letter $a^* \in \hat{A}$ exists for ρ and μ such that $E\rho(X_1, a^*) < \infty$ and that $\sup_{x \in A} \inf_{y \in \hat{A}} \rho(x, y) = 0$.

Corresponding to any LCLF l, three universal lossy data compression schemes are presented in this paper: one is with fixed rate, another is with fixed distortion, and a third is with fixed slope.

A fixed rate universal lossy data compression scheme. Fix $R > 0$. Let $N(R, l)$ be the smallest positive integer such that the set $\{y^n \in \hat{A}^n : l(y^n) \leq nR\}$ is nonempty for all $n \geq N(R, l)$. Let $B_n(l)(n \geq N(R, l))$ consist of all $y^n \in \hat{A}^n$ such that $l(y^n) \leq nR$. In our fixed rate universal lossy data compression scheme, each source sequence $x^n \in A^n$ is quantized into a closest member y^n of $B_n(l)$. There are two different ways for the encoder to encode x^n: (1) The encoder can transmit the index of y^n in $B_n(l)$ using a binary string of length $\lfloor nR \rfloor$; or (2) the encoder can transmit the binary codeword associated with y^n via the LCLF l, adding some dummy digits to ensure overall codeword length $\lfloor nR \rfloor$.

A fixed distortion universal lossy data compression scheme. Fix $D > 0$. For each $n \geq 1$, we think of the entire set $\hat{A}^n$ as a codebook of dimension n and list the elements y^n of $\hat{A}^n$ in order of nondecreasing lossless codeword length $l(y^n)$. For each $x^n \in A^n$, the encoder maps x^n into the binary codeword associated with y^n via the LCLF l, where y^n is the first element in $\hat{A}^n$ such that $\rho_n(x^n, y^n) \leq D$.

A fixed slope universal lossy data compression scheme. Let $\lambda > 0$ be fixed. Our fixed slope universal lossy data compression scheme works as follows: For each $x^n \in A^n$, the encoder first searches the first element y^n in $\hat{A}^n$ which minimizes the cost function $n^{-1} l(y^n) + \lambda \rho_n(x^n, y^n)$ over the whole set $\hat{A}^n$, where $\hat{A}^n$ is assumed to be ordered in some order, and then encodes x^n into the binary codeword associated with y^n. After receiving the binary codeword, the decoder can completely recover y^n and output y^n as a reproduction of x^n. In this way, the resulting rate $r_n(x^n, l, \lambda)$ in bits per sample is then $n^{-1} l(y^n)$; and the resulting distortion $\rho_n(x^n, l, \lambda)$ per sample is $\rho_n(x^n, y^n)$.

II. Optimality

The fixed rate or fixed distortion lossy data compression algorithm mentioned above is just the extension of the corresponding one in [1] to the case of any LCLF. Under some mild conditions on l, similar results to [1] can be proved. In the following, therefore, we focus only on the fixed slope lossy data compression algorithm.

A LCLF l is said to satisfy Condition A if for any stationary, ergodic process $\{Y_i\}_1^\infty$ taking values in $\hat{A}$, $n^{-1} l(Y^n)$ converges with probability one to the entropy rate of $\{Y_i\}_1^\infty$.

Theorem 1 *Let $\lambda > 0$. Let μ be a stationary, ergodic source with the random output $X = \{X_i\}_1^\infty$. If l satisfies Condition A, then as $n \to \infty$,*

(i) $r_n(X^n, l, \lambda) + \lambda \rho_n(X^n, l, \lambda) \to R_\lambda(\mu) + \lambda D_\lambda(\mu)$ *almost surely, where* $D_\lambda(\mu) = \inf\{D | D \geq 0,\ R'_+(D, \mu) > -\lambda\}$ *and* $R_\lambda(\mu) = R(D_\lambda(\mu), \mu)$.

(ii) $r_n(X^n, l, \lambda)(\rho_n(X^n, l, \lambda))$ *converges almost surely to* $R_\lambda(\mu)(D_\lambda(\mu))$, *provided* $(D_\lambda(\mu), R_\lambda(\mu))$ *is the only point on the rate distortion curve such that* $R'_-(D_\lambda(\mu), \mu) \leq -\lambda \leq R'_+(D_\lambda(\mu), \mu)$.

Particularly, Theorem 1 holds for the k-th order arithmetic codeword length function $L_{A,k}$(i. e. , $l = L_{A,k}$) if k is allowed to go to infinity. During the process of proving Theorem 1, we also obtain a very strong sample converse theorem for variable length source coding which implies Kieffer's sample converse theorem and strong converse theorem as corollaries.

The main advantage of this fixed slope universal lossy data compression scheme over the fixed rate(fixed distortion) universal lossy data compression scheme lies in the fact that it converts the encoding problem to a search problem through a trellis and then permits one to use some sequential search algorithms to implement it. Simulation results with the kth order arithmetic codeword length function as a LCLF and the M-algorithm as a sequential search algorithm show that this fixed slope universal algorithm, combined with suitable search algorithms, might be implementable in practice.

*This work was supported in part by NSF Grants NCR-9205265, NCR-9216975, IRI-9005849, and IRI-9310670.

[1]Dept. of Math., Nankai University, Tianjin 300071, P.R. China.

[2]Commun. Science Institute, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565, USA

[3]Dept. of Elec. Engr., Cornell University, Phillips Hall, Ithaca, NY 14853, USA

References

[1] E.-H. Yang and J. C. Kieffer, "Simple universal lossy data compression schemes derived from Lempel-Ziv algorithm," *Submitted to IEEE Trans. Inform. Theory for Publication.*

A Lossy Data Compression Based on an Approximate Pattern Matching

Tomasz Luczak[1] and Wojciech Szpankowski[2]
Mathematical Institute, Polish Academy of Science, 60-769 Poznań, Poland
Dept. Computer Science, Purdue University, W. Lafayette, IN 47907, U.S.A.

Abstract — **A practical suboptimal (variable source coding) algorithm for lossy data compression is presented. This scheme is based on an approximate string matching, and it extends lossless Wyner-Ziv data compression scheme.**

I. Introduction and Main Results

We consider a stationary and ergodic sequence $\{X_k\}_{k=-\infty}^{\infty}$ taking values in a binary alphabet $\Sigma = \{0,1\}$. We write X_m^n to denote $X_m X_{m+1} \dots X_n$. As a measure of fidelity we consider the Hamming distance (however, other fidelity criteria can be easily accommodated into our main results) defined as $d_n(x_1^n, \widetilde{x}_1^n) = (1/n)\sum_{i=1}^n d_1(x_i, \widetilde{x}_i)$ where $d_1(x,\widetilde{x}) = 0$ for $x = \widetilde{x}$ and 1 otherwise $(x, \widetilde{x} \in \Sigma)$. We assume that the maximum allowed distortion is D, and by $R(D)$ we denote the rate-distortion (cf. [1]).

We propose a practical suboptimal lossy data compression scheme that extends the Lempel-Ziv scheme. Our scheme reduces to the following approximate pattern matching problem: Let the "training sequence" or "database sequence" x_1^n be given. Find the longest L_n such that there exists $1 \le i_0 \le n$ in the database satisfying $d(x_{i_0}^{i_0-1+L_n}, x_{n+1}^{n+L_n}) \le D$. This naturally extends Wyner and Ziv [5] (cf. also [4]) idea to lossy situation (cf. also [3]) which is subject of this work.

Actually, the real engine behind this study (and its algorithmic issues) is a probabilistic analysis of an approximate pattern matching problem which we discuss next. Our probabilistic results are confined to the *stationary mixing model* in which two random events defined on two σ-algebra separated by g symbols behave like independent events as $g \to \infty$. We denote by $\alpha(g)$ the *mixing coefficient*, and assume that $\alpha(g) \to 0$ as $g \to \infty$.

It turns out that behavior of L_n is related to two other quantities, namely the shortest path s_n and the height H_n defined in the sequel. The *height* H_n is the length of the longest substring in the database X_1^n for which there exists another substring in the database within distance D. More precisely: the height is equal to the largest N for which there exist $1 \le i,j \le n$ such that $d(X_i^{i-1+N}, X_j^{j-1+N}) \le D$. Let now $\mathcal{W}_k$ be the set of words of length k, and $w_k \in \mathcal{W}_k$. Then, the *shortest path* s_n is the longest k such that for *every* $w_k \in \mathcal{W}_k$ there exists $1 \le i \le n$ such that $d(X_i^{i-1+k}, w_k) \le D$.

The asymptotic behaviors of L_n, H_n and s_n depend on generalized Rényi entropies $r_b(D)$ that we define below. We write $B_D(w_k)$ for a ball of radius D of sequences from $\mathcal{W}_k$, that is, $B_D(w_k) = \{x_1^k : d(x_1^k, w_k) \le D\}$.

Definition: For any $-\infty \le b \le \infty$

$$r_b(D) = \lim_{k\to\infty} \frac{-\log E P^b(B_D(X_1^k))}{bk}$$

where for $b = 0$ we understand $r_0(D) = \lim_{b\to 0} r_b(D)$, that is,

$$r_0(D) = \lim_{k\to\infty} \frac{-E\log P(B_D(X_1^k))}{k} ,$$

provided the above limits exist.

Using the subadditive ergodic theorem, we can prove that the entropies $r_b(D)$ exist in a stationary mixing model. The main result of the paper is summarized below.

Theorem. *In a mixing model with the mixing coefficient tending to zero the following holds:*

$$\lim_{n\to\infty} \frac{L_n}{\log n} = \frac{1}{r_0(D)} \qquad \text{(pr.)}$$

But, L_n does **not** *converge almost surely to any limit and actually the following is true*

$$\liminf_{n\to\infty} \frac{L_n}{\log n} = \frac{1}{r_{-\infty}(D)} \quad , \quad \limsup_{n\to\infty} \frac{L_n}{\log n} \ge \frac{2}{r_1(D)} \qquad \text{(a.s.)}$$

for the Markovian model. In the Bernoulli model, the last inequality can be replaced by equality.

In a related paper Steinberg and Gutman [3] analyzed the so called waiting time, defined as the number N_l such that the beginning substring of length l reoccurs approximately in the string for the first time after N_l symbols. The authors of [3] proved that for a stationary ergodic sequence $\limsup_{l\to\infty} \log N_l/l \le R(D/2)$ (pr.). As a corollary to our main result we show that in the mixing model $\lim_{l\to\infty} \log N_l/l = r_0(D)$ (a.s.), which ultimately settles the problem of [3].

References

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, (1971).

[2] D. Ornstein and P. Shields, Universal Almost Sure Data Compression, *Annals of Probability*, 18, 441-452 (1990).

[3] Y. Steinberg and M. Gutman, An Algorithm for Source Coding Subject to a Fidelity Criterion, Based on String Matching, *IEEE Trans. Information Theory*, 39, 877-886 (1993).

[4] W. Szpankowski, Asymptotic Properties of Data Compression and Suffix Trees, *IEEE Trans. Information Theory*, 39, 1647-1659 (1993).

[5] A. Wyner and J. Ziv, Some Asymptotic Properties of the Entropy of a Stationary Ergodic Data Source with Applications to Data Compression, *IEEE Trans. Information Theory*, 35, 1250-1258 (1989).

[1] On leave from Department of Discrete Mathematics, Adam Mickiewicz University, Poznań, Poland.

[2] Supported by NSF Grants NCR-9206315 and CCR-9201078.

The Gold-Washing Algorithm(II): Optimality for ϕ-Mixing Sources*

Zhen Zhang[1] and En-hui Yang[2]

Abstract — **Two versions of the Gold-Washing data compression algorithm, one with codebook innovation interval and the other with finitely many codebook innovations, are considered. The corresponding optimality results are proved for stationary, ϕ-mixing sources.**

I. Description of Algorithms

In their recent paper [1], Zhang and Wei proposed a universal lossy data compression algorithm called Gold-Washing(GW) algorithm. Let A and $\hat{A}$ be our source alphabet and reproduction alphabet, respectively. Fix $R > 0$ and let $L = \lfloor 2^{nR} \rfloor$. For each $n \geq 1$, the GW algorithm acts like an adaptive vector quantizer when it is applied to encode a source sequence $x = \{x_i\}_1^\infty$ from A. It first parses the source sequence $x = \{x_i\}_1^\infty$ into non-overlapping source words of length n $x^n(t) = (x_{(t-1)n+1}, x_{(t-1)n+2}, \cdots, x_{tn})$, $t = 1, 2, \cdots$, and then uses a codebook $C_n(t-1)$ which changes slowly in time to quantize $x^n(t)$. Each codebook $C_n(t-1)$ consists of an ordered list of $2L$ entries. Each entry in the first half(denoted by $C_n^1(t-1)$) of $C_n(t-1)$ is merely an n-length reproduction sequence called a codeword from $\hat{A}$, whereas each entry in the second half of $C_n(t-1)$ consists of a codeword from $\hat{A}$ and a counter associated with the codeword. When the codebook $C_n(t-1)$ is used to quantize the source word $x^n(t)$, the encoder maps $x^n(t)$ to a smallest index for which the corresponding codeword yields the smallest distortion among $C_n(t-1)$. After $x^n(t)$ is encoded, the codebook $C_n(t-1)$ is innovated and changed to $C_n(t)$. The innovation operation of $C_n(t-1)$ is as follows. (Assume an index i is assigned to $x^n(t)$ by the encoder.)

- S1 If $i > L$, the counter associated with the i-th codeword in $C_n(t-1)$ is incremented by 1.
- S2 If the counter associated with the $(L+1)$-th codeword in $C_n(t-1)$ is $\geq n^\zeta$ prior to the execution of S1, then a randomly selected codeword from $C_n^1(t-1)$ is discarded and, at the same time, the $(L+1)$-th codeword in $C_n(t-1)$ is promoted into $C_n^1(t-1)$; otherwise, the $(L+1)$-th entry in $C_n(t-1)$, including the codeword and counter, is discarded and the first L entries in $C_n(t-1)$ remain unchanged.
- S3 Each entry from the $(L+2)$-th position to the $2L$-th position is moved one step forward.
- S4 Finally, a new randomly selected codeword according to a prescribed distribution occupies the $2L$-th vacant position and its counter is set to 0; the resulting codebook is denoted by $C_n(t)$ and used to quantize $x^n(t+1)$.

In S2, n^ζ is a threshold and ζ is an number > 2. Initially, the codebook $C_n(0)$ is selected arbitrarily and all counters in the second half of $C_n(0)$ are set to zero. Knowing the initial codebook, the new random codeword in S4, and the discarded codeword in S2 when promotion occurs, the decoder performs the codebook innovation operation in the exact same way as the encoder does.

It was proved in [1] that the above mentioned GW algorithm is optimal for memoryless sources. In this paper, our aim is to investigate the asymptotic optimality of the GW algorithm for stationary, ϕ-mixing sources. Accordingly, we shall consider the following two versions of the GW algorithm.

GW algorithm with codebook innovation interval k: This version of the GW algorithm is similar to the original one mentioned above except that this time the encoder innovates its codebook only when $t = (k+1)m$, $m = 1, 2, \cdots$. In other words, the time interval between two consecutive codebook innovations is k; during the time period from $t = (k+1)(m-1)+1$ to $t = (k+1)m$, the codebook is held fixed and used to quantize source words $x^n((k+1)(m-1)+1), \ldots, x^n((k+1)m)$; only after the source word $x^n((k+1)m)$ is encoded, the codebook is innovated according to the codebook innovation operation (S1–S4) and then is held fixed(including the counters in the second part of the codebook) and reused for the next time period of length k.

GW algorithm with finitely many codebook innovations: This version is a variant of the GW algorithm with codebook innovation interval k where after finitely many codebook innovations, the codebook is held fixed and reused to quantize the incoming successive source words.

II. Optimality Results

Let $\rho : A \times \hat{A} \to [0, +\infty)$ be a single-letter distortion measure. Given a stationary, ergodic source μ with random output $\{X_i\}_1^\infty$, let $D(R)$ denote its distortion rate function with respect to the fidelity criterion $\{\rho_n\}$, where $\rho_n(x^n, y^n) = n^{-1} \sum_{i=1}^n \rho(x_i, y_i)$ for $x^n \in A^n$ and $y^n \in \hat{A}^n$. If a stationary, ergodic source μ with random output $X = \{X_i\}_1^\infty$ is encoded by the GW algorithm with codebook innovation interval $k(n)$, the expected distortion per symbol is defined by

$$\rho(n, \mu) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\rho_n(X^n(t), C_n(t-1)) , \qquad (1)$$

where $\rho_n(X^n(t), C_n(t-1))$ is the minimum of $\rho_n(X^n(t), y^n)$ over all $y^n \in C_n(t-1)$ and "E" denotes the expectation with respect to $X^n(t)$ and $C_n(t-1)$. The following is our optimality result concerning the GW algorithm with codebook innovation interval $k(n)$.

Theorem 1 *Let μ be a stationary, ϕ-mixing source having the blowing-up property and whose ϕ-mixing coefficients satisfy $\phi(k(n)n)L^{-1} \to 0$ as $n \to \infty$, then*

$$\rho(n, \mu) \to D(R) \quad as \ n \to \infty.$$

When μ is a strong mixing Markov(or finite-state) source, Theorem 1 can be strengthened to almost sure convergence. Similar results hold for the GW algorithm with finitely many codebook innovations.

References

[1] Z. Zhang and V. K. Wei, "An on-line universal lossy data compression algorithm by continuous codebook refinement," *Submitted to IEEE Trans. Inform. Theory for publication.*

*This work was supported in part by National Sciences Foundation under grant NCR 9205265.

[1]Commun. Science Institute, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565.

[2]Dept. of Math., Nankai University, Tianjin 300071, P.R. China.

Universal Estimation of the Optimal Probability Distributions for Data Compression of Discrete Memoryless Sources with Fidelity Criterion

Hiroki Koga and Suguru Arimoto

Department of Mathematical Engineering and Information Physics,
Faculty of Engineering, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Abstract — **Output probability distributions of the test channels play important roles in data compression of discrete memoryless sources with fidelity criterion. In this paper a universal algorithm for estimating the output probability distributions is proposed. Sample size required by the algorithm is evaluated under a criterion of estimation similar to that of PAC learning in the computational learning theory.**

I. Introduction

Rate-distortion function describes a basic lower bound of compression efficiency asymptotically attainable by a data compression scheme with fidelity criterion. For a discrete memoryless source of finite alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_J\}$ it is defined as a minimum of the mutual information as follows:

$$R(\mathbf{p}, D) = \min_{W \in \mathcal{W}(\mathbf{p}, D)} I(\mathbf{p}; W), \tag{1}$$

where $\mathbf{p} = (p(a_1), p(a_2), \ldots, p(a_J))$ denotes a probability distribution of the source, $\mathcal{W}(\mathbf{p}, D)$ is the set of $J \times J$ stochastic matrices each element of which causes average distortion per symbol within D under a single-letter fidelity criterion $d : \mathcal{A} \times \mathcal{A} \to [0, \infty)$ satisfying $d(a_j, a_k) = 0$ if and only if $j = k$. The rate-distortion function is positive for all $D \in [0, D_{max})$, where $D_{max} \stackrel{\text{def}}{=} \min_k \sum_j p(a_j) d(a_j, a_k)$. Fix $\Delta \in (0, D_{max})$ arbitrarily and denote by W^* the test channel matrix achieving the minimum in (1). The probability distribution on $\mathcal{A}$ defined by $p^*(a_k) = \sum_{j=1}^{J} p(a_j) W^*(a_k|a_j), k = 1, 2, \ldots, J$ means the output probability distribution of the test channel corresponding to the distortion level Δ. In this note a universal estimation algorithm of $\mathbf{p}^*$ is proposed and sample size required by the algorithm is evaluated.

Suppose that another discrete memoryless source with the same alphabet $\mathcal{A}$ as well as the source to be compressed is available to the estimation algorithm. Denote by $\mathbf{q} = (q(a_1), q(a_2), \ldots, q(a_J))$ the probability distribution of another source called *auxiliary source*. Assume that $q(a_j) > 0$ for all $a_j \in \mathcal{A}$ satisfying $p^*(a_j) > 0$. For an arbitrarily fixed n let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$ be L n-tuples drawn independently from the source and $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$ be M n-tuples drawn from the auxiliary source. By using the two sets $\mathcal{X}$ and $\mathcal{Y}$, the algorithm outputs $\hat{\mathbf{p}}^*$ as an estimate of $\mathbf{p}^*$ satisfying

$$Prob\,(D(\mathbf{p}^* || \hat{\mathbf{p}}^*) > \varepsilon) < \delta \tag{2}$$

for any given $\varepsilon > 0$ and $\delta \in (0, 1)$ if n is sufficiently large, where *Prob* means the probability with respect to $\mathcal{X} \times \mathcal{Y}$. The criterion of estimation (2) is deeply related to a data compression scheme with fixed data-base proposed by Steinberg and Gutman [1] and analyzed in detail by Koga and Arimoto [2]. Moreover, imposing the criterion (2) is the first attempt to introduce a viewpoint of the PAC (*Probably Approximately Correct*) learning models proposed by Valiant [3] to data compression with fidelity criterion.

II. Main Results

It is assumed that the estimation algorithm can use an estimate of $\mathbf{p}$, denoted by $\mathbf{p}_e$, satisfying $\|\mathbf{p} - \mathbf{p}_e\|_1 = O(n^{-\beta_e})$ for any fixed $\beta_e \in (0, \frac{1}{2}]$. It estimates $\mathbf{p}^*$ in the following manner:

Algorithm 1 *1) Choose $\alpha > 0$ and $\beta \in (0, \beta_e)$ arbitrarily. Derive $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$ from the source and $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M\}$ from the auxiliary source. Fix an integer m_0 arbitrarily satisfying $1 \le m_0 \le M$.*

2) For all $m = 1, 2, \ldots, M$ define $\mathcal{N}(\mathbf{y}_m, \Delta)$ by

$$\mathcal{N}(\mathbf{y}_m, \Delta) = \{\mathbf{x} \in \mathcal{X} \mid d_n(\mathbf{x}, \mathbf{y}) \le \Delta \text{ and } \|\mathbf{p}_e - \mathbf{t}(\mathbf{x})\|_1 \le n^{-\beta}\}, \tag{3}$$

where d_n denotes distortion between n-tuples defined by d, and $\mathbf{t}(\mathbf{x})$ denotes the type of $\mathbf{x}$. Search for the integer m^* maximizing $|\mathcal{N}(\mathbf{y}_m, \Delta)|$.

3) If $|\mathcal{N}(\mathbf{y}_{m^*}, \Delta)| \ge n^{\alpha}$, output $\mathbf{t}(\mathbf{y}_{m^*})$. Otherwise, output $\mathbf{t}(\mathbf{y}_{m_0})$. □

Under the assumption that $\mathbf{p}^*$ is unique, lower bounds of L and M guaranteeing Algorithm 1 to meet the criterion (2) are established in the following theorem.

Theorem 1 *Let $R_{\mathcal{X}} = \frac{1}{n} \log_2 L$ and $R_{\mathcal{Y}} = \frac{1}{n} \log_2 M$. Then for any fixed $\Delta \in (0, D_{max})$, if the two inequalities*

$$\min_{\mathbf{q}' : D(\mathbf{p}^* || \mathbf{q}') \le \varepsilon} \min_{V \in \mathcal{V}(\mathbf{p}, \mathbf{q}', \Delta)} I(\mathbf{q}'; V) < R_{\mathcal{X}} < R(\mathbf{p}, \Delta), \tag{4}$$

$$R_{\mathcal{Y}} > D(\mathbf{q}_\varepsilon || \mathbf{q}) \tag{5}$$

are satisfied then there exists an integer n_0 satisfying that Algorithm 1 outputs $\hat{\mathbf{p}}^*$ meeting the criterion (2) for all $n > n_0$, where $I(\mathbf{q}'; V)$ denotes the mutual information, $\mathcal{V}(\mathbf{p}, \mathbf{q}', \Delta)$ denotes the set of $J \times J$ stochastic matrices satisfying $\sum_{k=1}^{J} q'(a_k) V(a_j|a_k) = p(a_j)$ for all $j = 1, 2, \ldots, J$ and $\sum_{j=1}^{J} \sum_{k=1}^{J} q'(a_k) V(a_j|a_k) d(a_j, a_k) \le \Delta$, and $\mathbf{q}_\varepsilon$ is a probability distribution on $\mathcal{A}$ that achieves the minimum in (4) with a stochastic matrix $V \in \mathcal{V}(\mathbf{p}, \mathbf{q}_\varepsilon, \Delta)$. □

References

[1] Y. Steinberg and M. Gutman, "An algorithm for source coding subject to a fidelity criterion, based on string matching," *IEEE Trans. on Inform. Theory*, vol. IT-39, No. 3, pp. 877–886, 1993.

[2] H. Koga and S. Arimoto, " Asymptotic properties of algorithms of data compression with fidelity criterion based on string matching," *Proc. of 1994 IEEE Int. Symp. Inform. Theory*, Tronheim, Norway, p. 264, 1994.

[3] L. G. Valiant, "A theory of the learnable," *Communication of the ACM*, vol. 27, pp. 1134–1142, 1984.

A Universal Data-Base for Data Compression

Jun Muramatsu† and Fumio Kanaya‡

†NTT Optical Network Systems Laboratories, Yokosuka-shi, 238-03 Japan.
‡Shonan Institute of Technology, Fujisawa-shi, 251 Japan.

Abstract — **A data-base for data compression is universal if in its construction no prior knowledge of the source distribution is assumed and is optimal if, when we encode the reference index of the data-base, its encoding rate achieves the optimal encoding rate for any given source: in the noiseless case the entropy rate and in the semifaithful case the rate-distortion function of the source. We construct a universal data-base for all stationary ergodic sources, and prove the optimality of the thus constructed data-base for a block-shift type reference and a single-shift type referrence.**

I. Introduction

We consider the case where both a sender and a receiver have the same data-base on their respective sides. In this case, we can transmit a source output in the following way: the sender refers to the data-base for the data string which matches the given source output and then encodes the reference index of the data string to send it out. The receiver then decodes the encoded index to retrieve the data string from the data-base and then uses it to represent the source output. There are two typical conceivable methods of referring to the data-base: one is a block-shift type reference and the other is a single-shift type reference. Either method can achieve data compression if the number of bits needed to encode the reference index relative to the data-base is smaller than that needed to represent the source output itself. Hereafter we refer to the number of bits divided by the sequence length of the source output as the encoding rate.

We construct an optimal universal data-base for ergodic sources. The construction of our data-base sequence relies entirely on the basic concept of the complexity function (cf. [1]): it is constructed by ordering data strings according to the increasing complexity. The obtained data-base sequence can be applied for both the block-shift type and the single-shift type reference cases.

It should be noted that this data-base can be proved optimal also for the fixed-rate universal code with distortion (cf. [3]).

II. Block-Shift Type Reference Case

Let $\hat{\mathcal{A}}$ be a finite set and let L be a complexity function in the almost sure sense which is defined in [1].

Definition 1 *Let elements of set $\hat{\mathcal{A}}^n$ be ordered according to the increasing complexity (ties may be broken in an arbitrary order). The mapping which maps an element of $\hat{\mathcal{A}}^n$ into its order is called an index function induced by L and is denoted by $\mathcal{L}_{L,n}$. A data-base sequence corresponding to the index function $\mathcal{L}_{L,n}$ is defined by*

$$\hat{u}^{n|\hat{\mathcal{A}}^n|} \equiv \mathcal{L}_{L,n}^{-1}(1) * \mathcal{L}_{L,n}^{-1}(2) * \cdots * \mathcal{L}_{L,n}^{-1}(|\hat{\mathcal{A}}^n|),$$

where notation $*$ is used to denote concatenation of strings.

Next, let $\mathcal{A}$ be a standard space and let ρ is a distortion function which satisfies some conditions stated in [1].

Definition 2 *A D-semifaithful index function $\mathcal{L}_{L,D,n}$ is defined by*

$$\begin{aligned}\mathcal{L}_{L,D,n}(x^n) &\equiv \min_{\hat{x}^n \in \hat{\mathcal{A}}_D^n(x^n)} \mathcal{L}_{L,n}(\hat{x}^n) \\ &= \min\{l;\ \hat{u}_{n(l-1)+1}^{nl} \in \hat{\mathcal{A}}_D^n(x^n)\}, \quad x^n \in \mathcal{A}^n,\end{aligned}$$

where $\hat{u}_i^j \equiv (\hat{u}_i, \ldots, \hat{u}_j)$ and

$$\hat{\mathcal{A}}_D^n(x^n) \equiv \left\{ \hat{x}^n \in \hat{\mathcal{A}}^n;\ \frac{1}{n}\sum_{i=1}^{n} \rho(x_i, \hat{x}_i) \le D \right\}.$$

Theorem 1 *For any $\hat{\mathcal{A}}$-valued stationary ergodic source $\hat{X}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \mathcal{L}_{L,n}(\hat{x}^n) = H_{\hat{X}} \quad \mu_{\hat{X}}\text{-a.s.},$$

and for any $\mathcal{A}$-valued stationary ergodic source X and $D \ge D_0$,

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \mathcal{L}_{L,D,n}(x^n) = R_X(D), \quad \mu_X\text{-a.s.},$$

where $H_{\hat{X}}$ and $R_X(D)$ is the entropy-rate of the source $\hat{X}$ and the rate-distortion fucntion of the source X, respectively.

III. Single-Shift Type Reference Case

We now consider the case when a data-base sequence is reffered to by the single-shift type reference.

Definition 3 *We define a function $\mathcal{S}_{L,n}$ be ginen by*

$$\mathcal{S}_{L,n}(\hat{x}^n) \equiv \min\{s;\ \hat{x}^n = \hat{u}_s^{s+n-1}\}, \quad \hat{x}^n \in \hat{\mathcal{A}}^n$$

and refer to it as the index function for the single-shift type reference. And we define a finction $\mathcal{S}_{L,D,n}$ be given by

$$\begin{aligned}\mathcal{S}_{L,D,n}(x^n) &\equiv \min_{\hat{x}^n \in \hat{\mathcal{A}}_D^n(x^n)} \mathcal{S}_{L,n}(\hat{x}^n) \\ &= \min\{s;\ \hat{u}_s^{s+n-1} \in \hat{\mathcal{A}}_D^n(x^n)\}, \quad x^n \in \mathcal{A}^n\end{aligned}$$

and refer to it as the D-semifaithful index function for the single-shift type reference.

Theorem 2 *For any $\hat{\mathcal{A}}$-valued stationary ergodic source $\hat{X}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \mathcal{S}_{L,n}(\hat{x}^n) = H_{\hat{X}} \quad \mu_{\hat{X}}\text{-a.s.},$$

and for any $\mathcal{A}$-valued stationary ergodic source X and $D \ge D_0$,

$$\lim_{n\to\infty} \frac{1}{n} \log_2 \mathcal{S}_{L,D,n}(x^n) = R_X(D), \quad \mu_X\text{-a.s.}$$

References

[1] J. Muramatsu and F. Kanaya, "Distortion-complexity and rate-distortion function," *IEICE Trans. Fundamentals*, vol.E77-A, no.8, pp.1224–1229, 1994.

[2] J. Muramatsu and F. Kanaya, "A universal data-base for data compression," to appear in *IEICE Trans. Fundamentals.*

[3] J. Muramatsu and F. Kanaya, "The dual quantity of the distortion-complexity and a universal data-base for fixed-rate data compression with distortion," submitted to *IEICE Trans. Fundamentals.*

UNIVERSAL COMPRESSION ALGORITHMS BASED ON APPROXIMATE STRING MATCHING

Ilan Sadeh
Math. Dept., Ben Gurion University, Beer Sheva, Israel, sade@indigo.bgu.ac.il

Two practical universal source coding schemes based on approximate string matching are proposed. One is an approximate fixed-length string matching data compression, and the other is an LZ-type quasi parsing method by approximate string matching. It is shown that in the former algorithm the compression rate converges to the theoretical bound of $R(D)$ for ergodic and stationary processes as the average string length tends to infinity. A similar result holds for the latter algorithm in the limit of the infinite database produced by the former algorithm. The main advantages of the proposed methods are the asymptotic behavior of the encoder implementation and the simplicity of the decoder. Practical results of image and voice compression will be presented.

Definition 1. *We look at the positive time at the sequence $u_0, u_1 \ldots$. Let L be the first index such that the string $u_0 \ldots u_{L-1}$ is not a substring of the data-base u_{-n}^{-1}. That L is equal to $L_n(u)$.*

Definition 2. *The random variable $N_l(\bar{u})$ for $l > 0$ is the smallest integer $N \geq l$ such that $u_0^{l-1} = u_{-N}^{l-1-N}$.*

Given alphabets U and V, a distortion measure is any function $d : |U \times V| \rightarrow \mathcal{R}^+$. Let $\rho_l(\bar{u}; \bar{v})$ denote the distortion for a block- the average of the per letter distortions for the letters that comprise the block, $\rho_l(\bar{u}; \bar{v}) = \frac{1}{l} \sum_{k=1}^{l} d(\bar{u}_k; \bar{v}_k)$.

Definition 3. *For each sample sequence $\bar{u}$ of length l, taken from the sequence u, we define a set $D - Ball$,* $D - Ball(\bar{u}) = \left\{ \bar{v} | \rho(\bar{u}, \bar{v}) \leq D \right\}$.

Definition 4. *For each sample sequence $\bar{u}$ we define the random variable* $DL_n(\bar{u}, v_{-n}^{-1}) = \max_{\bar{v}:\rho(\bar{u},\bar{v})\leq D} L_n(\bar{v}, v_{-n}^{-1})$.

Definition 5. *For each sample sequence $\bar{u}$ we define the random variable* $DN_l(\bar{u}, v_{-k}^{-1}) = \min_{\bar{v}:\rho(\bar{u},\bar{v})\leq D} N_l(\bar{v}, v_{-k}^{-1})$.

Data Compression Scheme A.

1. Verify the readiness of the decoder.
2. Take a string $\bar{u} = u_0^{l-1}$ of length l.
3. If u_0^{l-1} can be approximately matched up to tolerance D by a substring of v_{-n}^{-1}, encode it by specifying $DN_l(\bar{u}, v_{-n}^{-1})$ to the decoder. Add a bit as a header flag to indicate that there is a match. Append string $v_{-DN_l}^{l-1-DN_l}$ to database in decoder and encoder at position 0.
4. If not, indicate that there is no match and transmit to the decoder and append to the database in the encoder and decoder, the string v_0^{l-1}, which is the best D-Ball center, obtained by blockcoding on the current u_0^{l-1} string and is based on the accumulated empirical distribution in the past of u. Blockcoding algorithms are known in literature. The codeword is transmitted as is, without compression.
5. Shift the indices by l to the appropriate values. Update n to $n + l$. Repeat the process from step 1, with a new string of length l and a database v_{-n}^{-1}.

Limit Theorem A. *Given is a D-semifaithful database $v_{-\infty}^{-1}$ generated by Scheme A from a stationary ergodic process u. We assume that the system preserves ergodicity and stationarity. For all $\beta > 0$,* $\lim_{l\to\infty} \Pr\left\{ \left| \frac{\log DN_l(\bar{u}, v_{-\infty}^{-1})}{l} - R(D) \right| > \beta \right\} = 0$. *The average compression ratio attains the bound $R(D)$.*

Scheme B.

1. l=1.
2. Take the string of length l u_0^{l-1}.
3. If u_0^{l-1} can be approximately matched up to tolerance D by a substring of v_{-n}^{-1}, store a pointer N to that substring and increment l. Go to step 2.
4. If not, append to the data base track the string v_{-N}^{l-2-N} at position zero and further, and append the letter v_{l-1} - the reproducing letter which satisfies $d(u_{l-1}, v_{l-1}) = 0$. The encoding is done by the pointer to the string v_{-N}^{l-2-N}, the length $DL_n(u)$ and the last reproducing letter associated to the last source letter.
5. Repeat the process from step 1, where the database is appended with the chosen string denoted by $v_0^{DL_n}$.

Limit Theorem B. *Given is a suffix v_{-n}^{-1} taken from an infinite database generated by an encoder - decoder pair as described in Scheme A. At time zero switch to Scheme B. As the memory size - n tends to infinity, for the new sample sequence $\bar{u}$ encoded from the stationary ergodic input u by Scheme B, in probability,* $\lim_{n\to\infty} \left\{ \frac{\log n}{DL_n(\bar{u}, v_{-n}^{-1})} \right\} = R(D)$.

Certain Exponential Sums over Galois Rings and Related Constructions of Families of Sequences

Jyrki Lahtonen
Dept. of Math., Univ. of Turku, FIN-20500 Turku, Finland

Abstract — **Upper bounds for certain exponential sums over Galois rings are presented. The bound may be regarded as the Galois ring analogue of the so called Kloosterman sums and related exponential sums with a Laurent polynomial argument. An application of the bounds to the design of large families of polyphase sequences with good correlation properties is also given.**

I. Introduction

Let $\psi : GR(q,m) \to \mathbf{C}^*$ be an additive character of the characteristic $q = p^e$, (p prime) Galois ring of q^m elements. Let $\mathcal{T}$ denote the subset of $GR(q,m)$ consisting of the zero element and the powers of an element α of multiplicative order $p^m - 1$.

In [1] Kumar, Helleseth and Calderbank studied the exponential sums of the type

$$\sum_{x \in \mathcal{T}} \psi(f(x)),$$

where $f(x)$ is a polynomial with coefficients in the ring $GR(q,m)$. They apply the theory of the function fields of algebraic curves and their characters. The same technique will allow us to study such sums, where in place of the polynomial $f(x)$ we have a Laurent polynomial, i.e. we allow negative powers of x as well. Observe that this makes sense in $\mathcal{T}^* = \mathcal{T} \setminus \{0\}$ as all the elements in $\mathcal{T}^*$ are units of the ring. Our technique differs from the approach in [1] in the sense that we have utilized a Witt vector presentation of the Galois rings: For example we view the ring $GR(4,m)$ as the ring of Witt vectors $W_2(\mathbf{F})$ of length two over the field $\mathbf{F} = GF(2,m)$. The elements of $W_2(\mathbf{F})$ are ordered pairs $(\alpha_0, \alpha_1), \alpha_i \in \mathbf{F}$ and the ring operations of two such pairs are defined as

$$\begin{aligned}(\alpha_0,\alpha_1) \quad + \quad (\beta_0,\beta_1) \quad &= \quad (\alpha_0+\beta_0, \alpha_1+\beta_1+\alpha_0\beta_0),\\ (\alpha_0,\alpha_1) \quad * \quad (\beta_0,\beta_1) \quad &= \quad (\alpha_0\beta_0, \alpha_1\beta_0^2+\beta_1\alpha_0^2),\end{aligned}$$

where the arithmetical operations between the individual components are the usual *field* operations. Our set $\mathcal{T}$ consists then of the pairs $(\beta, 0), \beta \in \mathbf{F}$. Similarly the rings $GR(8,m)$ can be viewed as rings of Witt vectors of length three. For a description of the arithmetic of Witt vectors of arbitrary length and characteristic we refer the interested reader to Jacobson [3, section 8.10].

II. Results

We have proven the following results:

Theorem 1 *Let $q = 4$ and $\alpha, \beta \in GR(4,m)$ be arbitrary excluding the case $\alpha = \beta = 0$. Then*

$$\left| \sum_{x \in \mathcal{T}^*} \psi\left(\alpha x + \frac{\beta}{x}\right) \right| \le 4\sqrt{2^m}.$$

Theorem 2 *Let $q = 8$ and $\alpha_1, \beta_1 \in GR(8,m)$ and $\alpha_3, \beta_3 \in 4GR(8,m)$ be such that at least one of them differs from zero. Then*

$$\left| \sum_{x \in \mathcal{T}^*} \psi\left(\alpha_1 x + \frac{\beta_1}{x} + \alpha_3 x^3 + \frac{\beta_3}{x^3}\right) \right| \le 8\sqrt{2^m}.$$

As is the case with the usual Kloosterman sums, we have the additional result that the associated hybrid sums

$$\sum_{x \in \mathcal{T}^*} \psi(f(x))\,\chi(x)$$

have exactly the same bounds. Here $f(x)$ is any of the Laurent polynomial appearing in the above results and $\chi\left(\alpha^j\right) = \omega^{kj}$ is a multiplicative character of the group $\mathcal{T}^*, \omega = e^{2\pi i/(p^m-1)}$. Such hybrid sums can be used either to analyze the aperiodic correlation properties of the resulting family of sequences or to get an even larger family with a very large alphabet. After submitting this note I have learned that Helleseth, Kumar and Shanbhag have obtained more general versions of the above theorems [2]. However, they didn't consider the associated hybrid sums.

III. Applications to Sequence Design

The above character sums appear naturally as correlation values of certain families of sequences. To arrive at the families all one has to do is to select representatives of cyclically distinct classes of associated codewords of period $L = |\mathcal{T}^*| = 2^m - 1$. Our character sums yield families with the following parameters for all $m \ge 1$:

- Quaternary family of size L^3 and maximal correlation $1 + 4\sqrt{L+1}$,
- Eight-phase family of size L^7 and maximal correlation $1 + 8\sqrt{L+1}$,
- Polyphase family of size L^4 and maximal correlation $1 + 4\sqrt{L+1}$ and
- Polyphase family of size L^8 and maximal correlation $1 + 8\sqrt{L+1}$.

Here the 'polyphase' families have alphabets of sizes $4L$ and $8L$ respectively effectively filling in the unit circle of the complex plane.

References

[1] P.V. Kumar, T. Helleseth, A. Calderbank, "An Upper Bound for Weil Exponential Sums over Galois Rings and Applications", *IEEE Trans. Inf. Theory*, vol. 41, pp. 456–468.

[2] T. Helleseth, P.V. Kumar, A.G. Shanbhag, "An Upper Bound for Kloosterman Sums over Galois Rings", preprint.

[3] N. Jacobson, *Basic Algebra II*, San Francisco: Freeman, 1980.

Generalization of No Sequences

Jong-Seon No
Dept. of Electronic Eng., Kon-kuk University
93-1 MoJin-Dong, KwangJin-Gu, Seoul, 133-701, Korea

Abstract-In this paper, GMW sequences and families of No sequences are generalized. Generalized GMW sequences have ideal autocorrelation properties and balance properties and generalized No sequences also have optimal correlation properties in terms of Welch's lower bound. The linear spans of the generalized GMW sequences and generalized No sequences appear to be large although we do not at present have a closed-form expression for the linear span. A count of the numbers of cyclically distinct generalized GMW sequences and generalized No sequences that can be constructed is provided.

I. INTRODUCTION

In this paper, the generalization of GMW sequences and No sequences is introduced. In Section II, GMW sequences are generalized, those ideal full-period autocorrelation properties are derived, and a count of the number of cyclically distinct generalized GMW sequences that can be constructed is provided. It is also shown how the families of No sequences can be generalized in an identical fashion and optimal correlation properties are described in Section III. Here, the number of distinct families of generalized No sequences of given period is described, too.

II. GENERALIZATION OF GMW SEQUENCES

We can define generalized GMW sequences as follows:

Definition 1 : Let n and m_i, $i = 1, 2, ..., d$, be integers satisfying

$$m_d | n \text{ and } m_i | m_{i+1}, \text{ for } 1 \le i \le d-1. \quad (1)$$

A *generalized GMW sequence* is then defined as the multiple trace function sequence of period N given by

$$s_g(t) = tr_1^{m_1}\{[tr_{m_1}^{m_2}\{[tr_{m_2}^{m_3}\{\cdots\{[tr_{m_d}^{n}(\alpha^t)]^{r_d}\}\cdots\}]^{r_2}\}]^{r_1}\}, \quad (2)$$

where α is an element of order $N = 2^n - 1$ and for $1 \le i \le d$,

$$gcd(r_i, 2^{m_i} - 1) = 1, \ 1 \le r_i < 2^{m_i} - 1. \quad (3)$$

The generalized GMW sequence has the ideal full-period autocorrelation values and it can be counted as follows:

Theorem 1 : The number of cyclically different generalized GMW sequences of given period N is given by:

$$\mathrm{N_{gGMW}} = \frac{\phi(2^n - 1)}{n} \cdot \prod_{i=1}^{d} \frac{\phi(2^{m_i} - 1)}{m_i}, \quad (4)$$

where $\phi(\cdot)$ is Euler's *phi* function.

III. GENERALIZATION OF NO SEQUENCES

The definition of a generalized No sequence family is given as follows:

Definition 2 : Let n and m_i, $i = 1, 2, ..., d$, be integers satisfying

$$n = 2 \cdot m_d \text{ and } m_i | m_{i+1}, \text{ for } 1 \le i \le d-1. \quad (5)$$

A family of *generalized No sequences*

$$S_g = \{s_i(t) \mid 0 \le t \le N-1, \ 1 \le i \le 2^m\} \quad (6)$$

is a set of multiple trace function sequences defined as

$$s_i(t) = tr_1^{m_1}\{[tr_{m_1}^{m_2}\{[tr_{m_2}^{m_3}\{\cdots\{[tr_{m_d}^{n}(\alpha^{2t}) + \gamma_i \cdot \alpha^{T \cdot t}]^{r_d}\}\cdots\}]^{r_2}\}]^{r_1}\}, \quad (7)$$

where $N = 2^n - 1$, γ_i is in $GF(2^{m_d})$, $T = 2^{m_d} + 1$, and for $1 \le i \le d$,

$$gcd(r_i, 2^{m_i} - 1) = 1, \ 1 \le r_i < 2^{m_i} - 1. \quad (8)$$

The full-period correlation function of No sequences are the same as that of Kasami sequences. Counts for the number of cyclically distinct generalized GMW sequences and generalized No sequence families are the same.

REFERENCES

1. J. S. No, "A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span," Ph.D. dissertation, University of Southern California, Los Angeles, CA, USA, May, 1988.
2. J. S. No and P. V. Kumar, "A new family of binary pseudorandom sequences having optimal periodic correlation properties and large linear span," *IEEE Trans. Inform. Theory*, vol IT-35, no. 2, pp. 371-379, Mar. 1989.
3. A. Klapper, "d-form sequences: Families of sequences with low correlation values and large linear spans," *IEEE Trans. Inform. Theory*, vol. 41, pp. 423-431, Mar. 1995.

Codes for Optical Transmission at Different Rates

O. Moreno[1] and Svetislav V. Marić

Department of Mathematics and Computer Science, University of Puerto Rico, Rio Piedras, 00931.
Department of Electrical Engineering, City College of the City University of New York, New York, NY 10031.

Abstract — **Constructions for families of cyclic constant weight codes are presented to be used in fiber optic CDMA networks for multirate transmission. It is shown that the discussed code families satisfy the requirements for successful transmission of different data rates using the CDMA technique.**

I. Introduction

An (n, ω, λ)-optical orthogonal code (OOC) (see [1], [2],[3]) 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:

$$\sum_{k=0}^{n-1} x(k)x(k \oplus_n \tau) \leq \lambda \tag{1}$$

for all sequences $x(.) \in C$ and all integers $\tau \neq 0\,(mod\, n)$ and

$$\sum_{k=0}^{n-1} x(k)y(k \oplus_n \tau) \leq \lambda \tag{2}$$

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

$$\Phi(n, \omega, \lambda) \leq \lfloor \frac{A(n, 2\omega - 2\lambda, \omega)}{n} \rfloor \leq \frac{(n-1)(n-2)...(n-\lambda)}{\omega(\omega-1)...(\omega-\lambda)}. \tag{3}$$

II. Constructions

Codes with these properties have been called optical orthogonal codes in papers [1]-[4] in connection with applications for optical channels and cyclically permutable constant weight codes (see [5] and references there) in connection of constructing of protocol sequences for the multiuser collision channel without feedback.

In a multimedia environment different types of users transmit at different data rates [6]. As a most obvious example in Personal Communication Networks we have low rate-voice transmissions and high rate data-transmissions.

Note that in a multirate case a CPCW with a longer length corresponds to lower data bit rates and the smaller length CPCW corresponds to higher data bit rates. Hence for multimedia applications we need CPCW families with different lengths and weights. The code construction is complicated by the fact that now we need to establish not only correlation properties (value of λ) of one CPCW family but also cross-correlation properties of families of CPCW with different n and ω.

In [3], 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 codewords within each family 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 form 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.

In this talk, we will show that Construction $\mathcal{A}$ can be used to construct a nested chain of asymptotically optimum OOC's of lengths $n_0 = n$, $n_i | n_0$, $i \geq 1$. Using on-off keying as the method of data modulation, we show how this nested chain can be used to efficiently allow several users with *different information rates* to simultaneously transmit information. Decoding of the desired information stream is easily accomplished using correlation detection.

Such codes are relevant to multimedia communications.

[1] This research is supported in part by the National Science Foundation under Grant numbers RII-9014056, NCR-890505, and the Computational Mathematics Group of the EPSCoR of Puerto Rico Grant.

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] O. Moreno, Z. Zhang, P. V. Kumar and V. A. Zinoviev, "New Constructions of Optimal Cyclically Permutable Constant Weight Codes ", to appear in the *IEEE Transactions on Information Theory*, Vol. 41, No. 3, March 1995.

[4] 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.

[5] Nguyen Q. A, Lásló Gyorfi and James L. Massey, "Constructions of Binary Constant-Weight Cyclic Codes and Cyclically Permutable Codes", *IEEE Trans. Inform. Theory*, vol.38, No.3, pp.940-949, May. 1992.

[6] W. Verbiest, G. Van der Plas and J. G. Mestdagh, "FITL and B-ISDN: A Marriage with a Future", *IEEE Communications Magazine*, vol. 31, no.6, pp.60-66, June, 1993.

An Upper Bound for Extended Kloosterman Sums over Galois Rings

Abhijit G. Shanbhag
EEB 522, EE-Systems
Univ. South. Calif.
Los Angeles
CA 90089-2565

P. Vijay Kumar
EEB 534, EE-Systems
Univ. South. Calif.
Los Angeles
CA 90089-2565

Tor Helleseth
Dep. of Informatics
Univ. of Bergen
N-5020, Bergen
Norway

Abstract — **An upper bound for the extended Kloosterman sum over Galois rings is derived. This bound is then used to construct new, efficient sequence families with prime-power phase.**

I. Introduction

For a fixed prime p and integers $e, m,\ \ e \geq 2,\ m \geq 1$, let $R_{e,m}$ denote the *Galois Ring* of characteristic p^e and containing p^{em} elements. Let $\psi_{e,m}$ be a non-trivial additive character of $R_{e,m}$ and let $f(x)$ be a *non-degenerate* polynomial (i.e. no term in $f(x)$ has degree which is a multiple of p) over $R_{e,m}$ with *weighted degree* [1] D_f. Define $\mathcal{T}_{e,m} = \mathcal{T}^*_{e,m} \cup 0$ where $\mathcal{T}^*_{e,m}$ is a cyclic subgroup (the *Teichmuller* group) of order $p^m - 1$ of $R^*_{e,m}$. In [1], Kumar et al. prove

$$\left| \sum_{x \in \mathcal{T}_{e,m}} \psi_{e,m}(f(x)) \right| \leq (D_f - 1)\sqrt{p^m}. \tag{1}$$

This bound leads to new sequence families which compare very well with existing sequence families when maximum non-trivial correlation, alphabet size and family size are used as a basis for comparison. More precisely, let $\mathcal{F}_D$ denotes a maximal family of pairwise, cyclically distinct sequences each having period $p^m - 1$ from the set

$$\mathcal{S}_D = \left\{ \{T_{e,m}(f(\beta^t))\}_{t \in Z} \mid D_f \leq D \right\}$$

where β is a generator of $\mathcal{T}^*_{e,m}$ and $f(x) \in R_{e,m}[x]$ has weighted degree D_f. We then have the following bounds for the maximum non-trivial correlation C_{max} and the size of the family $\mathcal{F}_D$ as under

$$C_{max} \leq 1 + (D-1)\sqrt{p^m}$$

and

$$|\mathcal{F}_D| \geq p^{m(D - \lfloor \frac{D}{p^e} \rfloor - 1)}.$$

In this paper we obtain an upper bound for the *extended Kloosterman* sums, i.e. sums of the form

$$K_{e,m}(f_1, f_2) = \sum_{x \in \mathcal{T}^*_{e,m}} \psi_{e,m}(f_1(x) + f_2(x^{-1})),$$

where $f_1(x), f_2(x)$ are polynomials over $R_{e,m}$. These sums lead to new sequence designs for CDMA applications.

II. Bound on the extended Kloosterman sum

Let $f_1(x), f_2(x) \in R_{e,m}[x]$ have weighted degrees D_{f_1} and D_{f_2} respectively. Let $\psi_{e,m}$ be any non-trivial additive character of $R_{e,m}$. Using L-function techniques, we can express the exponential sum $\sum_{x \in \mathcal{T}^*_{e,m}} \psi_{e,m}(f_1(x) + f_2(x^{-1}))$ as a sum of $D_{f_1} + D_{f_2}$ complex numbers. These complex numbers can be shown to be the reciprocal roots of the zeta function of a function field over F_{p^m}. It follows from the *Riemann Hypothesis* for function fields, that the magnitude of each of these complex numbers is $\sqrt{p^m}$. Thus, using notation as above, we obtain the following theorem:

Theorem 1

$$\left| \sum_{x \in \mathcal{T}^*_{e,m}} \psi_{e,m}(f_1(x) + f_2(x^{-1})) \right| \leq (D_{f_1} + D_{f_2})\sqrt{p^m}.$$

III. Applications to sequence designs

We now restrict ourselves to the case when $p = 2$. Consider the set $\mathcal{S}_{D_1,D_2}$ of sequences defined via

$$\mathcal{S}_{D_1,D_2} = \left\{ \{T_{e,m}(f_1(\beta^t) + f_2(\beta^{-t}))\} \mid D_{f_1} \leq D_1,\ D_{f_2} \leq D_2 \right\}$$

where β is a generator of $\mathcal{T}^*_{e,m}$ and $f_i(x) \in R_{e,m}[x],\ i = 1, 2$ is non-degenerate with weighted degree D_{f_i}. Let the set

$$\mathcal{F}_{D_1,D_2} \subset \mathcal{S}_{D_1,D_2}$$

consist of a maximal family of pairwise, cyclically distinct sequences in $\mathcal{S}_{D_1,D_2}$ with each sequence having period $2^m - 1$. Using Theorem 1, it is easy to see that the maximum non-trivial correlation C_{max} of the family $\mathcal{F}_{D_1,D_2}$ is upper bounded via

$$C_{max} \leq 1 + (D_1 + D_2)\sqrt{2^m}. \tag{2}$$

The size of the family $\mathcal{F}_{D_1,D_2}$ can be lower bounded using the formula below:

$$|\mathcal{F}_{D_1,D_2}| \geq 2^{m(D_1 + D_2 - \lfloor \frac{D_1}{2^e} \rfloor - \lfloor \frac{D_2}{2^e} \rfloor - 1)}. \tag{3}$$

Note that

$$\lfloor \frac{D_1 + D_2 + 1}{2^e} \rfloor \leq \lfloor \frac{D_1}{2^e} \rfloor + \lfloor \frac{D_2}{2^e} \rfloor + 1.$$

In case of equality, we note that the corresponding bounds for the maximum non-trivial correlation and family size of $\mathcal{F}_{D_1+D_2+1}$ and $\mathcal{F}_{D_1,D_2}$ are equal.

References

[1] P.V. Kumar, T. Helleseth, and A.R. Calderbank, "An Upper Bound for Weil Exponential Sums over Galois Rings and Applications", *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 456–468, 1995.

[0]The work was supported in part by the National Science Foundation under Grant Number NCR-93-05017 and in part by the Norwegian Research Council for Science and the Humanities.

Optimization of the Ambiguity Function of Binary Sequences and their Mismatched Filters for Use in CTDMA Systems

Hans Dieter Schotten* and Jürg Ruprecht**

*Institut für Elektrische Nachrichtentechnik, RWTH Aachen, Melatener Strasse 23, D-52056 Aachen, Germany
Phone +49 241 80 7680, Fax +49 241 8888 196, EMail schotten@ient.rwth-aachen.de

**Swiss Telecom PTT, R&D, Mobile Communications / FE 422, CH-3000 Berne 29, Switzerland
Phone +41 31 338 5492, Fax +41 31 338 5174, EMail ruprecht@vptt.ch

Abstract — **For the application in a cellular common-code spread-spectrum multiple access system, here referred to as CTDMA, the ambiguity function of the binary spreading sequence and its mismatched despreading filter is optimized.**

I. Introduction

In cellular *Code Time Division Multiple Access (CTDMA)* systems [1], the user signals are first separated by a symbol-level TDMA scheme and then spread in a DS-CDMA fashion by a common (cell-specific) binary spreading sequence $s[\cdot]$ of length L with $s[n] \in \{-1, +1\}$ for $0 \le n < L$ and zero otherwise. At the receiver, the incoming signal is passed through the aperiodic inverse filter $v[\cdot]$ of $s[\cdot]$, which completely separates the users of the same cell and thus omits this kind of interference that usually appears in cellular CDMA. The filter $v[\cdot]$, two-sided infinite in length, is well approximated by a filter $w[\cdot]$ of length $N \approx 3L$, which still achieves a sufficient user separation by minimizing the aperiodic correlation sidelobes $C_{sw}[m] = \sum_n s[n]\, w[n+m]$, $m \neq 0$. Different techniques to design $w[n]$ have been discussed in the literature: *Truncation* of $v[\cdot]$ is conceptually simple [2], *linear programming (LP)* optimizes the peak/off-peak (POP) ratio $\rho_{sw} = C_{sw}[0] / \max_{m \neq 0} |C_{sw}[m]|$, and the *least-square (LS) algorithm* minimizes the sidelobe energy of $C_{sw}[\cdot]$.

II. System Analysis

These filter design techniques neglect possible Doppler frequency shifts that occur in cellular applications due to velocity differences Δv. The corresponding effect at the receiver output is described by the ambiguity function

$$A_{sw}[m, \xi] = \sum_n \mathrm{e}^{-j2\pi\xi n}\, s[n]\, w[n+m]$$

with $\xi = T_c f_d$. Here, T_c is the chip duration, $f_d = 2\Delta v f_0 / c$ the Doppler shift, f_0 the carrier frequency and $c \approx 3 \cdot 10^8 \frac{m}{s}$ the speed of light. With $f_0 \approx 2\,\mathrm{GHz}$, $T_c \approx 1\,\mu s$ and $\Delta v_{max} = 30 \ldots 300\,\frac{m}{s}$, maximum values of $\xi_{max} \approx 5 \cdot 10^{-4} \ldots 5 \cdot 10^{-3}$ are obtained. In order to investigate the degradation due to these Doppler shifts (for other results, especially for larger Doppler shifts, cf. [3, 4]), we have computed the generalized POP-ratio

$$\rho_{sw}(\xi_{max}) = \frac{\min_{|\xi| \le \xi_{max}} A_{sw}[0, \xi]}{\max_{m \neq 0, |\xi| \le \xi_{max}} |A_{sw}[m, \xi]|},$$

where the filters $w[\cdot]$ of length $N = 3L$ have been determined using the LP technique. For $\xi_{max} = 10^{-4}$, the Doppler effect causes a noticeable degradation of $\rho_{sw}(\xi_{max})$, and for $\xi_{max} = 5 \cdot 10^{-3}$, the loss in $\rho_{sw}(\xi_{max})$ can exceed 20 dB as shown in the table below that lists the $\rho_{sw}(\xi_{max})$-values. Especially sequences with best noise performance [5] (cf. #1,#2,#3), which also provide very good $\rho_{sw}(0)$-values, seem to be Doppler sensitive. Others (cf. #4) are less sensitive.

#	L	$s[\cdot]$ (hex)	$v[\cdot]$	pedestrian $\Delta v_{max}=0$ $\xi_{max}=0$	car $\approx 30\frac{m}{s}$ $=5 \cdot 10^{-4}$	train $\approx 60\frac{m}{s}$ $=1 \cdot 10^{-3}$	airplane $\approx 300\frac{m}{s}$ $=5 \cdot 10^{-3}$
1	20	05D39	$w[\cdot]$	40.070 dB	37.97 dB	34.64 dB	22.00 dB
2	25	073F536	$w[\cdot]$	40.828 dB	37.90 dB	33.97 dB	20.86 dB
3	30	09BF8EB5	$w[\cdot]$	42.408 dB	38.35 dB	33.84 dB	20.34 dB
4	15	2DE4	$w[\cdot]$	30.982 dB	30.79 dB	30.25 dB	23.49 dB
5	15	2980	$w[\cdot]$	41.279 dB	38.89 dB	35.33 dB	22.56 dB
6	15	2980	$\overline{w}[\cdot]$	40.100 dB	39.11 dB	36.26 dB	25.33 dB

III. Doppler Tolerant Filters

We will first search for sequences $s[\cdot]$ with large $\rho_{sw}(\xi_{max})$-values and then design receiver filters $w[\cdot]$ of length $N = 3L$ with optimized Doppler performance. To simplify the search in the first step, the ambiguity function is expressed as

$$|A_{sw}[m, \xi]|^2 = \sum_n \sum_l s[n] s[l]\, w[n+m] w[l+m] \cos(2\pi\xi(n-l))$$
$$\approx C_{sw}^2[m] - 4\pi^2\xi^2 \left(C_{sw}^{(2)}[m] C_{sw}[m] - \left(C_{sw}^{(1)}[m] \right)^2 \right),$$

where we approximated $\cos(x) \approx 1 - x^2/2$ ($|x| \le 0.1$ yields less than 5% error) and where $C_{sw}^{(t)}[m] = \sum_n s[n] w[n+m] n^t$. Since this is a quadratic equation in ξ, only the cases $\xi = 0$ and $\xi = \xi_{max}$ must be considered. Moreover, this approximation leads to an efficient criterion that allows an exhaustive search up to lengths $L \approx 40$.

In the second step, we determined Doppler tolerant filters $\overline{w}[n]$ by adding constraints on $|\sum_n s[n]\, \overline{w}[n+m] \cos(2\pi\xi_{max} n)|$ and $|\sum_n s[n]\, \overline{w}[n+m] \sin(2\pi\xi_{max} n)|$, or on $|C_{s\overline{w}}^{(1)}(m)|$. Both approaches result in a reduced degradation of $\rho_{sw}(\xi_{max})$ with increasing ξ_{max}. For $\xi_{max} = 5 \cdot 10^{-3}$, the improvement of $\rho_{sw}(\xi_{max})$ may exceed 3 dB (compare #5 with #6). Nevertheless, the complete degradation of the $\rho_{sw}(\xi_{max})$ caused by Doppler frequency shifts cannot be compensated by mismatched filters.

References

[1] J. Ruprecht, F.D. Neeser, M. Hufschmid, "Code time division multiple access: An indoor cellular system", *Proc. VTC'92*, Denver, 1992.

[2] J. Ruprecht, *Maximum-Likelihood Estimation of Multipath Channels*, ISBN 3-89191-270-6, Hartung Gorre Verlag, Konstanz, Germany, 1989.

[3] A.J. Zejak, E. Zentner, P.B. Rapajić, "Doppler optimized mismatched filters", *Electronic Letters*, pp. 558-560, Vol 27, March 1991.

[4] G.S. Mitchell, R.A. Scholtz, "The Use of Doppler Tolerant Reference Signals in Time Synchronization Applications", *28th Asilomar Conf. on Signals, Systems, and Computers*, pp. 450-454, October 1994.

[5] J. Ruprecht, M. Rupf, "On the search and construction of good invertible binary sequences", *Proc. 1994 IEEE International Symposium of Information Theory ISIT'94*, p. 73, Trondheim, Norway, June 1994.

Optimal Sequence Sets Meeting Welch's Lower Bound

Wai Ho Mow[1]

Dept. of Elect. & Comp. Eng., Univ. of Waterloo, Waterloo, Ontario, CANADA N2L 3G1

Abstract — **Welch's lower bounds on total periodic and odd correlation energy of an equi-energy set of sequences are presented. It is shown that both bounds are simultaneously achieved precisely when the sequence set forms an aperiodic complementary sequence set, which has been extensively studied and is of independent interests. Then a lower bound closely related to an approximate SNR formula of Pursley for asynchronous DS/SSMA is derived. Our results are an extension of the works of Massey and Mittelholzer for synchronous DS/SSMA.**

I. Introduction

In spite of the fact that the existing theory of sequence design concerns mainly with the maximum periodic correlation magnitude, it is well-known that the inter-sequence aperiodic cross-correlation energy (i.e. between any two users) are more interesting than the maximum periodic (or even aperiodic) cross-correlation magnitude from the pragmatic viewpoint because they determine the average SNR of an asynchronous DS/SSMA system under proper assumptions [6],[7].

In order to maximize the average SNR of an asynchronous DS/SSMA system by proper choice of signature sequences, sets of binary sequences are typically numerically optimized with respect to the average interference parameter (AIP), which can be accurately approximated by the total aperiodic cross-correlation energy (i.e. sum over all pairs of distinct sequences). In the last two decades, many numerical results about binary sequences with optimized AIP were reported (c.f. [2] and the references therein).

Welch's bound is essentially a lower bound on the total-even-moment of inner products of any set of equi-energy sequences, though it is usually formulated as a bound on maximum inner-product magnitude. Recently, Massey [3] identified the necessary and sufficient condition for a sequence set to meet Welch's bound on the total inner-product energy. This result was subsequently elaborated by Massey and Mittelholzer [4] for application in synchronous DS/SSMA systems. In particular, the uniformly good property of the Welch-Bound-Equality (WBE) sequence sets guarantees that all inter-sequence inner-product energy of such sequence sets simultaneously achieve the same value. This property means that the use of WBE sequence set as the signature sequences for a synchronous DS/SSMA system results in the minimum worse-case interuser interference variance, and is very desirable from an application viewpoint.

II. Main Results

This work is an extension of the results of [3] and [4] to asynchronous DS/SSMA systems, which are considered to be more practical due to the removal of the assumption of ideal sequence synchronization. The following theorems state our main results.

[1]This work was supported by the Croucher Foundation Fellowship 1994/95.

Let X be an equi-energy set of K complex-valued sequences of length L.

Theorem 1 *(Welch's bound on total periodic correlation energy) Let X_S be the sequence set obtained by including all cyclically shifted versions of every sequence in X. Then the total inner-product energy of X_S is at least K^2L^3, with equality if and only if X is a periodic complementary sequence set.*

Theorem 2 *(Welch's bound on total odd correlation energy) Let $X_{\tilde{S}}$ be the sequence set obtained by including all negacyclically shifted versions of every sequence in X. Then the total inner-product energy of $X_{\tilde{S}}$ is at least K^2L^3, with equality if and only if X is a odd complementary sequence set.*

Theorem 3 *(Bound for asynchronous DS/SSMA) Let $C_{i,j}(t)$ denote the aperiodic cross-correlation at phase shift t between the ith and jth sequences in X. Then*

$$\max_{0\le j<K}\left\{\sum_{\substack{i=0\\ i\neq j}}^{K-1}\sum_{t=1-L}^{L-1}|C_{i,j}(t)|^2+\sum_{\substack{t=1-L\\ t\neq 0}}^{L-1}|C_{j,j}(t)|^2\right\}\ge (K-1)L^2,$$

with equality if and only if the sequence set forms an aperiodic complementary sequence set.

Theorem 3 is closely related to the approximate SNR formula of Pursley[6] for asynchronous DS/SSMA. Preliminary forms of Theorems 1 and 2 were presented in [5]. A discussion on binary linear cyclic codes that almost achieve Welch's bound on total periodic correlation energy can be found in [1].

References

[1] A. R. Hammons, Jr., and P. V. Kumar, "On a Recent 4-Phase Sequence Design for CDMA," *IEICE Trans. Commun.*, vol. E76-B, No. 8, pp. 804–813, Aug. 1993.

[2] K. H. A. Kärkkäinen and P. A. Leppänen, "Comparison of the performance of some linear spreading code families for asynchronous DS/SSMA systems," In *Proc. 1991 IEEE Military Communications Conference (MILCOM'91)* (McLean, VA), pp. 784–790, Nov. 4–7 1991.

[3] J. L. Massey, "On Welch's bound for the correlation of a sequence set," *in 1991 IEEE International Symposium on Information Theory (ISIT' 91)*, pp. 385.

[4] J. L. Massey and T. Mittelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in *Sequences II: Methods in Communication, Security, and Computer Science*, R. Capocelli, et al., Eds. New York: Springer-Verlag, 1993.

[5] W. H. Mow, "A study of correlation of sequences," PhD Thesis, Dept. of Information Engineering, the Chinese University of Hong Kong, Shatin, Hong Kong, May 1993.

[6] M. B. Pursley, "Performance evaluation for phase-coded spread-spectrum multiple-access communication — part I: system analysis," *IEEE Trans. Commun.*, vol. COM-25, pp. 795–799, 1977.

[7] 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.

New Signal Design Method by Coded Addition of Sequences

Naoki SUEHIRO

Institute of Applied Physics, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305, Japan

Abstract —

A method of "coded addition of sequences" is proposed for the signal design with many codewords for synchronous or approximately synchronized CDMA systems.

I. Coded Addition of Sequences

The method can be explained using small examples.

We can obtain a 4-phase good code of wordlength 2

$$[(1,j),(1,-j),(j,1),(-j,1),(-1,-j),(-1,j),(-j,-1),(j,-1)]$$

Then, from orthogonal vectors

$$\begin{aligned}\mathbf{x}_1 &= (\ 1\ \ 1\ \ 1\ \ -1\ )\\ \mathbf{x}_2 &= (\ 1\ \ 1\ \ -1\ \ 1\ ),\end{aligned}$$

we can obtain eight vectors by "coded addition" of vectors with above 4-phase code as follows:

$$\begin{aligned}
\mathbf{x}_1 + j\mathbf{x}_2 &= (\ 1+j\ \ 1+j\ \ 1-j\ \ -1+j)\\
\mathbf{x}_1 - j\mathbf{x}_2 &= (\ 1-j\ \ 1-j\ \ 1+j\ \ -1-j)\\
j\mathbf{x}_1 + \mathbf{x}_2 &= (\ 1+j\ \ 1+j\ \ -1+j\ \ 1-j)\\
-j\mathbf{x}_1 + \mathbf{x}_2 &= (\ 1-j\ \ 1-j\ \ -1-j\ \ 1+j)\\
-\mathbf{x}_1 - j\mathbf{x}_2 &= (\ -1-j\ \ -1-j\ \ -1+j\ \ 1-j)\\
-\mathbf{x}_1 + j\mathbf{x}_2 &= (\ -1+j\ \ -1+j\ \ -1-j\ \ 1+j)\\
-j\mathbf{x}_1 - \mathbf{x}_2 &= (\ -1-j\ \ -1-j\ \ 1-j\ \ -1+j)\\
j\mathbf{x}_1 - \mathbf{x}_2 &= (\ -1+j\ \ -1+j\ \ 1+j\ \ -1-j)
\end{aligned} \qquad (1)$$

The Euclid distance between any two of these vectors is always 4, except for the case of the two vectors are inverse each other. Farthermore, all of these vectors are orthogonal to both of

$$\begin{aligned}\mathbf{x}_3 &= (\ \ 1\ \ -1\ \ 1\ \ 1\ )\\ \mathbf{x}_4 &= (\ -1\ \ \ 1\ \ 1\ \ 1\ ).\end{aligned}$$

Above method of "coded addition of vectors" also can be used to the row vectors in the IDFT matrix in following formula of the method of signal making for approximately synchronized CDMA[1]. Because $(1\ j)$ is also an orthogonal sequence, a formula

$$\sqrt{3}F_6^{-1}\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 1 & j & -j & -1 & -1 & -j & j & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & j & -j & -1 & -1 & -j & j\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
j & -j & 1 & 1 & -j & j & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & j & -j & 1 & 1 & -j & j & -1 & -1
\end{bmatrix}$$

$$= \begin{bmatrix}
w^3 & w^{21} & w^3 & w^{21} & w^{15} & w^9 & w^{15} & w^9 & w^3 & w^{21} & w^3 & w^{21} & w^{15} & w^9 & w^{15} & w^9\\
w^1 & w^7 & w^{13} & w^{19} & w^{13} & w^{19} & w^1 & w^7 & w^5 & w^{11} & w^{17} & w^{23} & w^{17} & w^{23} & w^5 & w^{11}\\
w^{11} & w^5 & w^{11} & w^5 & w^{23} & w^{17} & w^{23} & w^{17} & w^{19} & w^{13} & w^{19} & w^{13} & w^7 & w^1 & w^7 & w^1\\
w^9 & w^{15} & w^{21} & w^3 & w^{21} & w^3 & w^9 & w^{15} & w^{21} & w^3 & w^9 & w^{15} & w^9 & w^{15} & w^{21} & w^3\\
w^{19} & w^{13} & w^{19} & w^{13} & w^7 & w^1 & w^7 & w^1 & w^{11} & w^5 & w^{11} & w^5 & w^{23} & w^{17} & w^{23} & w^{17}\\
w^{17} & w^{23} & w^5 & w^{11} & w^5 & w^{11} & w^{17} & w^{23} & w^{13} & w^{19} & w^1 & w^7 & w^1 & w^7 & w^{13} & w^{19}
\end{bmatrix}$$

$$= [\mathbf{x}_{11}\ \mathbf{x}_{12}\ \mathbf{x}_{13}\ \mathbf{x}_{14}\ \mathbf{x}_{15}\ \mathbf{x}_{16}\ \mathbf{x}_{17}\ \mathbf{x}_{18}\ \mathbf{x}_{21}\ \mathbf{x}_{22}\ \mathbf{x}_{23}\ \mathbf{x}_{24}\ \mathbf{x}_{25}\ \mathbf{x}_{26}\ \mathbf{x}_{27}\ \mathbf{x}_{28}]$$

prepares eight polyphase codewords for a user, where $w = \exp(\frac{2\pi j}{24})$. In this case, the user 1 can be assigned 8 pseudo-periodic sequences of length $6 + 2L$:

$$[\mathbf{x}'_{11}\ \mathbf{x}'_{12}\ \mathbf{x}'_{13}\ \mathbf{x}'_{14}\ \mathbf{x}'_{15}\ \mathbf{x}'_{16}\ \mathbf{x}'_{17}\ \mathbf{x}'_{18}],$$

where

$$\mathbf{x}'_{11} = {}^t[\ w^{17}\ \ w^3\ \ w^1\ \ w^{11}\ \ w^9\ \ w^{19}\ \ w^{17}\ \ w^3\ ],$$

when $L = 1$.

The Euclid distance between any two among $[\mathbf{x}_{11} \ldots \mathbf{x}_{18}]$ is the same except for the case that these two are inverse each other. On the other hand, the crosscorrelation function between $\mathbf{x}'_{1i}$ and $\mathbf{x}_{2j}$ is 0 for -1, 0 and 1 shift terms.

II. Discussion

For a synchronous CDMA system, a signal design without co-channel interference is realized by using rows of a unitary matrix. For an approximately synchronized CDMA system, a signal design without co-channel interference is also realized by using the pseudo-periodic sequences proposed by the author[1].

However, in real system, the information transmission rate and the number of users, which can use the system in the same time, are important. So, a user should be assigned many signals, each of which are without co-channel interference to the signals of other users, so that the user can use the assigned signals as codewords.

In this paper, a method of "coded addition of sequences" was proposed for the signal design with many codewords for synchronous or approximately synchronized CDMA systems.

Acknowledgements

The author wishes to thank Prof. Noriyoshi Kuroyanagi for his discussion. The author also wishes to thank SCAT (Support Center for Advanced Telecommunications Technology Research) for the research fund.

References

[1] N.Suehiro, "A signal design without co-channel interference for approximately synchronized CDMA systems," IEEE Journal on Selected Areas in Communications, vol. 12, June 1994.

An Upper Bound for the Aperiodic Correlation of Weighted-Degree CDMA Sequences[1]

Abhijit G. Shanbhag
EEB 522, EE-Systems
Univ. South. Calif.
Los Angeles
CA 90089-2565

P. Vijay Kumar
EEB 534, EE-Systems
Univ. South. Calif.
Los Angeles
CA 90089-2565

Tor Helleseth
Dep. of Informatics
Univ. of Bergen
N-5020, Bergen
Norway

Abstract — **An upper bound for a hybrid exponential sum over Galois rings is derived. This bound is then used to obtain an upper bound for the maximum aperiodic correlation of some recently constructed weighted degree sequence families over Galois Rings. The bound is of the order of $\sqrt{L}\log L$ where L is the period of the sequences.**

I. Introduction

For a fixed prime p and integers $e, m,\ e \geq 2,\ m \geq 1$, let $R_{e,m}$ denote the *Galois Ring* of characteristic p^e and containing p^{em} elements. Let $\psi_{e,m}$ be a non-trivial additive character of $R_{e,m}$ and let $f(x) \in R_{e,m}[x]$ be *non-degenerate* with *weighted degree* D_f [1]. Define $\mathcal{T}_{e,m} = \mathcal{T}^*_{e,m} \cup 0$ where $\mathcal{T}^*_{e,m}$ is a cyclic subgroup of $R^*_{e,m}$ of order $p^m - 1$. In [1], Kumar et al. prove

$$\left| \sum_{x \in \mathcal{T}_{e,m}} \psi_{e,m}(f(x)) \right| \leq (D_f - 1)\sqrt{p^m}. \qquad (1)$$

Consider the set $\mathcal{S}_D$ of sequences defined via

$$\mathcal{S}_D = \left\{ \{T_{e,m}(f(\beta^t))\}_{t \in Z} \mid D_f \leq D \right\}$$

where β is a generator of $\mathcal{T}^*_{e,m}$. Let the set

$$\mathcal{F}_D \subset \mathcal{S}_D$$

consist of a maximal family of pairwise, cyclically distinct sequences in $\mathcal{S}_D$ with each sequence having period $2^m - 1$. Using (1), it is easy to see that the maximum non-trivial correlation C_{max} of the family $\mathcal{F}_D$ has the upper bound

$$C_{max} \leq 1 + (D-1)\sqrt{p^m}.$$

The family $\mathcal{F}_D$ compares very well with existing sequence families when C_{max}, alphabet size and family size are used as a basis for comparison. In this paper, we obtain an upper bound to the maximum aperiodic correlation of the family $\mathcal{F}_D$. The aperiodic correlation is often more relevant than periodic correlation in CDMA applications.

II. Bound on a hybrid exponential sum

Let $f(x) \in R_{e,m}[x]$ have weighted degree D_f. Let $\chi_{e,m}$ be an arbitrary multiplicative character with order dividing $p^m - 1$. Using L-function techniques, we can express the hybrid exponential sum $\sum_{x \in \mathcal{T}_{e,m}} \psi_{e,m}(f(x))\chi(x)$ as a sum of D_f complex numbers. These complex numbers can be shown to be the reciprocal roots of the zeta function of a function field over F_{p^m}. It follows from the *Riemann Hypothesis* for function fields, that the magnitude of each of these complex numbers is $\sqrt{p^m}$. Thus, we have

Theorem 1

$$\left| \sum_{x \in \mathcal{T}_{e,m}} \psi_{e,m}(f(x))\chi_{e,m}(x) \right| \leq D_f\sqrt{p^m}.$$

III. Bound on aperiodic correlation

The aperiodic correlation $\theta_{1,2}(\tau)$ between any two p^e-ary sequences $s_1(t)$ and $s_2(t)$ of period N, is defined via

$$\theta_{1,2}(\tau) = \sum_{t=max\{0,-\tau\}}^{min\{N-1,N-1-\tau\}} \omega^{s_1(t+\tau)-s_2(t)},\ \omega = exp(i2\pi/p^e).$$

The computation of the aperiodic correlation distribution $\theta_{i,j}(\tau),\ 1 \leq \tau \leq N$, of $\mathcal{F}_D$ reduces to obtaining the distribution of the exponential sum values $\{\sum_{t=max\{0,-\tau\}}^{min\{N-1,N-1-\tau\}} \psi_{e,m}(f(\beta^t)) \mid D_f \leq D\}$.

Using Theorem 1, and using similar techniques as in [2] (see also [4], [3]), we can bound the maximum non-trivial aperiodic correlation θ_{max} of $\mathcal{F}_D$ as under

Theorem 2

$$|\theta_{max}| < D\sqrt{p^m}\,(\ln(p^m) + 1).$$

Acknowledgements

The authors would like to acknowledge J. Lahtonen for bringing the connection (see [2]-[4] below) between aperiodic correlation and hybrid exponential sums to their notice.

References

[1] P.V. Kumar, T. Helleseth, and A.R. Calderbank, "An Upper Bound for Weil Exponential Sums over Galois Rings and Applications", *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 456–468, 1995.

[2] S. Litsyn and A. Tietavainen, "Character Sum Constructions of Constrained Error-Correcting Codes", *AAECC*, 5:45–51, 1994.

[3] R.J. McEliece, "Correlation Properties of Sets of Sequences derived from Irreducible Cyclic Codes", *Information and Control*, 45:18–25, 1980.

[4] D.V. Sarwate, "An Upper Bound on the Aperiodic Correlation Function for a Maximal Length Sequence", *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 685–687, 1984.

[1]The work was supported in part by the National Science Foundation under Grant Number NCR-93-05017 and in part by the Norwegian Research Council for Science and the Humanities.

Construction of Signal Sets with Constrained Amplitude Spectrum with Upper Bounds on Cross-Correlation

Girish Chandran and Jules S. Jaffe[1]

Marine Physical Laboratory, Scripps Institution of Oceanography, Univ. of California, San Diego, La Jolla, CA 92093-0238

Abstract — **A new method for designing signals with a given time-bandwidth product is introduced. These signals in the set have a flat amplitude spectrum and have low cross-correlation function values and lie on a signal parameter space ellipse. Upper bounds for the cross-correlation between signals in the set is derived.**

I. Introduction

There are many applications where there is a need for synthesizing signal sets which have low values of cross-correlation at all lags and low values of autocorrelation at nonzero lags. While prolate spheroidal functions are "essentially" time and band-limited, and are orthogonal, the cross-correlation between the signals is not zero for all lags [1]. In an asynchronous system with no cooperation among users/targets, uniformly low cross-correlation values between signals are important. In an imaging context, using signals whose spectrum is broad enough to cover the nulls in the backscattering spectrum of targets ensures reasonable signal to noise ratio [2].

II. Design Constraints

Let $\mathbf{S} = \{s_1(t), s_2(t), ..., s_N(t)\}$ be a set of complex envelopes of signals which are $L_2(\frac{-T}{2}, \frac{T}{2})$ with a corresponding set of Fourier transforms $\mathcal{S} = \{\mathcal{S}_1(f), \mathcal{S}_2(f), ..., \mathcal{S}_N(f)\}$. The design specifications are as follows:

Condition 1:

$$\int_{\frac{-T}{2}}^{\frac{T}{2}} \mid s_i(t) \mid^2 dt = 1; \quad i = 1, ..., N \tag{1}$$

Condition 2: For some $\kappa > 0$ and for all $\tau \leq T$

$$\mid R_{i,j}(\tau) \mid < \kappa; \quad i, j = 1, ..., N; \quad i \neq j \tag{2}$$

where the cross-correlation $R_{i,j}(\tau)$ between signals $s_i(t)$ and $s_j(t)$ is defined as

$$R_{i,j}(\tau) = \int_{-T}^{+T} s_i(t) s_j^*(t-\tau) dt; \quad \tau \leq T$$

Condition 3: For i = 1,2,...,N

$$\mid \mathcal{S}_i(f) \mid = \begin{cases} \alpha_1 & ; \mid f \mid \leq W \\ \alpha_2(f) & ; \mid f \mid > W \end{cases} \tag{3}$$

where α_1 is a constant and $\alpha_2(f)$ is positive function. Let $\alpha_1 = \frac{1}{\sqrt{2W}} - \delta_1$ and $\alpha_2(f) = \delta_2$, where δ_1, $\delta_2 > 0$ are very small, such that $\int_{-W}^{W} \mid \mathcal{S}_i(f) \mid^2 df = 1 - \epsilon$. The signals are "essentially" band-limited with the amplitude of the Fourier transform as specified.

Since the area under the squared magnitude of the cross-correlation function is fixed because of (3), it can be reasoned that the cross-correlation function should be a constant function with a support $[-T, T]$ to achieve uniformly low values of cross-correlation.

[1]This work was supported by the National Science Foundation under grant OCE 89-14300

III. Solution to the Design problem

It has been shown heuristically that for signals with quadratic phase functions in the time and frequency domains the shape of the complex envelopes will be rectangular [3]. Let

$$\mathcal{S}_i(f) = \mid S_i(f) \mid e^{j(a_i f^2 + b_i f + c_i)} \tag{4}$$

By selecting the quadratic coefficients carefully we can also ensure that the difference between the phase functions of two signals, which determines the cross-correlation property, is quadratic. To arrive at a rule to pick the quadratic coefficients the usual definitions of the rms duration γ and rms bandwidth β are used [3]. Using these definitions, it can be shown that the quadratic coefficients lie on an ellipse, i.e.,

$$\frac{a_i^2}{(\frac{\pi\gamma}{\beta})^2} + \frac{b_i^2}{\gamma^2} = 1 \tag{5}$$

IV. Upper Bound for Cross-Correlation

Let the real and imaginary parts of a Fresnel integral be $C(x) = \int_0^x \cos(\frac{\pi t^2}{2}) dt$ and $S(x) = \int_0^x \sin(\frac{\pi t^2}{2}) dt$. $R_{i,j}$ is a continuous function of τ, Δa, Δb and Δc, where $\Delta a = a_i - a_j$; $\Delta b = b_i - b_j$; $\Delta c = c_i - c_j$. Since $R_{i,j}(\tau)$ is a continuous function so is $\mid R_{i,j}(\tau) \mid$. This means that $\max_{\tau \leq T} \mid R_{i,j}(\tau) \mid$ exists and is finite.

Theorem: $\max_{|\tau| \leq T} \mid R_{i,j}(\tau) \mid \leq 2.3(\frac{1}{2W}\sqrt{\frac{\pi}{2\Delta a}})$

Proof:

$$\mid R_{i,j}(\tau) \mid = \frac{1}{2W}\sqrt{\frac{\pi}{2\Delta a}}[C(x_1) + jS(x_1) - C(x_0) - jS(x_0)] \tag{6}$$

where $x_1 = \sqrt{\frac{2}{\pi\Delta a}}((2\pi\tau + \Delta b) + W\Delta a)$ and $x_0 = \sqrt{\frac{2}{\pi\Delta a}}((2\pi\tau + \Delta b) - W\Delta a)$. $\max_{x_0, x_1 \in (-\infty, +\infty)}[\mid C(x_1) - C(x_0) \mid] \leq 1.6$. Also, $\max_{x_0, x_1 \in (-\infty, +\infty)}[\mid S(x_1) - S(x_0) \mid] \leq 1.6$. Thus

$$\mid C(x_1) + jS(x_1) - C(x_0) - jS(x_0) \mid \leq 2.3$$

Corollary: For a given duration and bandwidth for the signal set, the cross-correlation between two signals that are furthest apart along the semi-minor axis, in the set is bounded by $\frac{1}{\sqrt{TW}}$.

It can be shown that it is possible to trade-off the number of signals on the signal parameter ellipse for better cross-correlation properties between signals in the set.

References

[1] D. Slepian and H.O. Pollack, "Prolate spherodial wave functions, Fourier analysis, and uncertainty-I," Bell Syst. Tech. J., **40**, 43-63 (1961)

[2] J.S. Jaffe, E. Reuss, D. McGehee, and G. Chandran, "FTV, a Sonar for Tracking Macrozooplankton in 3-dimensions," to be published in Deep Sea Research

[3] A.W. Rihaczek, "*Principles of High-Resolution Radar*," California: Peninsula, 1985

ON GRÖBNER BASES OF THE ERROR-LOCATOR IDEAL OF HERMITIAN CODES

Xuemin Chen, I. S. Reed and T. Helleseth*

1. ERROR-LOCATOR IDEALS FOR HERMITIAN CODES

Consider error-correcting codes constructed from an affine version of the Hermitian curve. Let $K = GF(q)$ and let $m = \sqrt{q}+1$ be an integer. In this case the affine version of the Hermitian curve, $C(x,y) = x + x^{m-1} - y^m$, is irreducible, regular, and has exactly $n = q\sqrt{q}$ rational points, given by $P_n = \{(x_1,y_1),(x_2,y_2),.....,(x_n,y_n)\}$. The genus g of this curve is given by $g = (m-1)(m-2)/2$. The total degree ordering(TDO) $<_t$ of the pairs (a,b) of the positive integers is chosen as follows :

$$(0,0) <_t (1,0) <_t (0,1) <_t (2,0) <_t (1,1) <_t (0,2) <_t \cdots.$$

In the TDO let j be a positive integer such that $m-2 \le j \le \lfloor \frac{n-1}{m} \rfloor$ and let $\phi_0(x,y)$, $\phi_1(x,y)$,...,$\phi_u(x,y)$ denote the monomials $x^a y^b$ for $(a,b) \le_t (0,j)$. The Hermitian code $\mathcal{C}$ is then defined by its parity check matrix H :

$$H = \begin{pmatrix} \phi_0(x_1,y_1) & \cdots & \phi_0(x_n,y_n) \\ \phi_1(x_1,y_1) & \cdots & \phi_1(x_n,y_n) \\ \vdots & & \vdots \\ \phi_u(x_1,y_1) & \cdots & \phi_u(x_n,y_n) \end{pmatrix}. \quad (1)$$

The dimension and the designed distance of the code $\mathcal{C}$ satisfy $k = n-(mj-g+1)$ and $d^* = mj-2g+2 \le d$, respectively, where d denotes the true minimum distance of the code $\mathcal{C}$.

In the decoding situation a received word $\mathbf{r}$ is the sum of a codeword $\mathbf{c}$ and an error vector $\mathbf{e}$. The syndrome vector $\mathbf{s}$ is computed as usual by $\mathbf{s} = \mathbf{r}H^T$. Assume that $v = wt(\mathbf{e}) \le t$, where $t = \lfloor (d-1)/2 \rfloor$. Also, assume that an error which occurs in the i-th coordinate of $\mathbf{r}$ is denoted by $e_i (\neq 0)$. Then the error-location set of $\mathbf{e}$ is defined by $EP_{xy} = \{(x_i,y_i) : i \in Z_n \text{ and } e_i \neq 0\}$, where $Z_n = \{i : 1 \le i \le n\}$. It follows from (1) that $s_{ab} = \sum_{i \in Z_e} e_i x_i^a y_i^b$ for $a+b \le j$ are the known syndromes for the errors of the Hermitian code $\mathcal{C}$, where $Z_e = \{i : i \in Z_n \text{ and } e_i \neq 0\}$ is called the error-location index set. The decoding problem is to use these syndromes s_{ab} to determine the $v (\le t)$ error positions (x_i,y_i) and the corresponding error values e_i for $i \in Z_e$.

Usually, the determination of the error positions is based on the observation that if any polynomial, $f(x,y) = \sum_{v+w \le h} f_{vw} x^v y^w$, has the same error positions as the received word among its zeros, then $\sum_{v+w \le h} f_{vw} s_{a+v,b+w} = 0$. This implies that the procedure for determining the error positions is independent of the method needed to find the error values. The error-locator ideal of $\mathbf{e}$ is defined next.

Definition 1 *The polynomial ideal,*

$$I_{\mathbf{e}}(x,y) = \{f(x,y) \in K[x,y] : f(x_i,y_i) = 0 \text{ for all } i \in Z_e\},$$

is called the error-locator ideal of the error vector $\mathbf{e}$.

2. DETERMINING GRÖBNER BASES OF THE ERROR-LOCATOR IDEAL

For brevity, define the following polynomials :

$$f_{ab} = E_1 X_1^a Y_1^b + E_2 X_2^a Y_2^b + ... + E_v X_v^a Y_v^b - s_{ab}, \quad (2)$$

$$h_j = c(X_j,Y_j), \quad (3)$$

$$l_{1j} = X_j^q - X_j,\ l_{2j} = Y_j^q - Y_j,\ l_{3j} = E_j^{q-1} - 1, \quad (4)$$

over the set of variables X_j, Y_j, E_j for $1 \le j \le v$. For a received word $\mathbf{r} = \mathbf{c} + \mathbf{e}$ with $v = wt(\mathbf{e}) \le t$, the problem of decoding Hermitian codes is equivalent to solving for the common zeros of the following set of multivariate non-linear equations : $f_{ab} = 0$ *for* $a + b \le j$, and the equations, $h_j = 0$, $l_{1j} = 0$, $l_{2j} = 0$, $l_{3j} = 0$ for $j = 1,2,...,v$.

Consider the polynomial ring $K[X_1,Y_1,E_1,...,X_v,Y_v,E_v]$ and the following set of polynomials : $F = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, where the sets $\mathcal{F}_j$ are given by $\mathcal{F}_1 = \{f_{ab} : a+b \le j\}$, $\mathcal{F}_2 = \{h_j : 1 \le j \le v\}$, and $\mathcal{F}_3 = \{l_{ij} : 1 \le i \le v, 1 \le j \le v\}$ with the polynomials f_{ab}, h_j and l_{ij} being defined by (2),(3) and (4), respectively. Thus, the problem of decoding Hermitian codes is equivalent to a determination of the variety $V(F)$ or its equivalent $V(I(F))$. The key observation is the following relation between the ideal $I(F)$ and the error-locator ideals $I_{\mathbf{e}}(X_j,Y_j)$:

Theorem 1 $I(F) \cap K[X_j,Y_j] \subset I_{\mathbf{e}}(X_j,Y_j)$ *for* $j = 1,2,...,v$, *and* $V(I_{\mathbf{e}}(X_j,Y_j)) = V(I(F) \cap K[X_j,Y_j])$ *for* $j = 1,2,...,v$.

In order to solve for the error-locations from the error-locator ideal $I_{\mathbf{e}}(X_j,Y_j)$, one needs to determine a set of generators for this ideal. First, define the projection sets $EP_x = \{\alpha : (\alpha,\beta) \in EP_{xy}\}$ and $EP_y = \{\beta : (\alpha,\beta) \in EP_{xy}\}$. Next, define the "purely lexicographical"(PLEX) ordering of the m-tuples $(a_1,a_2,...,a_m)$ as follows : $(0,0,...,0) <_p (1,0,...,0) <_p (2,0,...,0) <_p \cdots <_p (0,1,...,0) <_p (0,2,...,0) <_p \cdots$. Theorem 1 implies the following important theorem for the normalized reduced Gröbner basis(NRGB) of $I(F)$:

Theorem 2 *Let* G_p *be the NRGB of* $I(F)$ *w.r.t. PLEX ordering exponents of the monomials* $X_1^{a_1}Y_1^{a_2}E_1^{a_3}...X_v^{a_{3v-2}}Y_v^{a_{3v-1}}E_v^{a_{3v}}$. *Then* $G_p \cap K[X_1,Y_1] = \{g_2(X_1,Y_1), g_1(X_1)\}$ *and* $V(G_p \cap K[X_1,Y_1]) = EP_{xy}$, *where* $g_2(x,y) \in K[x,y]$ *and* $g_1(x) \in K[x]$.

The above theorems provide an approach for producing from $I(F)$ a minimal set of generators for the ideal $I(F) \cap K[X_j,Y_j]$. Following this approach, a decoding method based on Buchberger's algorithm [4] is developed as follows:

Decoding Method :

(1) Initialize : Give F and set $v = 0$.

(2) Set $v = v+1$ and apply the Buchberger algorithm(w.r.t. PLEX ordering) to F.

(3) If $|V(F)| = 0$ and $v < t$, goto (2); otherwise, find $G_p \cap K[X_1,Y_1]$, where G_p is the set of generator polynomials obtained by Buchberger's algorithm.

(4) Determine the error positions by solving $G_p \cap K[X_1,Y_1]$ for $V(G_p \cap K[X_1,Y_1])$.

(5) Solve for the error magnitudes e_i.

REFERENCES

[1] X.Chen, I. S. Reed, T. Helleseth and T. K. Truong, "Use of Gröbner Bases to Decode Binary Cyclic Codes up to the True Minimum Distance", IEEE Trans. on Inform. Theory, Vol.40,No.5,pp.1654-1661,Sept. 1994.

[2] —,"General principles for the algebraic decoding of cyclic codes", IEEE Trans. on Inform. Theory,Vol.40,No.5,pp.1661-1663, Sept. 1994.

[3] —, "Algebraic decoding of cyclic codes : a polynomial ideal point of view", Contemporary Mathematics,Vol.168,AMS,1994.

[4] B. Buchberger, "Gröbner Bases : An Algorithmic Method in Polynomial Ideal Theory", N. K. Bose (ed.) Multidimensional Systems Theory, pp.184-232, D. Reidel Publishing Company, 1985.

*X.Chen was with the department of electrical engineering, University of Southern California(USC), Los Angeles. He is now with Advanced Development of Communication Division, General Instrument Corp.,6262 Lusk Blvd.,San Diego,CA 92121. I. S. Reed is with the Department of electrical engineering,USC,LA,CA 90089-2565. T. Helleseth is with the department of informatics, University of Bergen, Høyteknologisenteret, N-5020 Bergen, Norway.This work was supported by the NSF under Grant NCR-9016340 and the Norwegian Research Council for Science and the Humanities

A New Approach to Determine a Lower Bound of Generalized Hamming Weights Using an Improved Bezout Theorem

Gui-Liang Feng and T. R. N. Rao
The Center for Advanced Computer Studies, USL
Lafayette, LA. 70504, USA

Summary

In this paper, a new approach to determine a lower bound for the generalized Hamming weights of algebraic-geometric (AG) codes is discussed.

Let *LS* be a location set and let $H \triangleq \{ \mathbf{h}_1, ..., \mathbf{h}_n \}$ be a well-behaving sequence of monomials based on *LS*. Let $I_{\{ \mathbf{h}_{r_1}, \cdots, \mathbf{h}_{r_p} \}}$ be a subset of *LS* called a *maximal partially linearly dependent location set*, on which $\mathbf{h}_{r_p}$ is consistently and partially linearly dependent on its previous monomials. Define $D_{\{ \mathbf{h}_{r_1}, \cdots, \mathbf{h}_{r_p} \}} = | I_{\{ \mathbf{h}_{r_1}, \cdots, \mathbf{h}_{r_p} \}} |$. It is called the *consistent dependent-degree* of monomials $\mathbf{h}_{r_1}, \cdots, \mathbf{h}_{r_p}$. We define $D_p^{(r)} \triangleq \max \{ D_{\{ \mathbf{h}_{i_1}, \mathbf{h}_{i_2}, \cdots, \mathbf{h}_{i_p} \}} \mid 1 \le i_1 < i_2 < \cdots < i_p \le r \}$.

Theorem: For a linear code C_r defined by $\mathbf{H}_r = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_r]^T$, if there is some d^* such that $D_{r-d^*+h+1}^{(r)} < d^* - 1$, then the generalized Hamming weight d_h is equal to or greater than d^*.

Thus, the determination of a lower bound of the generalized Hamming weights reduces to the calculation of $D_p^{(r)}$. Using an improved Bezout theorem, for the AG codes defined by a large class of plane curves, the value of $D_p^{(r)}$ can be easily determined. In the following we show one example. Let the curve be a Hermitian curve over GF(2^4): $x^5 + y^4 + y = 0$. We have the following well-behaving sequence *H*:

$H = \{ 1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3, x^4, x^3y, x^2y^2, xy^3, x^5, x^4y, x^3y^2, x^2y^3, ... \} = \{ x^iy^j \mid 0 \le i \le 15, 0 \le j \le 3 \}$.

Let us consider C_{16}, i.e., $r = 16$. The first 16 monomials are as follows: $\{ 1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3, x^4, x^3y, x^2y^2, xy^3, x^5, x^4y \}$. Using the calculation of $D_p^{(r)}$, we have the following values.

$D_1^{(16)} = 21$ $D_2^{(16)} = 17$ $D_3^{(16)} = 16$ $D_4^{(16)} = 13$
$D_5^{(16)} = 12$ $D_6^{(16)} = 10$ $D_7^{(16)} = 9$ $D_8^{(16)} = 8$
$D_9^{(16)} = 7$ $D_{10}^{(16)} = 6$ $D_{11}^{(16)} = 5$ $D_{12}^{(16)} = 4$
$D_{13}^{(16)} = 3$ $D_{14}^{(16)} = 2$ $D_{15}^{(16)} = 1$ $D_{16}^{(16)} = 0$.

From these values and the above theorem, we have $d_1(C_{16}) \ge 12$, $d_2(C_{16}) \ge 15$, $d_3(C_{16}) \ge 16$, $d_4(C_{16}) \ge 19$, $d_5(C_{16}) \ge 20$, $d_6(C_{16}) \ge 21$, $d_7(C_{16}) \ge 23$, and $d_h(C_{16}) \ge h + 16$, for $h = 8, 9, 10, 11, ..., 48$.

Using this new approach, some more efficient linear codes with the minimum distances 4, 5, 6 and any lengths over $GF(2^m)$, and some more efficient AG codes have also been constructed in this paper.

References

[1] V. K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. on Information Theory* Vol. IT-37, pp. 1412-1428, Sept., 1991.

[2] K. Yang, P. V. Kumar, and H. Stichtenoth, "On the Weight Hierarchy of Geometric Goppa Codes," *IEEE Trans. on Information Theory*, Vol. IT-40, pp. 913-920, May 1994.

[3] G. L. Feng and T. R. N. Rao, "A Simple Approach for Construction of Algebraic Geometric Codes from Affine Plane Curves," *IEEE Trans. on Information Theory* Vol. IT-40, No.4, pp. 1003-1012, July 1994.

[4] G. L. Feng and T. R. N. Rao, "Improved Geometric Goppa Codes, Part I: Basic Theory" to appear in *IEEE Trans. on Information Theory.*

Fast Erasure-and-Error Decoding of Any One-Point AG Codes up to the Feng-Rao Bound

Shojiro Sakata [1]

Dept. Comp. Sc. & Inf. Math., Univ. of Elect.-Comm., Chofu, Tokyo 182, JAPAN

Recently fast decoding methods ([1] [2] [3], etc.) of algebraic-geometric (AG) codes have been proposed as applications of Sakata algorithm (the multidimensional Berlekamp-Massey algorithm) [4]. Similar but distinct fast decoding algorithms have been presented by [5] [6], etc. Among them, [3] [5] [6] give fast decoding methods for generic one-point AG codes from any algebraic curves in the projective space. These methods are more efficient than the original Feng-Rao decoding method [7]. In particular, [3] is concerned with the multidimensional syndrome array instead of with the syndrome matrix, and employs a unique scheme of majority logic to find the unknown syndrome values necessary for decoding up to half the Feng-Rao bound (designed distance) d_{FR} in the framework of Sakata algorithm, where d_{FR} is greater than or equal to the Goppa bound d_G in general [8].

To improve the probability of correct decoding, it is desirable to devise an efficient decoding algorithm which can correct both errors and erasures. Skorobogatov and Vlăduţ [9] were the pioneers of erasure-and-error decoding of AG codes. Their method can correct t errors and τ erasures such that $2t + \tau < d_G - g$, where g is the genus of the curve difining the AG code. Extending their error-only decoding method [7], Feng and Rao [10] gave an erasure-and-error decoding method which can correct t errors and τ erasures such that $2t + \tau < d_{FR}$.

In this paper we propose a fast erasure-and-error decoding method based on a unification of our error-only decoding method [3] and the algorithm [11] for finding a minimal polynomial vector set of a vector of multidimensional arrays. Our main concern is how to find the unknown syndrome values and the error locations in addition to the given erasure locations more efficiently than the Feng-Rao's scheme based on matrix calculations [10].

We take a one-point AG code (over a finite field K) $C := \{(c_1, \cdots, c_n) \in K^n | \sum_{j=1}^{n} c_j f(P_j) = 0, f \in L(mP_\infty)\}$ from an irreducible nonsingular projective curve $\mathcal{C}$, where $L(mP_\infty)$ is a linear subspace of the algebraic function field $K(\mathcal{C})$ which is composed of functions f having a single pole of order $o(f) \leq m$ at P_∞. In fast decoding of AG codes, we manipulate two kinds of entities, i.e., functions f ($\in K[\mathcal{C}] := \cup_{m \geq 0} L(mP_\infty)$) (treated as multivariate polynomials) and multidimensional syndrome arrays. For our purpose, a kind of vectoral notation or data structure is crucial. That is, while we can represent a multidimensional (error or erasure) syndrome array u as an array vector $(u^{(1)}, \cdots, u^{(\lambda)})$ having λ component 1D arrays $u^{(i)}$, $1 \leq i \leq \lambda$, we represent each (error locator or erasure locator) polynomial (i.e., function) f as a polynomial vector $(f^{(1)}, \cdots, f^{(\lambda)})$ having λ component univariate polynomials $f^{(i)}$, $1 \leq i \leq \lambda$, where λ is the smallest nonzero nongap (pole order) of functions $f \in K[\mathcal{C}]$. Including an algorithm (Algorithm 1) similar to that presented in [11], we can construct a fast erasure-and-error decoding algorithm consisting of two stages. In the first stage, we find a system of λ erasure locator polynomial vectors by applying Algorithm 1 to the erasure syndrome array vector, and by using its result, we modify the errata (i.e., error plus erasure) syndrome array vector. Then, in the second stage, we can find unknown errata syndrome values by invoking a kind of majority logic for the modified errata syndrome array vector with the aid of Algorithm 1, and finally we obtain a system of λ errata locator polynomials. The computational complexity is of order $\mathcal{O}(\lambda n^2)$.

[1] This work was supported by the Science Foundation of the Japanese Educational Ministry under Grant No.20064157.

References

[1] J. Justesen, K.J. Larsen, H. Elbrønd Jensen, and T. Høholdt, "Fast decoding of codes from algebraic plane curves", *IEEE Transactions on Information Theory*, vol.39, pp.37–45, Jan. 1993.

[2] S. Sakata, J. Justesen, Y. Madelung, H. Elbrønd Jensen, and T. Høholdt, "A fast decoding method of AG codes from Miura-Kamiya curves C_{ab} up to half the Feng-Rao bound", *Proc. 1993 IEEE Workshop on Information Theory*, Shizuoka, Japan, June 1993; *Finite Fields and Their Applications*, vol.1, pp.83–101, Jan. 1995.

[3] S. Sakata, H. Elbrønd Jensen, and T. Høholdt, "Generalized Berlekamp-Massey decoding of algebraic geometric codes up to half the Feng-Rao bound", Sept. 1994; submitted for *IEEE Transactions on Information Theory*.

[4] S. Sakata, "Extension of the Berlekamp-Massey algorithm to N dimensions", *Information and Computation*, vol.84, pp.207–239, Feb. 1988.

[5] R. Kötter, "Fast generalized minimum distance decoding of algebraic geometric and Reed Solomon codes", Linköping, Sweden, Aug. 1993.

[6] G.L. Feng, V.K. Wei, T.R.N. Rao, and K.K. Tzeng, "Simplified understanding and efficient decoding of a class of algebraic-geometric codes", *IEEE Transactions on Information Theory*, vol.40, pp.981–1002, 1994.

[7] G.L. Feng and T.R.N. Rao, "Decoding algebraic-geometric codes up to the designed minimum distance", *IEEE Transactions on Information Theory*, vol.39, pp.37–45, 1993.

[8] C. Kirfel and R. Pellikaan, "The minimum distance of codes in an array coming from telescopic semigroups", Preprint presented at the Fourth Workshop on Arithmetic Geometry and Coding Theory", Luminy, France, June, 1993.

[9] A.N. Skorobogatov and S.G. Vlăduţ, "On the decoding of algebraic-geometric codes", *IEEE Transactions on Information Theory*, vol.36, pp.1051–1060, 1990.

[10] G.L.Feng and T.R.N. Rao, "Erasure-and-error decoding of algebraic-geometric codes", *Proceedings of 1993 IEEE Information Theory Workshop*, Shizuoka, Japan, June, 1993.

[11] S. Sakata, "Finding a minimal polynomial vector of a vector of nD arrays", *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, Proceedings of AAECC-9*, New Orleans, USA: Lecture Notes in Computer Science, 539, Springer Verlag, pp.414–425, 1991.

The (64,32,27) Hermitian Code and Its Application in Fading Channels

X. Chen and I.S Reed

Introduction

In a trellis-coded modulation(TCM) scheme, a transmitted message is determined by the current received bit and a number of previously received bits. Therefore, if the decoder makes a mistake, errors have the possibility of propagating. Such a propagation of errors is considered to be a drawback in some channels such as mobile radio channels with slow-shadowing fading. In such a case errors due to the shadowing can affect the decoding process of the symbols within an unshadowed time period and lead to long-term error propagation. Thus, in such a scenario, block-coded modulation(BCM) may have advantages because the decoding of a received code block is independent of any other blocks. The commonly used BCM schemes included the extended Reed-Solomon(RS) codes combined with M-ary Phase-Shift Keying (MPSK) signaling for the bandwidth-limited fading channels. For example, the (16,8,9) extended RS code, defined over $GF(2^4)$, is coded with a 16-PSK signal set. In recent years one of the most exciting developments in the field of error correcting codes is the construction and decoding of algebraic geometry(AG) codes. It is shown in [1] that a sequence of Hermitian codes can be found by the use of results from AG which generalize the original construction of the RS codes. It is proved that under certain conditions that there exist "good" codes within this class of codes. Further, as an example, van Lint and Springer claim for any practical channel that, the specific AG code, namely the (64,32,27) Hermitian code, has a considerably better performance than the corresponding (16,8,9) extended RS code. In this paper,the (64,32,27) Hermitian code and its application in fading channels is discussed.

Definition and Encoding of the (64,32,27) Hermitian Code

To construct the (64,32,27) Hermitian code, consider the Hermitian curve of degree $m=5$,i.e.$C(x,y)=x^5+y^4+y=0$ over $GF(2^4)$. This curve has exactly $n=64$ rational points and the genus of this curve is $g=6$. The set of these rational points can be computed from the cyclic group of order 15, generated by the irreducible polynomial $\pi(x)=x^4+x+1$Denote this set by$P_{64}=\{(x_1,y_1),(x_2,y_2),...,(x_{64},y_{64})\}$. Since $64>8\times5$, a Hermitain code C can be defined by its parity check matrix as follows :

$$H=\begin{pmatrix} \phi_1(x_1,y_1) & ..... & \phi_1(x_{64},y_{64}) \\ \phi_2(x_1,y_1) & ..... & \phi_2(x_{64},y_{64}) \\ . & . & . \\ . & . & . \\ \phi_{32}(x_1,y_1) & ..... & \phi_{32}(x_{64},y_{64}) \end{pmatrix},$$

where $\phi_1(x,y),\phi_2(x,y),...,\phi_{32}(x,y)$ denote the monomials x^ay^b for $(a,b)\le_t(0,8)$ and $a<5$ with $\le_t$ being the total ordering. Therefore, the dimension of C is $k=32$. The designed minimum distance of this code is defined to be $d^*=32-6+1=27$. The true minimum distance d of C satisfies $d\ge d^*$ since $g=6<32$. Finally the result $d=d^*=27$ is determined from Theorem 5 of [1].A transform encoding method of the (64,32,27) Hermitian code C is given by the following theorem :

Theorem 1 Let $c(x,y)=\sum_{i=1}^{32}m_i\phi_i(x,y)$,where for $i=1,2,...,32$ the m_i are the message symbols.Then $c=(c(x_1,y_1),c(x_2,y_2),...,c(x_{64},y_{64}))$ is a codeword in C.

This theorem can be proved by Theorem 1 in [1] from the fact that the code C defined above is a self-dual code. A method for recovering the message symbols is considered next. Let $c=(c_1,c_2,...,c_{64})$, $c_i\in GF(2^4)$ be the codeword encoded by the above encoding method, and let$d_i=\sum_{j=1}^{64}c_jx_j^5y_j^{14}\phi_i^{-1}(x_j,y_j)$, for $i=1,2,...,32$,where the (x_i,y_i) and $\phi_i(x_i,y_i)$ are defined as above. Then, the following theorem holds :

Theorem 2 The message symbol vector m of the codeword c satisfies $m=(d_1,d_2,...,d_{32})$.

This theorem is verified readily by a computer search.

Successive-Erasure Minimum-Distance Decoding of the (64,32,27) Hermitian Code

A fast error-only-decoding algorithm for the Hermitian codes is developed in [2] by Feng and Rao. Then the Feng-Rao algorithm is generalized to an error-and-erasures decoding(EED) in [3]. In this paper,a new Successive-Erasure Minimum- Distance Decoding for the (64,32,27) Hermitian code is discussed. Let $r=(r_1,r_2,...,r_{64})$ be the received word corresponding to c. Consider the decoder be a full maximum-likelihood detector which stores all 16 likelihood functions for each received symbol r_i, i.e. $p(r_i|c_j)$ for all j, where $p(\bullet|\bullet)$ denotes the conditional probability density function. Based on this information the detector provides an estimate $\hat{c}_i$ for each r_i such that $p(r_i|c_j)$ is the greatest. Next define an estimate of the log likelihood ratio to be $L(\hat{c}_i)=\ln\frac{p(r_i|\hat{c}_i(r_i))}{\sum_{c_j\neq\hat{c}_i}p(r_i|c_j)}$. Then the decoding algorithm can be summarized as follows :

Algorithm (1) Succesively erase pairs of symbols with the lowest $L(\hat{c}_i)$ values and apply the Feng-Rao EED algorithm to the estimated word $\hat{c}=(\hat{c}_1,\hat{c}_2,...,\hat{c}_{64})$ with erasures. (2) Iterate (1) $\frac{d+1}{2}=14$ times. During each iteration of this process an estimate of the transmitted codeword is obtained and stored. (3) The decoder chooses that single codeword for which $p(r|\hat{c})$ is the greatest,i.e. the codeword closest to the received vector r in likelihood distance.

The (64,32,27) Hermitian-Coded 16-PSK Scheme

A block coded MPSK scheme is developed by combining the (64,32,27) Hermitian code, defined over $GF(2^4)$, with a 2^4-PSK signal set. In this combination the rate of the coded scheme is the same as the uncoded 2^4-PSK. But the time diversity of the coded scheme is determined by the minimum Hamming distance ($d=27$)of the code. Since the minimum Hamming distance of the (64,32,27) Hermitian code is much larger than the corresponding (16,8,9) extended RS code, a high coding gain is expected for the new (64,32,27) Hermitian-coded 16 PSK scheme. In evaluating the error bounds of this coded scheme on a Rayleigh fading channel at the bit-error rates around 10^{-5}, more than a 26 dB coding gain, compared to uncoded QPSK, is obtained by the use of the new successive-erasure minimum-distance decoder.

References

[1] H.Stichtenoth, "A note on Hermitian codes over $GF(q^2)$," IEEE Trans. Inform. Theory, Vol. 34, pp.1345-1348, Sept. 1988.

[2] G. L. Feng and T. R. N. Rao, "Decoding algebraic-geometric codes up to the designed minimum distance", IEEE Trans. Inform. Theory, Vol. 39, pp.37-45, Jan. 1993.

[3] G. L. Feng and T. R. N. Rao,"Erasure-and-Errors decoding of algebraic-geometric codes", IEEE workshop on Information Theory, 1993.

1 This work was supported in part by the NSF under Grant NCR-9016340. X. Chen was with the Department of Electrical Engineering, University of Southern California, Los Angeles. He is now with the Advanced Development of Communication Division, General Instrument Corp., 6262 Lusk Blvd., San Diego, CA 92121.I.S.Reed is with the the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2565.

New Construction of Codes from Algebraic Curves[1]

B.-Z. Shen and K.K. Tzeng
Department of Electrical Engineering and Computer Science
Lehigh University, Bethlehem, PA 18015 USA

Abstract — **A new construction of linear codes from algebraic curves is introduced. In essence, the construction is of the BCH type, namely, it is to extend the method of constructing BCH codes to the construction of codes from algebraic curves. As a consequence, a new class of codes is constructed without relying much on algebraic geometry. A comparison to algebraic-geometric codes from Hermitian curves showed that our codes typically have much larger minimum distance at higher code rate. In particular, compared to Hermitian codes on $H(2^a)$, which have length 2^{3a}, then, at higher code rate, our codes have minimum distance at least $2^{\lfloor a/4 \rfloor}$ times greater than that of the Hermitian codes. Examples have also shown that, for the same code length and designed minimum distance, our codes can have higher dimension compared to codes constructed from the approach given by Feng and Rao.**

Constructing linear codes from algebraic curves is a relatively new technique for obtaining codes of better rate or higher minimum distance, as well as codes of longer length. It was proved by Tsfasman, Vlăduţ and Zink [1] that from algebraic curves a sequence of codes which exceeds the Gilbert-Varshamov bound can be constructed using Goppa's construction. Codes constructed from Goppa's approach is now called algebraic-geometric (AG) codes. Lately, much work has been done toward non-algebraic-geometric or simplified construction of AG codes [2, 3, 4]. Most recently, based on their simplified approach of AG codes, Feng and Rao constructed improved AG codes [5].

In this paper, a new method constructing of linear codes from algebraic curves is introduced. In essence, the construction is of the BCH type, namely, it is to extend the method of constructing BCH codes to the construction of codes from algebraic curves. As a consequence, a new class of codes is constructed without relying much on algebraic geometry. A comparison to algebraic-geometric codes from Hermitian curves showed that our codes typically have much larger minimum distance at higher code rate. In particular, compared to Hermitian codes on $H(2^a)$, which have length 2^{3a}, then at higher code rate, our codes have minimum distance at least $2^{\lfloor a/4 \rfloor}$ times greater than that of the Hermitian codes. Examples have also shown that, for the same code length and designed minimum distance, our codes can have higher dimension compared to codes constructed from the approach given by Feng and Rao [5].

A brief description of the construction follows:

Let α be a primitive element of $GF(q^2)$. For $i = 0, \ldots, q^2 - 1$, denote $\alpha_0 = 0$, $\alpha_i = \alpha^{i-1}$, for $i \geq 1$. Let $\beta_{i,1}, \ldots, \beta_{i,q}$ be the q solutions of $y^q + y = \alpha_i^{q+1}$ over $GF(q^2)$. Then we have q^3 distinct pairs $(\alpha_i, \beta_{i,j})$, which correspond to all the rational points of the Hermitian curve $H(q)$: $U^{q+1} + V^{q+1} + W^{q+1} = 0$, except a point at infinity. Let n be an integer $0 < n \leq q^3$ and $\theta = \lfloor n/q^2 \rfloor$, then $n = \theta q^2 + \xi$. We denote every $x \in GF^n(q^2)$ by $(x_{0,1}, \ldots, x_{q^2-1,1}, x_{0,2}, \ldots, x_{q^2-1,2}, \ldots, x_{0,\theta}, \ldots, x_{q^2-1,\theta})$ if $\xi = 0$, and by $(x_{0,1}, \ldots, x_{q^2-1,1}, \ldots, x_{0,\theta+1}, \ldots, x_{\xi-1,\theta+1})$ if $\xi \neq 0$. Let δ be a positive integer. Define

$$R_v := \{0, 1, \ldots, \lfloor \frac{\delta}{v} \rfloor - 1\} \text{ for } v \leq \delta$$

and

$$R := \bigcup_{v=1}^{\delta_n} \{(u, v-1) | u \in R_v\}, \text{ where } \delta_n = \min\{\delta, \lceil n/q^2 \rceil\}.$$

Then, we define the following linear code:

$$C(n, \delta) := \{c \in GF^n(q^2) | \sum_{j=1}^{\lceil n/q^2 \rceil} \sum_{i=0}^{k_j} c_{i,j} \alpha_i^\mu \beta_{i,j}^\nu = 0, (\mu, \nu) \in R\}$$

where $k_j = q^2 - 1$ for $j = 1, \ldots, \theta$ and $k_{\theta+1} := \xi - 1$ when $\xi \neq 0$.

Theorem *$C(n, \delta)$ has minimum distance $d \geq \delta + 1$. The dimension of $C(n, \delta)$ is $\geq n - \delta - \delta^*_{n,\delta}$, where $\delta^*_{n,\delta} := \sum_{i=2}^{\delta_n} \lfloor \delta/i \rfloor$, with equality holds for $\delta \leq q^2$.*

we shall call $\delta+1$ the designed minimum distance of $C(n, \delta)$.

Example Consider the Hermitian curve $H(8)$ over $GF(64)$. Then $C(512, 10)$ is a $(512, 487, 11)$ code. A one point AG code on $H(8)$ of length 512 and dimension 487 has actual minimum distance 7 [6]. Moreover, from the same curve, Feng and Rao's improved geometric Goppa codes of length 512 and designed minimum distance 11 has dimension at most 484.

[1] This work was supported by the National Science Foundation under Grants NCR-9406043.

References

[1] M. Tsfasman and S. Vlăduţ and T. Zink, "Modular Curves, Shimura Curves and Goppa Codes, Better than Varshamove-Gilbert Bound," *Math. Nachrichten*, vol. 109, pp. 21-28, 1982.

[2] T. Yaghoobian and I. Blake, "Hermitian Codes as Generalized Reed-Solomon Codes," *Designs, Codes and Cryptography*, vol. 2, pp.5-7, 1992.

[3] G.L. Feng, V.K. Wei, T.R. Rao and K.K. Tzeng, "Simplified Understanding and Efficient Decoding of a Class of Algebraic-Geometric Codes," *IEEE Trans. Inform. Theory*, vol. 40, pp.981-1002, July 1994.

[4] G.L. Feng and T.R.N Rao, "A Simple Approach for Construction of Algebraic-Geometric Codes from Affine Plane Curves," *IEEE Trans. Inform. Theory*, vol. 40, pp.1003-1012, July 1994.

[5] G.L. Feng and T.R.N Rao, "Improved Geometric Goppa Codes, Part I: Basic Theory," preprint, 1994.

[6] K. Yang and P.V. Kumar, "On the true minimum distance of Hermitian codes," *Coding Theory and Algebraic Geometry-3*, Lect. Notes in Math. 1518, pp.99-107, 1991.

A fast parallel decoding algorithm for general one-point AG codes with a systolic array architecture

Masazumi Kurihara and Shojiro Sakata[1]

Dept. of Comp. Sci. & Info. Math., Univ. of Electro-Communications, Tokyo, 182, Japan

Abstract — **In this paper we propose a fast parallel decoding algorithm for general one-point algebraic geometric(1-pt AG) codes with a systolic array architecture(SAA). This algorithm is able to correct up to half the Feng-Rao bound and the time complexity is $O(n)$ by using a series of $O(n)$ processors where n is the code length and each processor is composed of τ cells for the smallest non-zero and non-gap value τ.**

Our decoding algorithm is a parallel version of the decoding algorithm given in [5], which is a special version of multi-dimensional Berlekamp-Massey(multi-D BM) algorithm. This algorithm is implemented with a SAA. In [6], we recently presented a parallel version of 1D BM algorithm with a SAA which can be applied to decoding of Reed-Solomon codes and BCH codes. In this paper we present a scheme which is motivated by the systolic algorithm[6]. To implement the parallel computation, we introduce a concept of a discrepancy polynomial having discrepancies as coefficients of its terms in the multi-D BM algorithm.

Let X be a curve with genus g over a finite field $\mathbf{F}$. $P_1, \ldots, P_n$ and Q are distinct $\mathbf{F}$-rational points on X. $D := P_1 + \cdots + P_n$ and $G := mQ$. $\mathcal{T}$ is the set of all non-gap values at Q and $\tau := \min\{t \in \mathcal{T} | t \neq 0\}$. For each $0 \leq i \leq \tau - 1$, $v_i := \min\{t \in \mathcal{T} | t \equiv i (\mathrm{mod} \tau)\}$ and $v_\tau := \tau$. $\{v_1, \ldots, v_\tau\}$ is the minimal set of generators for the semi-group $\mathcal{T}$ under addition. For $1 \leq i \leq \tau$, let ψ_i be a function in $L(\infty Q)$ with $-\mathrm{v}_Q(\psi_i) = v_i$ where $\mathrm{v}_Q(\cdot)$ denotes the valuation at Q. $\vec{v} := (v_1, \ldots, v_\tau)$, $\vec{e}_0 := (0, \ldots, 0)$, $\vec{e}_1 := (1, 0, \ldots, 0), \ldots, \vec{e}_\tau := (0, \ldots, 0, 1) \in \mathbf{Z}_+^\tau$ where $\mathbf{Z}_+$ is the set of all non-negative integers. $\Sigma := \mathbf{Z}_+^\tau$ and $\widetilde{\Sigma} := \{\vec{e}_i + k\vec{e}_\tau | k \in \mathbf{Z}_+, 0 \leq i \leq \tau - 1\}$. For any $\vec{p} \in \Sigma$, $\psi^{\vec{p}} = \psi_1^{p_1} \cdots \psi_\tau^{p_\tau} \in L(\infty Q)$ and $-\mathrm{v}_Q(\psi^{\vec{p}}) = \sum_{i=1}^{\tau} p_i v_i =: (\vec{p} \cdot \vec{v})$.

The general 1-pt AG code C of length n over $\mathbf{F}$ is defined as follows: For $\vec{c} \in \mathbf{F}^n$, $\vec{c} \in C$ iff $\sum_{j=1}^{n} c_j \psi^{\vec{p}}(P_j) = 0$ for all $\psi^{\vec{p}} \in L(mQ)$, i.e., all $\vec{p} \in \Sigma$ s.t. $(\vec{p} \cdot \vec{v}) \leq m$. d_{FR} denotes the Feng-Rao designed distance defined in [3]. Let $(e_j)_{1 \leq j \leq n}$ be an error vector. For all $\vec{p} \in \Sigma$, we define the syndrome as $S_{\vec{p}} := \sum_{\mu=1}^{\nu} e_{j_\mu} \phi^{\vec{p}}(P_{j_\mu})$ where $\nu \leq \lfloor (d_{\mathrm{FR}} - 1)/2 \rfloor$. All $S_{\vec{p}}$ are known for $(\vec{p} \cdot \vec{v}) \leq m$, but $S_{\vec{p}}$, $m+1 \leq (\vec{p} \cdot \vec{v})$, are unknown. To correct up to $\lfloor (d_{\mathrm{FR}} - 1)/2 \rfloor$ errors, we must find the values of unknown syndromes $S_{\vec{p}}$ s.t. $m+1 \leq (\vec{p} \cdot \vec{v}) \leq N := d_{\mathrm{FR}} + 3g - 2$ from [1]. Using the majority scheme[5], however, we can find the values of them.

For each $0 \leq i \leq \tau - 1$ and $\vec{s}_i \in \{\vec{e}_i + k\vec{e}_\tau | k \in \mathbf{Z}_+\}$, we introduce the following generator polynomial and its discrepancy polynomial. $f^{(i)}(x) := \sum_{\vec{k}} f_{\vec{k}}^{(i)} x^{\vec{k}}$ and $d^{(i)}(x) := \sum_{\vec{n}} d_{\vec{n}}^{(i)} x^{\vec{n}} \in \mathbf{F}[x]$ where $\vec{k} \in \widetilde{\Sigma}$ s.t. $(\vec{k} \cdot \vec{v}) \leq (\vec{s}_i \cdot \vec{v})$, $\vec{n} \in \vec{s}_i + \widetilde{\Sigma}$ s.t. $(\vec{n} \cdot \vec{v}) \leq N$, and $d_{\vec{n}}^{(i)} := \sum_{\vec{k}} f_{\vec{k}}^{(i)} S_{\vec{k}+\vec{n}-\vec{s}_i}$. Moreover, for the above i, $\vec{n}$ and $0 \leq j_i \leq \tau - 1$, we consider the auxiliary polynomials $g^{(j_i)}(x)$ and $e^{(j_i)}(x)$ with span $\vec{c}_{j_i}$ where $\vec{n} - \vec{s}_i - \vec{c}_{j_i} = k\vec{e}_\tau, |k| \in \mathbf{Z}_+$. We set their initial data as follows: For each $0 \leq i \leq \tau - 1$, $\vec{s}_i := \vec{e}_i$, $f^{(i)}(x) := x^{\vec{s}_i}$ and $d^{(i)}(x) := \sum_{\vec{n}} S_{\vec{n}} x^{\vec{n}}$, $\vec{c}_{j_i} := \vec{e}_{j_i} - \vec{e}_\tau$ and $g^{(j_i)}(x) = e^{(j_i)}(x) := \emptyset$.

We consider the following systolic array (see Fig. 1). The systolic array is composed of a series of N processors where each processor is composed of τ cells. In each cell 1D BM algorithm is practiced not per a polynomial but per a term of the polynomial. All processors receive/send the data from the left-neighboring processor/to the right-neighboring processor, synchronously. We call a unit of synchronized operations a beat where each beat is composed of a fixed small number of arithmetic operations over $\mathbf{F}$, which is assumed to take $O(1)$ time complexity. The number of beats necessary for executing our algorithm is at most $3N$. To correct up to $\lfloor (d_{\mathrm{FR}} - 1)/2 \rfloor$ errors, our algorithm achieves an optimal $O(n)$ computing time by using a series of $O(n)$ processors where we assume $O(n) = O(N)$. Each processor has $O(\tau)$ space complexity, and thus the total time and space complexity is $O(\tau n^2)$. In general, $\tau < n$, e.g. for codes from Hermitian curves, $O(\tau) = O(n^{1/3})$. Moreover, in [4], Kötter proposes a parallel Berlekamp-Massey type algorithm for Hermitian codes, which time complexity is $O(n^2)$ by using τ processors where each processor is composed of $O(n)$ registers. Thus, Kötter's total complexity is $O(\tau n^3)$.

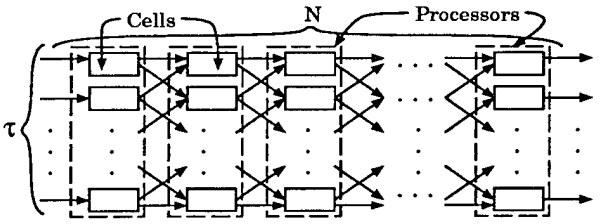


Fig. 1: A systolic array

[1]Email: kurihara@cs.uec.ac.jp and sakata@cs.uec.ac.jp

REFERENCES

[1] A.N.Skorobogatov and S.G.Vlăduţ : "On the decoding of algebraic-geometric codes," *IEEE Trans. IT*, vol.36, no.5, pp.1051-1060, 1990.

[2] G-L Feng and T.R.N. Rao, "Decoding algebraic-geometric codes up to the designed minimum distance," *IEEE Trans. IT*, vol.39, no.1, pp.37-45, 1993.

[3] C.Kirfel and R. Pellikaan, "The minimum distance of codes in an array coming form telescopic semigroups," presented at the Fourth Workshop an Arithmetic Geometry and Coding Theory, France, 1993.

[4] R.Kötter, "A fast parallel Berlekamp-Massey type algorithm for Hermitian codes," private communication, 1994.

[5] S. Sakata, H.E. Jensen, T. Høholdt, "Generalized Berlekamp-Massey decoding of algebraic geometric codes up to half the Feng-Rao bound", submitted for *IEEE Trans. IT*

[6] S.Sakata and M.Kurihara, "A parallel implementation of Berlekamp-Massey algorithm with a systolic architecture," (in Japanese) Proc. of SITA94 , pp.453-456, 1994.

[7] M.Kurihara and S.Sakata: "A fast parallel decoding algorithm for one-point AG-codes with a systolic architecture," Proc. of SITA94, pp.449-452, 1994.

Effective Construction of Self-Dual Geometric Goppa Codes

Gaétan Haché
INRIA, Domaine de Voluceau - BP 105, 78153 Le Chesnay Cedex, France
email: Gaetan.Hache@inria.fr

I. Introduction

The first criterion of self-duality for geometric Goppa codes has been given by Driencourt and Michon [1] for codes constructed from elliptic curves. More general criterions can be found in [2, 4, 5, 6, 8]. Our aim is to effectively construct self-dual geometric Goppa codes. One example is given at the end which was done using the implementation of the Brill-Noether algorithm written in AXIOM by the author (see [3]).

II. Self-Dual Geometric Goppa Codes and Class group

Denote by $\mathbb{F}_q$ the finite field of q elements. For $a = (a_1, a_2, \ldots, a_n) \in \mathbb{F}_q^n$ and $b = (b_1, b_2, \ldots, b_n) \in \mathbb{F}_q^n$ we have the *outer product* $a*b = (a_1b_1, a_2b_2, \ldots, a_nb_n) \in \mathbb{F}_q^n$. A linear code $C \subset \mathbb{F}_q^n$, n even, is said *quasi self-dual* if there exists a vector $a = (a_1, a_2, \ldots, a_n) \in \mathbb{F}_q^n$, $a_i \neq 0$, such that $a * C = C^\perp$. Note that if for each a_i there exists b_i such that $b_i^2 = a_i$ then the code $b * C$ is self-dual with $b = (b_1, b_2, \ldots, b_n)$. If $\mathbf{char}\mathbb{F}_q = 2$ then such b_i always exists.

For the rest of this abstract, F denotes an algebraic function field in one variable of genus g with full constant field $\mathbb{F}_q$. Denote by $\mathbb{P}_F$ the set of places of F and by $\mathcal{D}_F$ the set of divisors of F. The *class group* of F is the factor group $\mathcal{C}_F^0 := \mathcal{D}_F^0/\mathcal{P}_F$ where $\mathcal{D}_F^0$ is the subgroup of $\mathcal{D}_F$ consisting of all divisors of degree zero and $\mathcal{P}_F$ the subgroup of principal divisors. The group $\mathcal{C}_F^0$ is finite. Its order $h_F = h$ is called the *class number* of F (see [7, V.1.3]). To compute the class number, one can use the Zeta-function of an algebraic function field in one variable (see [7, V.1.15 and V.1.17]).

Proposition 1 *Let the divisor $D := P_1 + P_2 + \cdots + P_n$ be the sum of $n = 2k$ pairwise distinct places of F of degree 1. Assume that the class group $\mathcal{C}_F^0 := \mathcal{D}_F^0/\mathcal{P}_F$ is cyclic of order $h \neq 0(\mathbf{mod}2)$ and have a generator A with disjoint support from that of D. Assume moreover that there exists a divisor B of degree 1 with disjoint support from that of D. Then there exists an integer $m \in \{0, 1, \ldots (h-1)\}$ such that with the divisor*

$$G := (k + g - 1)B + mA$$

the geometric Goppa code

$$C_{\mathcal{L}}(D, G) := \left\{(f(P_1), f(P_2), \ldots, f(P_n)) \in \mathbb{F}_q^n \mid f \in \mathcal{L}(G)\right\}.$$

is quasi self-dual.

Proof: Since $h \neq 0(\mathbf{mod}2)$, the divisor $2A$ is also a generator of the class group. Hence there exists an integer $m_1 \in \{0, 1, \ldots (h-1)\}$ such that

$$D \equiv nB + 2m_1A = 2(kB + m_1A).$$

For the same reason there exists an integer $m_2 \in \{0, 1, \ldots (h-1)\}$ such that $(2g-2)B + 2m_2A$ is a canonical divisor. If we take $m \in \{0, 1, \ldots (h-1)\}$ with $m = m_1 + m_2 \ (\mathbf{mod}\ h)$ and set $G := (k + g - 1)B + mA$ we have

$$2G - D \equiv (2g-2)B + 2m_2A.$$

Hence $2G - D$ is a canonical divisor which implies that $C_{\mathcal{L}}(D, G)$ is quasi self-dual (see [9, Th. 3.1.46] or [6, Satz 1]). □

III. Example

Let F be the function field of the smooth plane quartic $\mathcal{X}$ defined by the following equation

$$X^3Z + X^2Y^2 + XY^3 + XZ^3 + Y^4 + YZ^3 + Z^4 = 0.$$

The genus of F is $g = 3$ and the class number over $\mathbb{F}_2$ is $h = 3$. Over $\mathbb{F}_2$, F has one place of degree 1, one place of degree 2 and 7 places of degree 4. Let P and Q be respectively the places of degree 1 and 2. The divisor $A := 2P - Q$ is non-principal, thus it is a generator of the class group $\mathcal{C}_F^0$. The intersection divisor of the curve $\mathcal{X}$ with any line is a canonical divisor (see [9, Prop. 2.2.7]). We take $K := 2P + Q$ as a canonical divisor which is the intersection divisor of the curve with the line $Z = 0$. Set the divisor $B := P$. Then $4B + 2A$ is equivalent to K. Among the seven places of degree 4 and considered as divisors, two are equivalent to $4B$, one is equivalent to $4B + A$, and the remaining four are equivalent to $4B + 2A$. Thus the sum of the seven places of degree 4, say D, is equivalent to $28B + 9A \equiv 28B$. Set $G := 16B + A$. Then $2G - D$ is a canonical divisor (see the proof of Proposition 1). Let $F_4 := \mathbb{F}_{2^4}F$. Let $D' := Con_{F_4/F}D$ and $G' := Con_{F_4/F}G$ (see [7, III.6.3 and V.1.9]). Then $C_{\mathcal{L}}(D', G')$ is a quasi self-dual $[28, 14, d \geq 12]$ code over $\mathbb{F}_{2^4}$. In fact $d = 12$ since by computing the generator matrix of the code we found a word of weight 12.

References

[1] Y. Driencourt and J.-F. Michon. "Remarques sur les codes géométriques," *C.R. Acad. Sc. Paris*, 301 Serie I(1):15-17, 1985.

[2] Driencourt Y. and H. Stichtenoth, "A criterion for self-duality of geometric codes", *Com. in Algebra*, 17(4):885-898, 1989.

[3] G. Haché and D. Le Brigand, "Effective construction of aglgebraic geometry codes," *Technical Report*, no. 2267, INRIA, May 1994.

[4] C. Munuera and R. Pellikaan, "Equality of geometric Goppa codes and equivalence of divisors," *Journal of Pure and Applied Algebra*, pages 229-252, 1993.

[5] C. Munuera and R. Pellikaan, "Self-dual and decomposable geometric Goppa codes", *EUROCODE 92*, 1993.

[6] W. Scharlau, "Selbstdual Goppa Codes", *Math. Nachr.*, 143:119-122, 1989.

[7] H. Stichtenoth, "Algebraic function fields and codes", *University Text*, Springer-Verlag, 1993.

[8] H. Stichtenoth, "Self-dual Goppa dodes", *Journal of Pure and Applied Algebra*, 55:199-211, 1988.

[9] M.Tsfasman and S. Vladut, "Algebraic-geometric codes", *Kluwer Academic Pub., Math. and its Appl.*, vol. 58, 1991.

On Codes Containing Hermitian Codes

Richard E. Blahut
University of Illinois, Urbana, IL 61801

Abstract — **It is shown that certain syndromes of a Hermitian code are not needed for decoding. These syndromes can be replaced by data symbols thereby increasing the dimension of the code without changing the designed minimum distance.**

I. Hermitian Codes and Hyperbolic Codes

Hermitian codes and hyperbolic codes are defined on the affine plane $GF(q)^2$. A hyperbolic code is defined for any q and is a two-dimensional cyclic code. A Hermitian code is defined for q an even power of two; it can be viewed as a shortened two-dimensional cyclic code. Whereas a hyperbolic code is defined on the full affine plane $GF(q)^2$, a Hermitian code is defined on a curve in the affine plane $GF(q)^2$. Using only the affine plane is so that the discussion can be organized around the formalism of the two-dimensional Fourier transform

$$C_{j'j''} = \sum_{i=0}^{n} \omega^{i'j'} \omega^{i''j''} c_{i'i''}.$$

The Hermitian polynomial

$$G(X,Y) = X^m - Y^{m-1} - Y$$

has $n = (m-1)^3 - (m-1)$ zeros in the affine plane over $GF((2^m)^2)$ of the form (γ, β) with γ and β both nonzero. These zeros of $G(X,Y)$ are used to define a code over $GF((2^{m-1})^2)$ with blocklength n and dimension $k = mJ - g + 1$ if $J \geq m$, and designed distance $d^* = n - k - g + 1$ where $g = \binom{m-1}{2}$.

A codeword is a vector $\mathbf{c}$ with components c_i for $i = 0, \ldots, n-1$, where i indexes the n points (i', i'') at which $G(\omega^{-i'}, \omega^{-i''}) = 0$. The spectrum $\mathbf{C}$ is required to satisfy

$$C_{j'j''} = 0 \text{ if } j' + j'' \leq J.$$

There are $\frac{1}{2}(J+1)(J+2)$ such (j', j''). The code is the set of such codewords.

II An Enlarged Code

We now enlarge the code to a new linear code that contains the Hermitian code. As in the previous section, a codeword is a vector with components c_i for $i = 0, \ldots, n-1$ where i indexes the n points (i', i'') at which $G(\omega^{-i'}, \omega^{-i''}) = 0$. For the enlarged code, the codewords satisfy

$$C_{j'j''} = 0 \text{ if } j' + j'' \leq J \text{ and } (j'+1)(j''+1) \leq d^* = n - k - g + 1.$$

Otherwise $C_{j'j''}$ is arbitrary. If the set $\{(j', j'') \mid (j'+1)(j''+1) \leq d^*\}$ is not contained in the set $\{(j', j'') \mid j' + j'' \leq J\}$, there will be fewer elements in the intersection than in the second set. Because there are fewer such (j', j'') than before, the constraints are weaker. Then there will be more codewords satisfying the new constraint so the dimension of the code is larger.

Syndrome $S_{j'j''}$ will be known only if $j' + j'' \leq J$ and $(j'+1)(j''+1) \leq d^*$. It follows from the two-dimensional form of Massey's theorem that each unknown syndrome can be inferred by a subsidiary calculation in the Sakata algorithm just at the time that it is needed. This uses an argument of Saints and Heegard in the case that $(j'+1)(j''+1) \leq d^*$, and uses an argument of Sakata et al. (based on the ideas of Feng and Rao) in the case that $j' + j'' \leq J$. Because the unknown syndromes that result from the new hyperbolic constraint can be inferred by the decoder there is no reduction in the designed distance. (Apparently the performance of this code cannot be found by the usual methods of algebraic geometry.) Feng and Rao showed that the Hermitian code has true minimum distance larger than its designed distance. We are probably taking up the same slack in another way, increasing the dimension by reducing the true minimum distance.

III. Syndrome Filling

The Sakata algorithm is a generalization of the Berlekamp-Massey algorithm to two dimensions, processing the two-dimensional syndromes in some fixed total order. The graded order works best for our purposes. The locator polynomial update rule is based on a two-dimensional version of Massey's theorem. At each iteration one or more discrepancies are computed using the current error-locator ideal. If one or more discrepancies are nonzero, then Massey's theorem describes how the size of the error-locator ideal must increase.

It follows from Massey's theorem that certain syndromes cannot be generated wrong by the Sakata recursion; otherwise the error-locator ideal would grow too large.

IV. Example

An example shows that the class of codes defined contains more than the usual Hermitian codes. We simply display one code in which the set $\{(j', j'') | (j'+1)(j''+1) \leq d^*\}$ is not contained in the set $\{(j', j'') | j' + j'' \leq J\}$.

We choose the Hermitian code over $GF(256)$, so $m = 17$ and $d^* = (4080 - 17J)$. Choose $J = 130$, then $d^* = 1870$. Then consider $(j', j'') = (17, 113)$. Next, observe that $(17, 113) \in \{(j', j'') | (j' + j'') \leq 130\}$. However $(j'+1)(j''+1) = 2358$. Therefore

$$(17, 113) \notin \{(j', j'') | (j'+1)(j''+1) \leq d^*\}.$$

This means that syndrome $S_{17,113}$ is not needed by the two-dimensional Berlekamp-Massey algorithm. Hence $C_{11,113}$ is made into a data component, thereby enlarging the code. In particular, the enlarged code has larger dimension with the same designed distance.

References

[1] S. Sakata, J. Justesen, Y. Madelung, H. Elbroend Jensen, and T. Hoeholdt. "Fast Decoding of AG Codes up to the Minimum Distance," The Technical University of Denmark, 1993.

[2] K. Saints and C. Heegard. "On Hyperbolic Cascaded Reed-Solomon Codes," San Juan, Puerto Rico, 1993.

[3] G.-L. Feng and T.R.M. Rao. "Decoding Algebraic-Geometric Codes up to the Designed Minimum Distance," *IEEE Transactions on Information Theory*, vol. IT-39, pp.37-45, 1993.

[4] R.E. Blahut. *Lectures in Information Theory*, EE162 Class Notes, Swiss Federal Institute of Technology, Spring 1993 (edited copy in preparation).

Multilevel Codes Based on Matrix Completion

Dariush Dabiri and Ian F. Blake
Dept. of Elect.& Computer Eng., Univ. of Waterloo, Waterloo, Ont. N2L 3G1 Canada

Abstract — **Based on matrix completion algorithms, new constructions for algebraic multilevel codes are given. The constructions have low computational complexity and can be used for channels with combinations of burst and random errors.**

I. Introduction

Combinations of random and bursty errors usually occur in many communication and storage systems. For example, consider a conventional concantenated coding system which is operating on a communication channel with a combination of random and bursty errors. The error process at the output of the convolutional decoder tends to be bursty, and depending on the channel output, the length of the bursts varies. Assuming that interleaving has been used, the output of the de-interleaver will produce errors of short burst lengths. In an ideal situation, where the interleaving depth is enough to remove all of the bursts, the outer code may view the error process at the output of the de-interleaver as a purely random error process. However, each codeword in an interleaving frame will have a different share in the number of errors produced by bursts of short lengths. Therefore, in each frame, the decoder for the outer code may fail to decode some blocks which are suffering from a large number of errors, while it is still capable of decoding the other blocks in the frame. One may benefit from re-decoding the inner convolutional code as a determinate state convolutional code, where the Viterbi decoder is re-initialized by known bits periodically [2]. In this way, the side information provided by successfully decoded blocks can be used to improve on the error correction capability of the inner convoutional code. The matrix completion approach presented in this paper can be used as a complementray rather than competing technique, since they can be used at the same time.

In the matrix completion technique, syndrome information for the algebraic outer codes are not provided explicitly. At the first level of decoding, each frame is viewed as a single codeword. However, at this levlel, no attempt will be made to find the error locations and the error values for each block. Instead, some syndrome information will be computed for each block in the frame. This crucial step in the decoding process, is achieved by a matrix completion algorithm which is similar to the Feng-Rao algorithm [1]. Now each block may be viewed as a member of an algebraic code. For some blocks, the computed syndrome information will be enough to remove all of the errors. For others, the combination of the known syndrome information and the determinate state Viterbi decoder is used to enhance the performance of the purely algebraic decoding algorithm. The efficiency of this scheme comes from the following facts:

- The success of the completion algorithm for the reconstruction of the syndrome information depends on the over all number of errors in the frame.
- Even if the number of errors in one block is beyond the error correction capability of the algebraic multilevel code, still one might be able to re-construct the syndrome information.
- The complexity of the completion algorithm depends mainly on the number of the blocks and the number of the syndromes which are to be re-constructed. The complexity grows only linearly with the size of the blocks. Therefore, the over all complexity of the decoding is much less than the complexity of decoding a large code of the same length.

II. Construction of the Multilevel Codes

In the multilevel coding architecture, each frame

$$\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N)$$

consists of N blocks. Each block '$\mathbf{c}_i$', $i = 1, 2, \ldots N$, is a vector of length n_1 over $GF(q)$. Assuming that '$\mathbf{c}$' is transmitted and the word

$$\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

is received, the error pattern

$$\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N)$$

is defined by $\mathbf{e}_i = \mathbf{r}_i - \mathbf{c}_i$. Let α be an element of the extension field $GF(q^s)$ such that the order of Γ, the cyclic group generated by α, is n_1. The parameter s is the smallest integer such that $\alpha \in GF(q^s)$. The one dimensional syndrome $s(\mathbf{e}_i, \gamma)$ for any $\gamma \in \Gamma$ is defined as

$$s(\mathbf{e}_i, \gamma) = \sum_{j=0}^{n_1-1} e_{i,j}\gamma^j.$$

A linear multilevel code is defined to be the vector space of codewords, $\{\mathbf{c}\}$, such that

$$\mathbf{s}(\mathbf{c}, \gamma) = (s(\mathbf{c}_1, \gamma), s(\mathbf{c}_2, \gamma), \ldots, s(\mathbf{c}_N, \gamma))$$

falls in some specific subspaces of $GF(q^s)^N$.

Here, we use a matrix completion algorithm to reconstruct $\mathbf{s}(\mathbf{c}, \gamma)$ for some specific values of γ.

References

[1] G. L. Feng and T. R. N. Rao, "Decoding algebraic-geometric codes up to the design minimum distance", *IEEE Trans. Inform. Theory.* vol. IT-38, 37-45, 1993.

[2] O. M. Collins, "Determinate State Convolutional Codes" *IEEE Trans. Comm.* vol. COM-41, 1785-1794, 1993.

Rate Distortion Functions and Effective Bandwidth of Queueing Processes

C. S. Chang[1] and J. A. Thomas[2]

[1]Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30043, Taiwan.
[2]IBM T.J. Watson Research Center, P.O.Box 704, Yorktown Heights, NY 10598.

Abstract — **Parallel to the definition of the rate distortion function for source coding, we define a rate distortion function for delay in a queueing system which gives the tradeoff between the capacity of the server and the delay or buffer overflow incurred. This function is decreasing and convex and it is shown to be equal to the "effective bandwidth" of the input source for exponentially vanishing buffer overflow probability.**

Recent work on the information theoretic capacity of a queue[1] and on theory of "effective bandwidth"[2] have prompted us to take a new look at the capacity-delay tradeoff in a single server discrete time queue and view this as analogous to a rate distortion function for the source. Conventional source coding theory would allow perfect reproduction of the source at any service rate greater than the average rate of the input packets; however, this would incur arbitrarily long delays. The tradeoff between the rate and the delay is difficult to calculate in general, but in the case when the distortion measure is the buffer overflow probability, then we show that the limiting value of the rate distortion function is the effective bandwidth of the source.

We consider a discrete time slotted queue, where the input and the output processes consist of a sequence of packets (e.g. ATM packets). Consider two (discrete-time) point processes $a_i = \{a_i(t), t = 1, 2, \ldots\}$, $i = 1$ and 2. Let $A_i(0,t)$, $i = 1$ and 2, be the number of arrivals in the interval $(0,t]$ of the point process a_i and $\tau_i(n)$, $i = 1$ and 2, be the arrival epoch of the n^{th} customer of the point process a_i. We define a class of distance functions between two point processes as follows:

$$\rho^q(a_1, a_2) \triangleq \limsup_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} Ef(A_1(0,t), A_2(0,t)) \quad (1)$$

for some function $f : \mathcal{R}^2 \mapsto \mathcal{R}$. A similar distortion measure can be defined using arrival instants. For instance, if a_1 is an arrival process to a queue and a_2 is the corresponding departure process, then $\rho^q(a_1, a_2)$ is the average queue length and $\rho^d(a_1, a_2)$ is the average delay when $f(z_1, z_2) = z_2 - z_1$.

We consider a queue with time varying capacity $c(t)$[3] where $c(t)$ is the maximum number of packets served in time slot t. Then the behaviour of the queue is governed by the following recursive equation: $q(t+1) = (q(t)+a(t+1)-c(t))^+$. We will call the sequence $\{c(t), t \geq 0\}$ an independent bandwidth allocation sequence if $\{c(t), t \geq 0\}$ is independent of the arrival process. Let $\mathcal{F}_c$ be the family of independent bandwidth allocation sequences that is stationary and ergodic with mean c. Also, let a (b) denote the arrival (departure) process. For the class of distance functions $\rho^q(a,b)$, we say that a distortion D is achievable at rate c if there is an independent bandwidth allocation sequence $\{c(t), t \geq 1\} \in \mathcal{F}_c$ such that $\rho^q(a,b) \leq D$. The rate distortion function $R^q(D)$ is defined to be the *minimum rate* c that the distortion D is achieved, i.e., $R^q(D) \triangleq \inf\{c : \rho^q(a,b) \leq D\}$. We define $R^d(D)$ similarly. One can show that the rate distortion functions $R^q(D)$ and $R^d(D)$ are both decreasing and convex in D.

In general, the calculation of $R^q(D)$ and $R^d(D)$ are difficult problems. However, motivated by the case of ATM networks with exponentially small buffer overflow probability, we consider the distance function

$$\rho_x^q(a,b) \triangleq \limsup_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} E1_{\{A(0,t)-B(0,t)\geq x\}}, \quad (2)$$

and let $R_x^q(D)$ denote the corresponding rate distortion function Note that $q(t) = A(0,t) - B(0,t)$ when $q(0) = 0$. If $a(t)$ is stationary and ergodic with mean $Ea(t) < c$, then $\rho_x^q(a,b) = \Pr(q(\infty) \geq x)$. In this case, we have the following asymptotic result, based on the theory of effective bandwidth[3]:

Theorem 1 *If the arrival process $a(t)$ is stationary and ergodic and it satisfies $\lim_{t\to\infty} \frac{1}{t} \log Ee^{\theta A(0,t)} = \Lambda(\theta)$, for all $\theta \geq 0$, then*

$$\lim_{x\to\infty} R_x^q(e^{-\theta x}) = a^*(\theta), \quad (3)$$

where $a^*(\theta)$ is called the effective bandwidth of the source, and is defined to be $a^*(\theta) = \frac{\Lambda(\theta)}{\theta}$

The concept of effective bandwidth has attracted considerable interest lately, and a number of papers develop a calculus of effective bandwidth that allows one to analyze superpositions of sources, outputs of queues, and networks of queues. The current work provides an new interpretation of some of these results, and provides a link between rate distortion theory and queueing theory.

References

[1] Ananthram, V. and Verdu, S., "Bits through queues", Proceedings of the IEEE International Symposium on Information Theory, Trondheim, Norway, 1994.

[2] C.S. Chang, "Stability, queue length and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, Vol.39, pp. 913-931, 1994.

[3] C.S. Chang and T. Zajic, "Effective bandwidths of departure processes from queues with time varying capacities," *IEEE INFOCOM'95*, Boston, pp. 1001-1009.

LARGE BURSTS DON'T CAUSE INSTABILITY

BRUCE HAJEK

Coordinated Science Laboratory and the
Department of Electrical and Computer Engineering
University of Illinois, Urbana, Illinois 61801, USA

Abstract – **It is shown that if a queueing network is stable with fluid arrival processes, then it is also stable for deterministically constrained bursty arrival processes of the same or smaller long-term rate.**

SUMMARY

Fluid models of queuing networks are among the simplest models to analyze, owing to the fact that calculus can be applied. At the same time, wider classes of network models are more flexible for modeling real traffic. It is thus useful to reduce questions about the more realistic models to questions about related fluid models. Such a reduction was recently achieved by J.G. Dai, who showed that stability of a fluid model implies stability (in the sense of Harris recurrence) of related multiclass networks with random service and interarrival processes of renewal type. The purpose here is to similarly reduce the question of stability for networks with input traffic satisfying *deterministic constraints* in the sense of Cruz (*IEEE IT Tranasactions*, January 1991) to a question of stability for a fluid model.

The network has d single server stations and K classes of traffic. Class l traffic is served at a unique station $s(l)$. Let C be the $d \times K$ matrix such that $C_{i,l} = 1$ if $s(l) = i$ and $C_{i,l} = 0$ otherwise. Upon completion of service at $s(l)$ the traffic of class l either becomes traffic of class l' for some other class l', in which case we write $l \to l'$, or it immediately exits the network. Let P denote the $K \times K$ matrix such that $p_{l,l'} = 1$ if $l \to l'$ and $p_{l,l'} = 0$ otherwise. A simple network with three stations and eight classes is shown in Figure 1 with $1 \to 2 \to 3$, $4 \to 5 \to 6$ and $7 \to 8$ and $s = (1, 2, 3, 2, 3, 1, 3, 2)$. It is assumed that the network is open, so that P^K is the zero matrix.

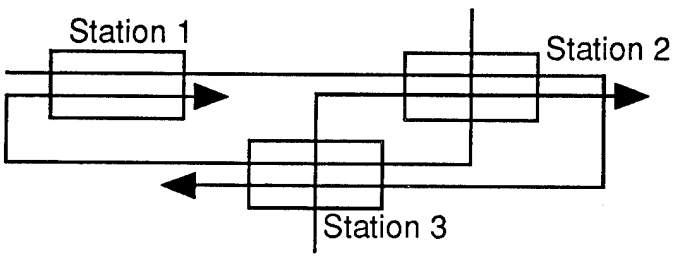


Figure 1: Sample network.

Exogenous traffic can enter the network as any class, though for the example given it might make sense for the exogenous arrival functions to be nonzero only for classes 1, 4 and 7. Let $E_l(t)$ denote the amount of exogenous class l traffic to enter the network during $[0, t]$. Traffic of class l can be served (at station $s(k)$) at a maximum rate $\mu_l = 1/m_l$ where $m_l > 0$. Let $M = diag(m_1, \ldots, m_K)$ and let e denote a column vector of all ones (with dimension depending on the context). The flow of traffic in the network is assumed to satisfy the following equations and conditions:

$$Q(t) = q + E(t) + (P^T - I)M^{-1}T(t) \quad (1)$$

$$I(t) = et - CT(t) \quad (2)$$

$$Q(t) \geq 0 \quad \text{for } t \geq 0 \quad (3)$$

$$\int_0^\infty (CQ(t) \wedge e)\, dI(t) = 0 \quad (4)$$

$$T(0) = 0, \quad T \text{ is right-continuous and nondecreasing} \quad (5)$$

$$I(0) = 0, \quad I \text{ is right-continuous and nondecreasing.} \quad (6)$$

The the following interpretations hold:

$Q_l(t)$ is the amount of class l traffic in the network at time t, and $Q_l(0) = q_0$.

$T_l(t)$ is the amount of work (where work is measured in units of time) done on class l traffic during $[0, t]$.

$(CT(t))_i$ is the amount of work done at station i during $[0, t]$.

$I_i(t)$ is the amount of idleness (measured in units of time) of the server at station i accumulated during $[0, t]$

$(CQ(t))_i$ is the amount of traffic at station i at time t.

The exogeneous traffic E is said to satisfy deterministic constraints with parameters $\alpha = (\alpha_l)$ and $\sigma = (\sigma_l)$, abbreviated to "E is DC(α, σ) traffic", if

$$0 \leq E_l(t) - E_l(s) \leq \alpha_l(t - s) + \sigma_l \quad 0 \leq t \leq s < \infty. \quad (7)$$

The network (C, P, m) is *totally stable* for DC(α, σ) traffic if there is a finite constant Γ so that whenever (E, q, Q, T, I) satisfy (1)–(7), then $\limsup_{t\to\infty} |Q(t)| \leq \Gamma$, where $|Q(t)| = \sum_l |Q_l(t)|$.

Theorem 1 *If the network (C, P, m) is totally stable for fluid traffic with input rate vector α, then it is totally stable for DC(α, σ) traffic, for any vector σ.*

ACKNOWLEDGEMENT

This work was supported by National Science Foundation Contract NCR 93-14253.

The Extension of Optimality of Threshold Policies in Queueing Systems with Two Heterogeneous Constant-Rate Servers

A. Traganitis
University of Crete
Heraklion, Greece

A. Ephremides
EE Dept. & ISR, U. of Maryland
College Park, MD

Abstract - We extend the threshold condition for optimality of the policy for activating the slow server in a 2-server queueing system when the service times are deterministic.

I. Introduction

We consider a queueing system composed of an infinite-size buffer and two servers S_1 and S_2 with constant, but different, service times T_1 and T_2, respectively, with $T_2 > T_1$. The arrival process is Poisson with rate $\lambda < \frac{1}{T_1} + \frac{1}{T_2}$. We wish to find the optimal policy for server activation that minimizes the average customer sojourn time in the system.

II. Background

This is a problem that has been well-studied for the case of exponential servers [1-3]. The optimal policy has been shown to consist of always keeping the fast server busy, as long as the queue is non-empty, and of activating the slow server when the queue size is greater than a threshold m_o, the value of which depends on λ and on the service rates. Extension of the result in more complicated systems has been, in general, difficult. The motivation for considering the deterministic case is that in packet-switched systems the transmission time of each packet is constant so long as the transmission bandwidth remains constant.

III. Approach

Let $x_o(t)$ denote the queue size at time t and $r_i(t)$ the residual service time of server $S_i, i = 1, 2$; clearly, $0 \leq r_i(t) \leq T_i$. The vector $x(t) = (x_o(t), r_1(t), r_2(t))$ is a Markovian state description of the system. We let $x_i(t)$ be 0, if $r_i(t) = 0$, and 1 otherwise. The total number of customers is then given by $|x(t)| = x_o(t) + x_1(t) + x_2(t)$. Let π be a control policy that decides at every $t \geq 0$ which idle server to activate based on $\{x(s), 0 \leq s \leq t\}$. The policy π is optimal if it minimizes the long-run average cost $J_\pi(x)$, where $J_\pi(x) = \limsup \frac{1}{T} E_x^\pi [\int_0^T |x(t)| dt]$ where x is the initial state.

Markov Decision Theory cannot be used easily to establish optimality conditions here because of the continuity of the transitions in the residual service times. However, we use the special features of the deterministic service times to obtain necessary conditions for the optimal Markovian policy that coincide with the properties of the optimal policy of the exponential service case.

Furthermore, we obtain lower and upper bounds to the threshold value for the class of threshold policies.

IV. Results

The optimal policy is shown to (i) activate at least one of the servers without delay if they are both idle, (ii) activate the fast server immediately if it is idle and the slow server is busy, and (iii) activate the fast server before the slow server if both are idle. Furthermore, the optimal Markov policy that activates the slow server when the system is in states $(y, r_1, 0), (z, r_1, 0)$ for $y < z$, must also activate the slow server for any state $(x, r_1, 0)$, for $y < x < z$.

The lower bound to the threshold value is given by $1 - \frac{r_1}{T_1} + (\frac{T_2}{T_1} - \frac{r_1}{T_1} - 1)(1 - \lambda T_1)$, and the upper bound by $1 + \frac{T_2 - r_1}{T_1}$. The method for computing the thresholds applies to the exponential case as well and the results are consistent with the exact threshold calculations in [3].

References

1. W. Lin, P.R. Kumar, "Optimal Control of a Queueing System with Two Heterogeneous Servers," IEEE Trans. on AC, Aug. 84, pp. 696-703.
2. J. Walrand, "On Optimal Control of a Queueing System with Two Heterogeneous Servers," System & Control Letters, May 84, pp. 131-134.
3. M. Rubinowich, "The Slow Server Problem: A Queue with Stalling," J. Applied Probability, Vol. 22, pp. 879-892, 1985.

Elimination of Bistability in Spread-Spectrum Multiple-Access Networks

Ramaswamy Murali and Brian L. Hughes[1]
Dept. Elect. & Comp. Engr., Johns Hopkins Univ.,
Baltimore, Maryland 21218, USA

Abstract — **This work considers packet radio networks in which users transmit using spread-spectrum multiple-access (SSMA) signaling, slotted ALOHA random access, and forward-error-correction (FEC). It is well known that these networks can exhibit bistable behavior similar to narrowband ALOHA systems. In this work, we analyze the impact of FEC parameters on the throughput, delay, and drift of slow frequency-hop (FH)/SSMA networks. We present exact expressions for throughput, delay, and drift, and, furthermore, characterize bistable systems by their first exit times (FET). Drift analysis suggests an approach to eliminating bistability that involves increasing both the average retransmission time and the blocklength. Numerical examples are provided to illustrate our approach. A noteworthy feature of our approach is that the elimination of bistability is achieved by careful selection of system parameters without using active control.**

I. Introduction

Consider a network in which a population of N transmitters (or *users*) and N receivers share a common radio channel. The network topology is assumed to be *fully-connected* with *paired-off* transmissions; each transmitter communicates with a single, unique receiver. Each user is fed by a bursty message source. Users transmit messages in the form of packets using spread-spectrum multiple-access (SSMA) signaling, slotted ALOHA random access, and forward-error-correction (FEC). We assume that feedback is present and that the feedback propagation delay is negligible in comparison to the packet transmission time.

It is well known that such networks can exhibit *bistable* behavior similar to narrowband ALOHA systems [1]. A bistable system possesses two locally stable equilibria with the system achieving high throughput and small delay at one (*operating point*) and low throughput and large delay at the other (*saturation point*). In practice, a bistable network can remain in saturation for large time periods, thus leading to poor performance. Consequently, the elimination of bistability, by which we mean removing the saturation point while retaining the throughput-delay performance at the operating point, is of importance in practical networks.

Various *stabilization* techniques, which aim to prevent the network from reaching saturation, have been examined in the literature. Notably, these techniques rely either on recursive retransmission control (e.g., [2]), wherein an estimate of the total number of backlogged users in the network is employed by each user to alter a design parameter, such as the retransmission probability, or code rate, or else on centralized control of user transmissions (e.g., [3]).

Our focus, in this work, is on investigating the impact of FEC code parameters on the throughput, delay, and drift of SSMA networks. For concreteness, we consider slow frequency-hop (FH)/SSMA signaling with Reed-Solomon erasure correction [4]. We present a model which is exact for *finite* user populations and expressions for throughput, delay, and drift. Moreover, we characterize bistable systems by their *first exit times* (FET), which is a measure of the average time taken by a bistable network to reach saturation, starting from zero user backlog. Also, a simpler limiting model is presented which may be used when both the number of users and the number of frequency bins are large. We then present our approach to eliminating bistability based on a drift analysis of the limiting model. Finally, numerical examples are provided to illustrate our approach.

II. Conclusions

The following four observations apply to a bistable FH/SSMA network:

(1) Increasing the code blocklength leads to higher throughput and lower delay at the operating point at the cost of smaller FET.

(2) Increasing the average time to retransmission leads to larger FET at the cost of lower throughput and higher delay at the operating point.

(3) Elimination of bistability necessitates increasing both the blocklength and the average time to retransmission.

(4) At fixed blocklength, further improvement in operating point performance can be achieved by optimizing over code rate.

From (3), we infer that it is possible to eliminate bistability by careful selection of network design parameters *without* the use of *active* (decentralized or centralized) control.

Acknowledgements

The authors wish to thank Mr. Syed R. Reza for his help in obtaining the numerical results.

References

[1] A. Polydoros and J. Silvester, "Slotted random access spread-spectrum networks: An analytical framework," *IEEE J. Selected Areas Commun.*, vol.6, pp. 989-1002, July 1987.

[2] B. Hajek, "Recursive retransmission control - application to a frequency-hopped spread-spectrum system," *Proc. 1982 Conf. Inform. Sci. Syst.*, Princeton, NJ, pp. 116-120, March 1982.

[3] A. Papasakellariou and B. Aazhang, "Stabilization and performance analysis of CDMA communication systems," *Proc. 1992 Conf. Inform. Sci. Syst.*, Princeton, NJ, March 1992.

[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.

[1]This work was supported by the National Science Foundation under grant NCR-9217457

Transient Analysis of Media Access Protocols by Diffusion Approximation

Qiang Ren
NEC USA, Inc.
4 Independence Way
Princeton, NJ08540, USA

Hisashi Kobayashi
Department of Electrical Engineering
Princeton University
Princeton, NJ08544, USA

Summary

In our earlier work [1, 2] we applied the diffusion approximation method to the steady state analysis of an ALOHA random access protocol, and non-Markovian queueing networks. More recently, we have successfully applied the diffusion model in analyzing the transient behavior of a statistical multiplexer, and in deriving a simple formula for the effective bandwidth of a bursty traffic source in an ATM (asynchronous transfer mode) network [3].

In this paper, we present a transient analysis of media access control (MAC) protocols such as slotted ALOHA and CSMA/CD (Ethernet) by formulating their queue behavior as an Ornstein-Uhlenbeck process $X(t)$. We also derive an important result on the transient mean $m_X(t) = E[X(t)]$: if the drift coefficient $\beta(x,t)$ is a linear function of the system congestion $x = X(t)$, then $m_X(t)$ that we can obtain under the assumption of homogeneous diffusion coefficient is an unbiased estimate of $m_N(t)$, the mean of the original process $N(t)$, and is independent of the diffusion coefficient $\alpha(x,t)$.

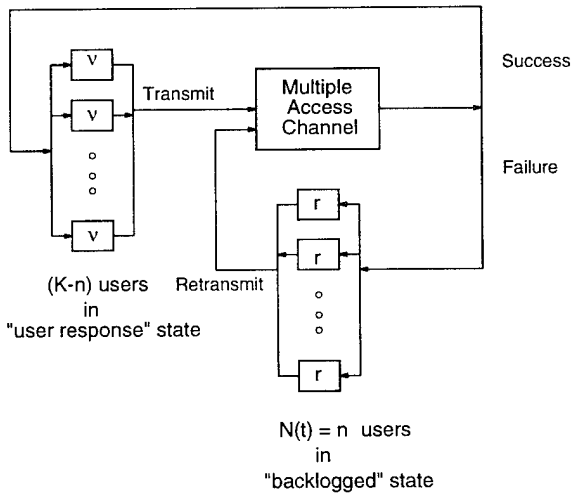


Figure 1: The number of backlogged users, $N(t)$, in the slotted ALOHA system

Figure 1 is a queueing model representation of the slotted ALOHA system being considered. There are K users in the system. $N(t)$ is the number of the "backlogged" users who are either engaged in actual transmission or waiting for a retransmission at the t-th slot time. The parameter r is the probability that a retransmission will take place at a given time slot. The remaining $K - N(t)$ users are in the "user response" state, ready to generate a new packet with probability ν in a given slot.

We approximate the integer-valued process $N(t)$ by a continuous-state diffusion process $X(t)$, which has the *drift* coefficient $\beta(x,t) = (K-x(t))\nu - S(x(t))$, and the *diffusion* coefficient $\alpha(x,t) = (K - x(t))\nu + S(x(t))$, where $S(x(t))$ represents the channel throughput when the process $X(t)$ is in state x at time t.

For analytical tractability, we simplify $\beta(x,t)$ and $\alpha(x,t)$ as follows: Suppose that $S(x)$ (hence $\beta(x,t)$, as well) can be approximated by a linear function of $x(t)$, i.e., $\beta(x,t) = \beta_0 - \beta_1 x(t)$; and that $\alpha(x,t)$ is nearly constant around some value $\overline{x}$, i.e., $\alpha(x,t) \approx \alpha_0$. Then the diffusion process $X(t)$ becomes an Ornstein-Uhlenbeck process, and we can readily obtain the time-dependent solution $f(x,t|x_0)$: it is a Gauss-Markov process with time-dependent mean $m_X(t) = \overline{x}(1 - e^{-\beta_1 t}) + x_0 e^{-\beta_1 t}$ and variance $\sigma_X^2(t) = \frac{\alpha_0}{2\beta_1}(1 - e^{-2\beta_1 t})$.

References

[1] Kobayashi, H., Y. Onozato and D. Huynh (1977), "An Approximation Method for Design and Analysis of an ALOHA System", *IEEE Trans. on Communications*, **COM-25**(1), pp.148-157. 1992, pp.262-271.

[2] Kobayashi, H. (1983), "Diffusion Approximations in Queue Analysis". In Louchard, G. and G. Latouche (eds.), *Probability Theory and Computer Science*, pp. 86-98, Academic Press.

[3] Kobayashi, H. and Q. Ren (1993), "A Diffusion Approximation Analysis of an ATM Statistical Multiplexer with Multiple Types of Traffic, Part II: Time-Dependent Solutions", *Proc. the 1993 IEEE Global Telecommunications Conference*, Vol.2, pp.772-777, Houston, Texas.

Multi-Media Multi-Rate CDMA Communications

Y.-W Chang, S. Yao, and E. Geraniotis
Dept. of Electrical Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742

Abstract — **In this paper we extend further the results of [1] and [2] to CDMA networks with truly multi-rate multi-media traffic. Voice, high and low priority data, and possibly videophone traffic all use DS/CDMA modulation but different information (data) rates.**

I. Introduction

In our recent work of [1] we introduced a Markovian formulation for the problem of optimal admission of voice users in a CDMA network; there was also data traffic in the CDMA network but it had lower priority over voice and was allowed to use the CDMA codes left unused by the voice users. In [2] we extended this work to CDMA networks with voice and multi-priority data; there are two classes of data users, those with high priority (same as voice) that require real-time delivery (but lower BER than voice) and those with low priority that are treated as in [1]. In the work of both [1] and [2] voice and data users transmit at the same data (information) rate and employ the same processing gain (number of chips per bit) in their DS/CDMA modulation. Finally, in our work of [3] we provided a preliminary analysis of a CDMA system supporting true multi-rate multi-media traffic. In [3] multiple chip rates are used to spread the user signals in proportion to their information data rates rather than spreading them over the entire frequency band. In this paper we extend further the results of [1] and [2] to CDMA networks with truly multi-rate multi-media traffic. Voice, high and low priority data, and possibly videophone traffic all use DS/CDMA modulation but different information (data) rates. Since the signals of all users are spread over the entire bandwidth, the different data rates result in different processing gains.

II. Multi-Rate CDMA System

The different processing gains affect the other-user interference in a more complicated way that in traditional CDMA interference evaluation. However, both an approximate analysis based on the Gaussian approximation and another analysis with any desirable accuracy based on the characteristic function method are carried out in order to determine the BER of the various traffic types as functions of the information rates, the overall bandwidth, and the number of users in each traffic class. This evaluation is used to determine the capacity region, that is, the maximum number of users that can be supported from each class so that the individual class BER requirements are met. This calculation is carried out for (i) voice and data of different information rates and BERs, (ii) voice and two types of data, all with different information rates and BERs, and (iii) videophone, voice, and data with different rates and BERs.

III. Code Allocation for Multi-Media Traffic

Three types of policies for optimal CDMA code allocation are derived: one that pertains to voice and low priority data, another that pertains to voice, high priority data, and low priority data, and a third one that pertains to videophone, voice, and low priority data.

For the first policy we present an optimal allocation scheme that determines the number of newly arrived voice calls that are accepted in the network so that the long-term blocking (rejection) rate of voice calls is minimized and the packet error probabilities of voice traffic remains within acceptable limits. The unused CDMA capacity is be used by data traffic and the reamining data traffic is queued. We consider two models for the effects of other-user interference: the threshold model and the graceful degradation model.

For the second policy we derive an optimal code allocation scheme that determines the number of newly arrived voice calls and data users with high priority that are accepted in the network so that the long-term weighted blocking rates of voice calls and data traffic is minimized and the packet error probabilities of these two traffic types are within acceptable limits. For the lower priority data we consider two policies. According to the first policy there are no CDMA codes reserved for these data, they get assigned CDMA codes only when the combined voice and high priority data traffic leaves certain codes unused. The second policy operates like the first except that there is also a small number of CDMA codes that are always assigned to low priority traffic. For both schemes the BER requirement for the low priority data traffic is met.

The performance measures are the average blocking rates and average throughputs of the voice calls and all data messages as functions of the offered voice and data traffic loads under the proposed optimal code allocation policy. The queueing delay and the packet loss probability of the low priority data traffic is also evaluated. A semi-Markov decision process (SMDP) with guaranteed BERs for voice and data traffic is used for formulating the system operation as a dynamic code assignment problem. A value-iteration algorithm is applied to this SMDP to derive the optimal policy.

The third policy has many similarities in its operation with the second policy, but it operates on videophone traffic instead of high priority data traffic; this fact gets reflected in the multi-state Markovian model used (more complicated than the one used for data) and the different BER requirements.

References

[1] W.-B. Yang and E. Geraniotis. "Admission Policies for Integrated Voice and Data Traffic in CDMA Packet Radio Networks." *IEEE JSAC*, pp. 654-664, May 1994.

[2] E. Geraniotis, Y.-W. Chang, and W.-B. Yang "Dynamic Code Allocation for Integrated Voice and Multi-Priority Data Traffic in CDMA Networks" *European Transactions on Telecommunications and Related Technologies (ETT)*, pp. 85-96, Jan-Feb 1995.

[3] T.-H. Wu and E. Geraniotis. "CDMA with Multiple Chip Rates for Multi-Media Communications." Proceedings of the *1994 IEEE Conference on Information Sciences and Systems*, pp. 992-997, Princeton March 1994.

Stability Analysis of Quota Allocation Access Protocols in Ring Networks with Spatial Reuse

Leonidas Georgiadis Wojciech Szpankowski[1] and Leandros Tassiulas[2]
IBM T. J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598, U.S.A.
Dept. Computer Science, Purdue University, W. Lafayette, IN 47907, U.S.A.
Electrical Engineering, Polytechnic University, Brooklyn, NY 11201, U.S.A

Abstract — **We consider a slotted ring that allows simultaneous transmissions of messages by different users, known as ring with *spatial reuse*. To alleviate fairness problems that arise in such networks, policies have been proposed that operate in cycles and guarantee that certain number of packets, called *quota*, will be transmitted by every node in every cycle. We provide *sufficient and necessary* stability conditions for such rings.**

I. Introduction

We consider a ring with spatial reuse, i.e., a ring in which multiple simultaneous transmissions are allowed as long as they take place over different links (cf. [1, 2, 3]). Time is divided in slots and each slot is equal to the smallest transmission unit, called packet. A node receiving a packet with destination another node in the ring (ring packet), may retransmit the packet in the outgoing link *in the same slot.*

While rings with spatial reuse have higher throughput than standard token passing rings, they also introduce the possibility that some overloaded nodes may block other nodes from accessing the ring. To avoid this problem, the following policy is proposed in [1, 2] for the operation of the ring. Each node is assigned a number called "quota". The policy operates in cycles. A node is allowed to transmit during a cycle as long as the number of transmitted packets does not exceed its assigned quota. A cycle ends when the quotas of all nodes are delivered to their destinations. In this way, the operation of a node with regular traffic requirements is not adversely affected by nodes that may become overloaded. An analysis of the throughput characteristics of this policy is presented in [3], where it is also shown that if the end-to-end throughput requirements result in aggregate traffic load for each link of the network less than one, then the node quotas can be selected to achieve these throughput requirements.

II. Main Results

The primary goal of this work is to obtain the stability region of the ring network with finite quota and to compare it with the maximum achievable stability region for such ring networks (cf. [3]). The second motivation is to extend our stability approach of multidimensional distributed systems developed in Georgiadis and Szpankowski [4, 5] and Szpankowski [6] to ring networks with spatial reuse. The conditions for stability are derived by means of a technique that is based on an application of mathematical induction, stochastic monotonicity properties and Loynes stability criteria. A special technique, based on the structure of the complement of the stability region and the construction of a dominant system, permits the derivation of the necessary stability conditions from the *instability* condition of a *dominant* system. The general steps of the above stability analysis have been applied to the analysis of other systems as well (cf. [4, 5, 6]). It should be stressed that this general construction of [4, 6] requires detailed and subtle modifications for almost every queueing network which may be far from trivial, and this analysis is a typical example. In addition we provide a decomposition and characterization of the instability region of the system.

The exact computation of the stability region depends on the *distribution* of the arrival processes and this often renders this computation intractable. The dependence on the distribution leads us to the introduction of the notions of the *essential* and *absolute* stability region. The first contains any arrival rate vector such that for every distribution with this arrival rate vector the network is stable. The second contains any arrival rate vector for which there exists some distribution with this arrival rate vector under which the network is stable. Both stability regions have interesting practical implications. If the arrival distribution is not known, then the essential stability region is essentially the operational region of the system. The absolute stability region specifies what is achievable when the arrival streams can be shaped to have the statistics which lead to higher throughput. We present a method based on linear programming that permits the development of upper and lower bounds on the stability region using only the knowledge of the average cycle lengths. For the case of two nodes we provide a closed-form expression for the region of arrival rates where the system is stable for any arrival distribution.

[1] Supported by NSF Grants NCR-9206315 and CCR-9201078.
[2] Research supported in part by NSF under grant NCR-9211417.

References

[1] I. Cidon and Y. Ofek. MetaRing - a full-duplex ring with fairness and spatial reuse. *IEEE Trans. Commun.*, 41, 110–120, 1993.

[2] R. M. Falconer and J. L. Adams. Orwell: A protocol for an integrated service local network. *Br. Telecom Technol. J.*, 3(4):21–35, 1985.

[3] L. Georgiadis, R. Guerin, and I. Cidon, Throughput properties of fair policies in ring networks. *IEEE/ACM Trans. Networking*, 1, 718-728, 1993.

[4] L. Georgiadis and W. Szpankowski, Stability of token passing rings. *Queueing System*, 11, 7-33, 1992.

[5] L. Georgiadis and W. Szpankowski, Stability analysis for yet another class of multidimensional distributed systems, *Proc. 11-th Intern. Conf. on Analysis and Optimization Systems. Discrete Event Systems*, Sophia Antipolis, 523-530, 1994.

[6] W. Szpankowski Stability conditions for some multiqueue distributed systems *Adv. Appl. Probab.*, 26, 498-515, 1994.

Capacity of Digital Cellular CDMA System With Adaptive Receiver

Ian Oppermann[1], Branka S. Vucetic and Predrag B. Rapajic
Department of Electrical Engineering, University of Sydney, Sydney, NSW, Australia. 2006

Abstract — **This paper presents the results of a capacity evaluation for a cellular code-division, multiple access (CDMA) system over a wide-band fading channel. The study compares the performance of a system based on a conventional matched filter with that based on an adaptive receiver. Trellis codes and various rate convolutional codes are investigated in an attempt to improve the system performance. Performance is measured in terms of the maximum number of simultaneous users per cell for a given bit error rate (BER). The channel model is developed from measured propagation data at 2.6 GHz in heavily built-up urban areas and includes the effect of both intra-cell and inter-cell interferers.**

I. Summary

One of the major problems associated with using direct sequence CDMA systems is the low channel efficiency. For single cell systems based on conventional, matched filter receivers, the efficiency is typically between 10 - 20% [1], [2] of the theoretical channel capacity. By using an adaptive linear receiver [1], [3], it is possible to increase the system efficiency of a single cell system significantly with only a moderate increase in receiver complexity. This is also the case for a multi-cell system operating over a multipath fading channel.

The poor efficiency of the matched filter system is primarily due to the multiple access interference (MAI) produced by competing users of the channel bandwidth. Measurements [4] indicate that MAI coming from adjacent cells contributes up to 40 % of the total interference, for equally loaded cells, experienced by a given user. For this reason it is important for a capacity study to examine a system which includes the interference from surrounding cells.

The efficiency of the cellular system will be defined as the maximum number of users that may be supported in one cell, while maintaining a specified BER, multiplied by the data rate of each user

$$\eta = \frac{M r_d}{B} \tag{1}$$

where M is the number of simultaneous users at a given bit error probability, r_d is the data rate and B is the spread spectrum bandwidth. For this calculation, it is assumed that the same number of simultaneous users exist in all cells and all users have the same data rate. This gives an indication of the performance of the system normalised to one cell.

The spreading sequence considered in this paper are GOLD sequences of length ($N = 15$ to 127). For code rates of 1/2, 1/4 and 1/8, these spreading ratios change to 63, 31 and 15 respectively in order to maintain constant bandwidth. With trellis coding, the spreading ratios do not change. The system considered has a BER of 10^{-3} and a data rate of 39.4 Kbps. The transmitted signal is band-limited by a raised cosine filter with a roll-off of 40 %. The average power of the combined signals of interferers from each adjacent cell is varied from -10 dB to -15 dB below the power of the signal of the user of interest. Due to imperfect power control, the signal power of the users in the cell of interest are assumed to be normally distributed with a variance of 1 dB.

Simulation results have shown that the uncoded system based on the adaptive receiver lead to a 4 to 5 fold improvement in system efficiency over the fading channel compared to the matched filter system.

When combining the systems with convolutional codes it was seen that, while offering significant advantages for the matched filter system, this form of coding actually reduces the system efficiency for the adaptive receiver system by restricting the maximum number of simultaneous users. As the code rate decreases from 1/2 to 1/8, the maximum number of users for the matched filter system eventually equals the number for the adaptive receiver system.

The uncoded matched filter system offers a very low efficiency with the maximum number of users well below the spreading ratio N. The gain from coding allows a larger number of simultaneous users however, the number is still well below the spreading ratio. The uncoded adaptive receiver system however offers a maximum number of users which is approximately 70 % of the spreading ratio. The decrease in spreading ratio which accompanies a decrease in code rate causes the maximum number of users to decrease linearly. The gain from coding is insufficient to increase the number of users to that of the uncoded case. This is true for all code complexities examined.

For both systems however, convolutional coding does offer the advantage of a reduction in the signal to noise ratio (SNR) required to achieve a given BER for lightly loaded systems.

Trellis codes conversely require no reduction in spreading ratio and so the effect described above is minimised. Trellis coding allows an increase in efficiency for both systems as well as the coding gain described above for lightly loaded systems.

It is therefore suggested that convolutional coding should not be considered when attempting to maximise efficiency for systems based on adaptive receivers. Rather, trellis codes offer a far better alternative.

[1]This work was supported in part by the Australian Telecommunications and Electronics Research Board (ATERB)

References

[1] P. Rapajic and B.S. Vucetic, "Adaptive Receiver Structures for Asynchronous CDMA Systems" *IEEE Journal On Selected Areas in Communications*, Vol. 12, No. 4, May 1994.

[2] A.J. Viterbi, "When Not to Spread-Spectrum - a Sequel" *IEEE Communications Magazine*, Vol. 23, No. 4, April 1985. pp. 1309-1319.

[3] P. Rapajic and B.S. Vucetic, "Linear Asynchronous Code Division Multiple Access Single User Receiver" *Proceedings of IEEE 2nd International Symposium on Spread Spectrum Techniques and Applications*, Yokohama, Japan, Nov 1992. pp 223 - 231.

[4] K. S. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver Jr and C.E. Wheatly III, "On the Capacity of a Cellular CDMA System", IEEE VT-40, No. 2, May 1991, pp 303-312.

Quadratic-Inverse Spectrum Estimates for Non-Stationary Processes

David J. Thomson
AT&T Bell Laboratories, Murray Hill, New Jersey 07974.

Abstract Quadratic-inverse spectrum estimates for locally white non-stationary processes are described.

I. Introduction

Most time-series data encountered in practice is non-stationary whereas most spectrum estimation methods assume stationarity, and this has resulted in many ad-hoc analysis methods. The multiple-window methods of spectrum estimation[1] is akin to linear inverse theory applied to the discrete Fourier transform. In the original derivation, both stationarity and "local whiteness" were assumed and the resulting estimates of spectra were simply squares of the linear inverse. Quadratic-inverse theory[2] eliminated the locally white assumption and gave stable second moments. The effects of limited non-statonarity on the variance of such estimates is known[3]. This talk describes a way of combining the non-stationary quadratic inverse theory with that of harmonizable processes.

II. Harmonizable Processes

Harmonizable processes may be written as generalized Fourier transforms,

$$x(t) = \int e^{i2\pi\eta t}\, d\xi(\eta)$$

with covariance function $\Gamma_c(t,t') = \mathbf{E}\{x(t)\,\overline{x}(t')\}$ and corresponding generalized spectral density $\gamma_c(\eta,\eta')\,d\eta\,d\eta' = \mathbf{E}\{d\xi(\eta)\,\overline{d\xi}(\eta')\}$. They are connected by the two-dimensional Wiener-Khintchine relation

$$\Gamma_c(t,t') = \int\int e^{i2\pi(t\eta - t'\eta')}\,\gamma_c(\eta,\eta')\,d\eta\,d\eta' .$$

Local stationarity is described by the spectrum parallel to the $\eta = \eta'$ diagonal, while global non-stationarity is on the orthogonal coordinate. Wigner-Ville and dynamic spectra are obtained from 45° coordinate rotations followed by *single* Fourier transforms[4]. For white non-stationary processes the covariance function is $\Gamma_c(t,t') = P(t)\,\delta(t-t')$ so the corresponding generalized spectrum is a function of $\eta - \eta'$ $P(t)$ is the expected power of the process at time t.

III. Estimation

To estimate $P(t)$ we use a multiple-window method. Thus one chooses a frequency f and a bandwidth W and describes the information contained in the band $(f - W, f + W)$ by the coefficients of a *locally* orthogonal expansion of Slepian functions. Given N samples from an observed sequence $x(t)$ one chooses a resolution bandwidth W and computes the eigencoefficients

$$x_k(f) = \sum_{n=0}^{N-1} e^{-i2\pi fn}\, v_n^{(k)}(N,W)\, x(n) \qquad (1)$$

where the $v_t^{(k)}(N,W)$ are the orthonormal discrete prolate spheroidal sequences. The extreme band-limiting properties of the Slepian sequences makes it almost irrelevant whether the white non-stationary model is valid globally, or only locally within the frequency band $(f - W, f + W)$. Denote the vector of eigencoefficients (1) by $\mathbf{X} = \mathbf{X}(f)$ and the covariance matrix of the eigencoefficients by $\mathbf{C}_{jk}(f) = \mathbf{E}\{x_j(f)\,\overline{x_k(f)}\}$.

To estimate limited non-stationarity, expand $P(t)$ on an orthonormal bases

$$P(f,t) = \sum_{l=0} p_l(f)\, A_l(t) .$$

The system whose kernel is the square of the *sinc* kernel defining the Slepian sequences

$$\alpha_l A_l(n) = \sum_{m=0}^{N-1} \left[\frac{\sin 2\pi W(n-m)}{\pi(n-m)}\right]^2 A_l(m)$$

has $4NW$ real eigensequences corresponding to significantly non-zero eigenvalues. The corresponding bases matrices defined by

$$\mathbf{A}_{jk}^{(l)} = \sqrt{\lambda_j \lambda_k} \sum_{n=0}^{N-1} v_n^{(j)}\, v_n^{(k)}\, A_l(n)$$

are *trace-orthogonal*, $tr\{\mathbf{A}^{(l)}\,\mathbf{A}^{(m)}\} = \alpha_l\,\delta_{l,m}$.

The covariance matrix $\mathbf{C}$ may now be written

$$\mathbf{C}(f) = \sum_{l=0} p_l(f)\,\mathbf{A}^{(l)}$$

so quadratic estimates of the expansion coefficients

$$\hat{p}_l(f) = \frac{1}{\alpha_l}\, tr\{\hat{\mathbf{C}}(f)\,\mathbf{A}^{(l)}\} = \frac{1}{\alpha_l}\,\mathbf{X}^{\dagger}(\mathbf{f})\,\mathbf{A}^{(l)}\,\mathbf{X}(f)$$

follow. $p_0(f)$ is proportional to the usual stationary spectrum $S(f)$, while $p_1(f)$ is, loosely, the time derivative if the spectrum $\dot{S}(f)$. It may be shown that these estimates are unbiased and, because $Var\{\hat{p}_l(f)\} \sim \alpha_l^{-1}$, time resolution of non-stationarity is essentially limited to $1/4\ W$.

References

[1] Thomson,D.J., *Spectrum Estimation and Harmonic Analysis,* Proc. IEEE **70,** 1055-96, (1982).

[2] Thomson, D.J., *Quadratic-Inverse spectrum estimates; applications to paleoclimatology,* Phil. Trans. R. Soc. Lond. A. **332,** 539-97, (1990).

[3] Thomson, D.J., *Nonstationary Fluctuations in stationary time-series,* Proc. SPIE **2027,** 236-244, (1993).

[4] Thomson, D.J., *An overview of Multiple-Window and Quadratic-Inverse Spectrum Estimation Methods,* Proc. ICASSP-94, VI-185 to VI-194, (1994).

Parameter Estimation and Order Determination in the Low-Rank Linear Statistical Model

D.W. Tufts, R.J. Vaccaro, and A.A. Shah[1]
Department of Electrical Engineering
The University of Rhode Island
Kingston, RI 02881 USA

Abstract — **We consider the problem of estimating the min-norm solution to a low-rank, linear statistical model. We calculate the statistics of the solution as a function of the statistical characterization of the matrix containing observation noise. We also present a new method for estimating the rank of the underlying noise-free matrix.**

I. Calculation of Bias and Variance

Consider the following linear model

$$\mathbf{y} = \mathbf{H}_{m\times n}\boldsymbol{\theta}\ , \quad m \geq n. \tag{1}$$

The parameter vector $\boldsymbol{\theta}$ is to be estimated from data contained in $\mathbf{H}$ and $\mathbf{y}$. In the absence of observation noise on the data, we assume that the rank of $\mathbf{H}$ is $r < n$. Thus we refer to (1) as a low-rank model, and we are interested in the minimum-norm solution for $\boldsymbol{\theta}$. In practice, the data will contain perturbations (noise), and we assume that both $\mathbf{H}$ and $\mathbf{y}$ are perturbed by additive noise so that the observed data are $[\mathbf{H}\ \mathbf{y}] + \mathbf{N}$. The estimate $\hat{\boldsymbol{\theta}}$ is obtained as the min-norm solution to the low-rank model equation $\hat{\mathbf{H}}\boldsymbol{\theta} = \hat{\mathbf{y}}$, where $[\hat{\mathbf{H}}\ \hat{\mathbf{y}}]$ is obtained from a singular value decomposition of $[\mathbf{H}\ \mathbf{y}] + \mathbf{N}$.

Our results are based on a perturbation expansion for the SVD of a finite-size matrix [1]. We have previously applied these matrix perturbation ideas adaptive detection [2], and performance analysis of array signal processing algorithms [3]. In order to be useful, the perturbed subspaces must not be "too far" from the unperturbed subspaces. This will be true if the noise matrix $\mathbf{N}$ is "small enough." We have quantified the concepts of "too far" and "small enough" in our previous analysis of the threshold effect [4]. First- and second-order expressions for the perturbed subspaces were derived in [5].

In this paper, we use the perturbation formulas to calculate the statistics (bias and variance) of the solution $\hat{\boldsymbol{\theta}}$ as a function of the statistical characterization of $\mathbf{N}$, the matrix containing observation noise. We stress that the perturbation formulas do not require the data record to become large. In addition, this approach can handle arbitrary correlation of the elements of $\mathbf{N}$.

II. Order Determination

In the estimation problem discussed above, a low–rank approximation to the data matrix is utilized to draw an inference. In doing so, it is implicitly assumed that the underlying true rank of the data matrix is known. In practice, this is seldom the case and the underlying true rank is unknown and needs to be determined.

Under high SNR conditions, the perturbed signal subspace is stable and is more or less determined by the underlying (noise-free) signal subspace. This in turn stabilizes the perturbed orthogonal subspace. Hence in different realizations of the data, the singular vectors change erratically, but the space spanned by them remains relatively unchanged. Thus the energy in the perturbed orthogonal subspace is also well defined. It is closely related to the noise energy in the orthogonal subspace. Using matrix perturbation approximations we quantify this idea and evaluate the distribution to be a central χ^2 with $(m-r)(n-r)$ degrees of freedom.

Based on this distribution, we can set a threshold T_r such that $S_r < T_r$ with a probability $1-\alpha$ where α is a small positive number. In other words, if the rank is r then S_r can be well explained by the noise energy alone, and will be below this threshold with high probability. If the rank is $r+1$ or greater then, due to the additional signal energy, S_r will exceed the threshold with high probability. Based on this idea, we develop a recursive procedure on the set of sums of squares of singular values of data matrix that is essentially a signal energy detection procedure in enlarging subspaces.

In order determination there are two basic types of error probabilities, error due to overestimation (false alarm) and error due to underestimation (miss). The proposed method allows the user to set a bound on the false alarm probability. The user can determine a value of SNR for which a prescribed value of probability of detection or probability of net error can be obtained. Thus, the user can specify the conditions under which performance goals, specified by error probabilities, can be obtained.

[1]This work was supported in part by NUWC, Division Newport, under contract N66604-94-C-B607. A.A. Shah is now with Wireless Access Inc., Santa Clara, CA.

References

[1] F. Li and R. J. Vaccaro, "Unified analysis for DOA estimation algorithms in array signal processing," *Signal Processing*, vol. 22, pp. 147–169, November 1991.

[2] I. Kirsteins and D. W. Tufts, "Adaptive detection using low rank approximation to a data matrix," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, pp. 55–67, January 1994.

[3] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: Further unification, simplification, and observations," *IEEE Transactions on Aerosp., Electron. Syst.*, vol. 29, pp. 1170–1184, October 1993.

[4] D. W. Tufts, A. C. Kot, and R. J. Vaccaro, *SVD and Signal Processing, II*, ch. The Threshold Effect in Signal Processing Algorithms Which Use an Estimated Subspace, pp. 300–321. R.J. Vaccaro, ed., Elsevier Science Publishers, 1991.

[5] R.J. Vaccaro, "A second-order perturbation expansion for the SVD," *SIAM J. Matrix Anal. Appl.*, vol. 15, pp. 661–671, 1994.

[6] A. Shah, *Fast and Effective Algorithms for Order Determination and Signal Enhancement.* PhD thesis, University of Rhode Island, Kingston, RI, 1994.

Blind Identification in Non-Gaussian Impulsive Environments[1]

Xinyu Ma (Student Member, IEEE) and Chrysostomos L. Nikias (Fellow, IEEE)
Signal and Image Processing Institute, Dept. of Elect. Eng. - Systems,
University of Southern California, Los Angeles, CA 90089

Abstract — **A new algorithm is proposed for blind channel identification in the impulsive signal environments, where the signals are modeled as symmetric α-stable processes. The Alpha-Spectrum, a new spectral representation based on fractional lower-order moments, is developed. Conditions for blind identifiability of any FIR channel (non-minimum phase, unknown order) are established using the properties of the Alpha-Spectrum.**

I. Introduction

The α-stable distributions are characterized by heavy tails and infinite variance, except for the Gaussian distribution($\alpha = 2$), which is the limiting case. By the Generalized Central Limit Theorem, they are the *only* class of distributions that can be the limiting distributions for sums of i.i.d. random variables. These properties make the α-stable distributions attractive models for impulsive data [1]. Most algorithms for blind identification of FIR channels are based on second- or higher-order statistics. When the signals are impulsive and modeled as α-stable processes, these algorithms fail. We propose a robust blind identification method based on a new spectral representation: the **α-Spectrum**, for the impulsive environments. We then prove the blind identifiability of any FIR channel (non-minimum phase, unknown order) driven by white SαS ($\alpha > 1$) processes.

II. Blind Identification with the α-Spectrum

Our new blind identification method is based on the properties of covariation, which plays a role analogous to covariance. For two jointly SαS random variables X and Y with $1 < \alpha \leq 2$, covariation is defined by $[X,Y]_\alpha = \int_S xy^{<\alpha-1>}\mu(ds)$,[2] where S is the unit circle and $\mu(\cdot)$ is the spectral measure of the SαS random vector (X,Y). The FLOM (fractional lower-order moment) estimator for covariation is $[X,Y]_\alpha = \frac{\mathbf{E}(XY^{<p-1>})}{\mathbf{E}(|Y|^p)}\gamma_y$, $p \geq 1$, where X, Y are both real or isotropic complex $S\alpha S$ random variables and γ_y is the dispersion of Y [2]. Properties of covariation include:

1. If X_1, X_2, Y are jointly SαS, then $[aX_1 + bX_2, Y]_\alpha = a[X_1,Y]_\alpha + b[X_2,Y]_\alpha$.
2. If Y_1, Y_2 are *independent* and X, Y_1, Y_2 are jointly SαS, then $[X, aY_1 + bY_2]_\alpha = a^{<\alpha-1>}[X,Y_1]_\alpha + b^{<\alpha-1>}[X,Y_2]_\alpha$.
3. If X, Y are independent, then $[X,Y]_\alpha = 0$.

[1]This work was supported by the Office of Naval Research under contract N00014-92-J-1034.

[2]The notation:

$$Y^{<p-1>} = \begin{cases} |Y|^{p-2}Y^* & Y\text{: complex} \\ |Y|^{p-1}\text{sign}(Y) & Y\text{: real.} \end{cases}$$

For a FIR channel $Y_n = \sum_{i=0}^{q} h_i X_{n-i}$, using the above properties, we have:

$$S_\alpha(z) \triangleq [Y_n, \sum_{i=-q}^{i=q} Y_{n-i}z^i]_\alpha = \gamma_x H\left((\frac{1}{z})^{<\alpha-1>}\right)(H(z))^{<\alpha-1>}, \quad (1)$$

where γ_x is the input dispersion and $H(z)$ is the filter transfer function. Eq.(1) is of fundamental importance. We name $S_\alpha(z)$ **the α-Spectrum**, with which, we can identify both the magnitude and phase responses of the channel. The magnitude can be obtained by letting $|z| = 1$, then $S_\alpha(e^{j\omega}) = \left|H(e^{j\omega})\right|^\alpha$. To obtain channel phase response, noticing the magnitude $|H(z)|$ and phase $\Phi(z)$ of any FIR channel can be expressed in terms of $A^{(m)} = \sum_{i=1}^{L_1} a_i^m$ and $B^{(m)} = \sum_{i=1}^{L_2} b_i^m$, where $\{a_i\}$ and $\{\frac{1}{b_i}\}$ are the zeros inside and outside the unit circle, respectively. $\frac{A^{(m)}+B^{(m)}}{m}$ determines $|H(e^{j\omega})|$ and $\frac{A^{(m)}-B^{(m)}}{m}$ determines $\Phi(e^{j\omega})$ [3]. Taking logarithm of both sides of Eq.(1), we have:

$$\log|S_\alpha(re^{j\omega})| = \alpha\log A_0 - \sum_{m=1}^{\infty}\frac{A^{(m)}\mu_m(r) + B^{(m)}\mu_m(\frac{1}{r})}{m}\cos(m\omega). \quad (2)$$

$$\Psi(re^{j\omega}) = \sum_{m=1}^{\infty}\frac{A^{(m)}\nu_m(r) - B^{(m)}\nu_m(\frac{1}{r})}{m}\sin(m\omega), \quad (3)$$

where $|S_\alpha(re^{j\omega})|$ and $\Psi(re^{j\omega})$ are the magnitude and phase of the α-spectrum, and $\mu_m(r) = r^{m(\alpha-1)} + (\alpha-1)r^{-m}$, $\nu_m(r) = r^{m(\alpha-1)} - r^{-m}$. $\frac{A^{(n)}-B^{(n)}}{n}$ can be obtained from either $|S_\alpha(re^{j\omega})|$ or $\Psi(re^{j\omega})$. Therefore the channel phase response is:

$$\Phi(e^{j\omega}) = \sum_{n=1}^{\infty}\left(\frac{\frac{2}{\pi}\int_0^\pi\left(\Psi(re^{j\omega}) + \Psi(\frac{1}{r}e^{j\omega})\right)\sin(n\omega)d\omega}{\nu_n(r) + \nu_n(\frac{1}{r})}\right)\sin(n\omega), \quad (4)$$

or

$$\Phi(e^{j\omega}) = \sum_{n=1}^{\infty}\left(\frac{\frac{2}{\pi}\int_0^\pi \log\frac{|S_\alpha(\frac{1}{r}e^{j\Omega})|}{|S_\alpha(re^{j\Omega})|}\cos(n\Omega)d\Omega}{\mu_n(r) - \mu_n(\frac{1}{r})}\right)\sin(n\omega), \quad (5)$$

Acknowledgements

We are thankful to Dr. Min Shao for fruitful discussions.

References

[1] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, NY, 1995.

[2] S. Cambanis and G. Miller, "Linear Problems in p-th Order Stable Processes", *SIAM J. Appl. Math.*, vol.41, no.1, pp. 43-69, Aug., 1981.

[3] C. L. Nikias and A. P. Petropulu, *Higher Order Spectral Analysis: A Nonlinear Signal Processing Framework*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

Algorithms for blind identification of digital communication channels

Javier Buisán Gómez del Moral and Ezio Biglieri[1]
Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy)
fax: +39 11 5644099 • e-mail: biglieri@polito.it

Abstract — **Blind identification consists of estimating the impulse response of a linear, time-invariant channel used for transmission of digital data by observing the channel output without knowledge of the transmitted symbol sequence.**

The aim of this paper is twofold. First we compare, in order to assess their applicability to the equalization of digital radio links affected by selective fading, some recently proposed algorithms based on the second-order statistics of the received signal. Further we show how one of these algorithms can be modified to account for correlated noise.

I. Introduction

By blind identification we mean here the estimate of the impulse response of a linear, time-invariant noisy channel used for transmission of digital data; this estimate is obtained by observing the channel output without knowledge of the transmitted symbol sequence.

The desirable features of the ultimate blind identification algorithm are the following:

- Low identification error in the presence of noise.
- Fast convergence.
- Computational simplicity.
- Insensitivity to data-symbol correlation.
- Insensitivity to noise correlation.
- Possibility to make it adaptive.

Tong, Xu, and Kailath [6, 7] have developed a blind identification algorithm (herewith referred to as TXK algorithm) which is based on an estimate of the autocorrelation function of the observed channel-output samples. This feature entails a convergence faster than other blind algorithms based on higher-order statistics [9], which is highly desirable when the channel is time-varying and its variations have to be tracked quickly in order to compensate for them. This algorithm converges globally, can resolve the non-minimum-phase zeros of the channel transfer function, and is robust with respect to timing recovery. However, it suffers from some drawbacks, viz.,

- It is computationally intensive, as it requires two singular-value matrix decompositions.
- It requires data symbols to be uncorrelated.
- It requires the noise to be uncorrelated.
- It is not adaptive.

More recently, an improved algorithm (herewith referred to as MDCM) was proposed by Moulines *et al.* [4]. The advantages of this new algorithm over TXK are:

- Lower computational complexity.
- Convergence even with correlated data symbols.
- Lower identification error for the same observation length.

Baccalá and Roy [1, 2] have proposed a new algorithm (herewith referred to as BR) that presents a significantly lower computational complexity and an identification error close to that of TXK and MDCM algorithms.

II. Our results

The aim of this paper is twofold. First we compare, by combining analysis and simulation, the TXK, MDCM, and BR algorithms, in order to assess their applicability to the equalization of digital radio links affected by selective fading. Our results show that in general these algorithms based on second order statistics outperform standard blind equalization in terms of convergence speed. Moreover, while the BR algorithm has a lower computational complexity, in this specific application the MDCM algorithm outperforms both TXK and BR in terms of robustness to source-data correlation and mean-square estimation error.

Further, we derive a modification of the MDCM algorithm in [4] in order to achieve blind identification in the presence of unknown correlated noise. Our algorithm is based on a matrix decomposition method proposed in [5].

[1]This research was sponsored by the Human-Capital and Mobility Program of the European Union.

References

[1] L. A. Baccalá and S. Roy, "A new blind time-domain channel identification method based on cyclostationarity," *Signal Processing Letters*, Vol. 1, No. 6, pp. 89–91, 1994.

[2] L. A. Baccalá and S. Roy, "Time-domain blind channel identification algorithms," *Proc. of 1994 CISS*, Princeton, NJ, March 1994.

[3] W. A. Gardner, "A new method of channel identification," *IEEE Trans. Commun.*, Vol. 39, No. 6, pp. 813–817, June 1991.

[4] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multipath channels," *IEEE Trans. Signal Processing*, submitted for publication, 1993.

[5] R. Rajagopal, P. Ramakrishna Rao, "DOA estimation with unknown noise fields: a matrix decomposition method," *IEE Proceedings-F*, vol. 138, No 5, October 1991.

[6] L. Tong, G. Xu and T. Kailath, "A new approach to blind identification and equalization of multipath channels," *Proc. of the 25th Asilomar Conference*, Pacific Grove, CA, pp. 856–860, Nov. 1991.

[7] L. Tong, G. Xu and T. Kailath, "Blind identification and equalization of multipath channels," *Proc. of ICC'92*, pp. 351.3.1–351.3.5, 1992.

[8] J. K. Tugnait, "Fractionally Spaced Blind Equalization And Estimation Of FIR Channels," *Proc. 1993 Intern. Conf. on Communications*, Geneva, Switzerland, May 23-26, 1993.

[9] J. K. Tugnait, "Fractionally spaced blind equalization of FIR channels under symbol timing offsets," *27th Conf. Signals Systems Computers*, Nov. 1-3, 1993, Pacific Grove, CA.

A CLASS OF ITERATIVE SIGNAL RESTORATION ALGORITHMS

Joseph P Noonan[1], Premkumar Natarajan[2] and Baaziz Achour[3]

[1,2]Department of Electrical Engineering and Computer Science , Tufts University , Medford , MA 02155, USA

[3]Qualcomm Inc., 6455 Lusk Blvd., San Diego , CA 92121, USA

Abstract — **Iterative methods have of late enjoyed increasing popularity in signal restoration problems. Inherent mathematical difficulties have led researchers to propose *ad hoc* solutions in many instances. The question of optimality of such solutions is an open one. This paper concerns this question for a class of iterative methods of signal restoration and offers a criterion for optimality based on information theory.**

I. Introduction

The signal restoration problem has classically been modelled as that of estimating the input z to a sytem , assuming that the distorting process h is specified or estimatable and that the distorted output u is available. Further generalizations incorporate any *a priori* knowledge about the solution , into the restoration process , in the form of constraint(s). A well defined system model in combination with robust techniques naturally leads to good results. More interesting is the case where the problem belongs to the class of *ill-posed* problems. Regularization theory is used to pose a corresponding *well-posed* problem , the solution to which is a close enough approximation of the solution to the ill-posed problem being considered [1].

II. Algorithm Derivation

The problem is formulated as the constrained minimization of a stabilizing functional [2] $\Omega(x)$. Recently , Noonan and Achour[3], [4] studied the use of the Itakuro-Saito distance from communication theory and the Kullback-Leibler measure from statistics as stabilizing functionals. They proposed a generalized mapping function based on the use of the Mutual Information Measure as the stabilizing functional and incorporated a priori noise variance information as a mean squared error constraint. Mathematically stated,

Minimize :

$$\Omega(\mathbf{u},\mathbf{z}) = \sum_{\mathbf{u}}\sum_{\mathbf{z}} P_{u,z}(u,z) ln\left[\frac{P_{u/z}(u/z)}{P_u(u)}\right]$$

Subject To :

$$\frac{1}{N}\sum_{i=1}^{N}(u - h * z)^2$$

The proposed generalized mapping function resulting from this optimization leads to the following general form ,

$$\varphi(z_{n+1}) = \varphi(z_n) exp\left(\lambda z_n'\left([u - z_n * h] * h_f\right)\right)$$

where h_f is the flipped i.e. time-reversed version of the distorting process h and z' is the partial of z with respect to $\varphi(z)$.

III. Special Cases of the Mapping

In this paper we demonstrate that various well known *ad hoc* algorithms are special cases of the proposed mapping function [3],[4]. In particular the pioneering Van-Cittert method and the popular Landweber restoration technique are shown to belong to this class of *optimum* algorithms.

The specific mapping $\varphi(z) = exp(z'(z * h_f))$ yields ,

$$z_n = z_n + \lambda(u - h * z_n)$$

which is the Van Cittert restoration algorithm with a particular smoothing function h, while the specific mapping $\varphi(z) = exp(z'z)$ yields ,

$$z_{n+1} = z_n + \lambda(u - z_n * h) * h_f$$

which is the Landweber restoration algorithm with a particular smoothing function h

The algorithm generated by the use of the Kullback-Leibler Measure is given by ,

$$z_n = z_n exp(\lambda([u - z_n * h] * h_f))$$

which can be generated from the generalized mapping function by the trivial mapping $\varphi(z) = z$

This provides a sound statistical argument for the use of these methods and establishes an optimality interpretation for their estimates. The Mutual Information Measure is based on the concepts of entropy and information content. Interestingly , the proposed mapping function is derived by normalizing the signal and identifying this as the probability density function of the signal itself. This method can thus be applied to signals whose probability structure is not fully known. The derived algorithm has been shown to be stable and robust[5]. There exists a strong condition for the convergence of the mapping in the general case. For specific cases the condition simplifies to a weaker problem specific condition.

Applications of the above algorithm are discussed. In particular noisy image restoration and high resolution estimation of spectra are presented. This work is a continuation of that presented in [3],[4], [5] .

References

[1] Tikhonov, A. N. and Arsenin,V.Y. *Solution of Ill-Posed Problems.* Wiley , New York , 1977.

[2] Karayannis, N. B. and Venetsanopoulos,A.N. "Restoration theory in image restoration - the stablizing functional approach." *IEEE Trans. Acoustics , Speech and Signal Processing.***ASSP-38** *(July , 1990) , 1155-1179.*

[3] Noonan, J. P. and Achour,B. "Two new Robust nonlinear signal restoration algorithms." *Digital Signal Processing* **2** *, 39-43 (1992).*

[4] Noonan, J. P. and Achour,B. "A Hybrid MIP algorithm for signal restoration." *IASTED International Conference On Adaptive Control , Signal Processing.* NY 1990 , pp. 93-97 .

[5] Achour, B. "Information Theoretic Criteria In Signal Restoration." *Ph.D Dissertation*-1991, Dept. of Electrical Engineering , Tufts University.

Stochastic processes and linear combinations of periodic clock changes

N.D.Aakvaag, A.Duverdier and B.Lacaze
ENSEEIHT/GAPSE, 2 rue Camichel, 31071 Toulouse, France

Abstract — **Stochastic processes subjected to a periodic clock change function will have weighted versions of its power spectrum reproduced at integer multiples of the jitter frequency. It has been shown that the original process may, in theory, be reconstructed without error by a suitable choice of correction filter [1]. In this paper we extend the results presented in [1] to the general case where the resulting process is a linear combination of N clock change functions.**

I. Definitions

Let $Z = \{Z(t), t \in \mathbb{R}\}$ be a random wide sense stationary process of zero mean and mean square continuous admitting an autocorrelation function $K_Z(\tau)$ and a Cramer-Loeve representation $\Theta_Z(\omega)$ [2]. The results regarding a single periodic clock change are given in [1]. We consider here the following extended definition:

$$W(t) = \sum_{n=1}^{N} Z(t - f_n(t+\theta))\, g_n(t+\theta) \qquad (1)$$

where θ is a random variable, independent of $Z(t)$ and uniformly distributed on $(0, 2\pi)$. We assume that the functions $f_n(\cdot)$ and $g_n(\cdot)$ are periodic with the same period $T_n = 2\pi/\omega_n$, and that the frequencies are related in the following manner:

$$\frac{\omega_n}{\omega_m} = \frac{p}{q} \qquad (p,q) \in \mathbb{N} \times \mathbb{N}^* \quad \forall (m,n) \in \{1..N\}^2 \qquad (2)$$

There then exists a frequency λ, the smallest multiple of the set $\{\omega_n\}$ such that the functions $f_n(\cdot)$ and $g_n(\cdot)$ are periodic in $T_\lambda = 2\pi/\lambda$.

II. Power Spectrum and Linear Periodic Filtering

Using the Cramer-Loeve representation of (1), we find

$$W(t) = \int_{\mathbf{R}} e^{i\omega t} \sum_{n=1}^{N} e^{-i\omega f_n(t+\theta)} g_n(t+\theta)\, d\Theta_Z(\omega) \qquad (3)$$

where the summation is periodic (of period T_λ). If it can be expressed in terms of its Fourier components

$$\sum_{n=1}^{N} e^{-i\omega f_n(u)} g_n(u) = \sum_{k=-\infty}^{\infty} \Phi_k(\omega)\, e^{ik\lambda u} \qquad (4)$$

then $W(t)$ is a cyclostationary process, stationarised by the phase θ. $W(t-\theta)$ then admits a continous series representation whose elements, the responses of $Z(t)$ through the linear invariant filters $\Phi_k(\omega)$, are jointly stationary [3]. This notation also demonstrates that $W(t-\theta)$ can be seen as the filtering of $Z(t)$ by a linear periodic filter [4], whose impulse response is given by

$$h(t, t-\tau) = \sum_{k=-\infty}^{\infty} e^{ik\lambda t} \int_{\mathbf{R}} \Phi_k(\omega)\, e^{i\omega\tau} d\omega \qquad (5)$$

The power spectrum of $W(t)$ follows directly from that of $W(t-\theta)$

$$dS_W(\omega) = \sum_{k=-\infty}^{\infty} |\Phi_k(\omega - k\lambda)|^2\, dS_Z(\omega - k\lambda) \qquad (6)$$

III. Reconstruction

A linear reconstruction of $Z(t)$ based on the observation of a frequency band centered around $k\lambda$ of $W(t)$, is the linear projection of $Z(t)$ on the Hilbert space spanned by this process. This reconstruction is ideal when the spectrum of $Z(t)$ is bounded such that

$$dS_Z(\omega) = 0 \qquad \forall \omega \notin \left(-\frac{\omega_\lambda}{2}, \frac{\omega_\lambda}{2}\right) \qquad (7)$$

It can be shown that this is a filtering operation, where the filter is given by

$$H_k(\omega) = \sqcap_k(\omega) / \Phi_k(\omega) \qquad (8)$$

where $\sqcap_k(\omega)$ is an ideal bandpass filter centered at $\omega = k\lambda$ and $\Phi_k(\omega)$ is given by

$$\Phi_k(\omega) = \sum_{n=1}^{N} \frac{1}{T_n} \int_0^{T_n} e^{-i\omega f_n(u) - ik\omega_n u} g(u)\, du \qquad (9)$$

IV. Example

Consider the case where $Z(t)$ is a sequence of independent bits, $N = 2$, and the periodic functions are given by

$$\begin{array}{ll} f_1(t) = \alpha \sin(\omega_1 t) & f_2(t) = \beta \sin(\omega_1 t/2) \\ g_1(t) = 1 & g_2(t) = \exp(i\omega_1 t/2) \end{array} \qquad (10)$$

In this case, the receiving filter, based on the baseband signal is given by

$$H_0(\omega) = \frac{1}{J_0(\alpha\omega) + J_1(\beta\omega)} \qquad (11)$$

where $J_n(\omega)$ is the n'th order Bessel function.

V. Conclusion

In this paper we have presented a generalisation of periodic clock changes. We developed the optimal receiving filter and found its analytical expression in a particular example.

References

[1] B.Lacaze and N.D.Aakvaag, *Stationary random functions and periodic clock changes*, ISIT'94

[2] H.Cramer and M.R.Leadbetter, *Stationary and related stochastic processes*, Wiley, New-York, 1967

[3] W.A.Gardner and L.E.Franks, *Characterisation of cyclostationary random signal processes*, IEEE Transactions on Information Theory, Vol.IT-21, pp.4-14, 1975

[4] L.E.Franks, *Polyperiodic Linear Filtering*, Cyclostationarity in Communication and Signal Processing : Gardner, IEEE Press, 1994

Near Optimum Filtering of Quantized Signals

Marcelo S. Alencar[1] and Jacob Scharcanski
Departamento de Engenharia Elétrica, Universidade Federal da Paraíba
58109-970 Campina Grande PB Brazil – E. mail: malencar@dee.ufpb.br

Abstract — **A near-optimum method for filtering out the quantization noise is presented. Use is made of the result that the spectrum of the quantization noise is related to the probability density function of the signal derivative.**

I. Introduction

The need for representing signals by a finite number of bits implies that quantization noise is present in almost all digital signal processing systems and inherently occurs in the analog-to-digital conversion process. The distortion error, or quantization noise, consists of the difference between the input to the quantizer and the discrete output signal.

In the following a near-optimum method for filtering out the quantization noise is presented. Use is made of the result that the spectrum of the quantization noise is remarkably related to the probability density function (pdf) of the signal derivative.

II. Near Optimum Filtering of Quantized Signals

For a signal $x(t)$, assumed stationary in the wide sense, the power spectrum density (PSD) of the quantization noise was shown to be given by [1]

$$S_N(w) = \frac{d^2}{2\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^3}p_{X'}(\frac{wd}{2\pi n}) \qquad (1)$$

where $p'_X(.)$ is the pdf of the derivative of the input signal.

Equation 1 demonstrates that the PSD of the quantization noise is related to the pdf of the derivative of the input signal. The convergence of the noise spectrum to Equation 1, as the stepsize decreases, is a result of a previous work [2].

In order to design the optimum filter, for extracting the signal corrupted by noise, there are two common approaches: minimization of the signal to quantization noise ratio (SQNR), that leads to the matched filter, or minimization of the mean square difference between the input signal and its estimate, leading to the theory of Wiener filtering.

It is useful to consider an estimator based on the Wiener filter, that minimizes the expected value $E[x(t)-\hat{x}(t)]^2$, where the estimator is given by

$$\hat{x}(t) = \int_{-\infty}^{\infty} h(t-\tau)[x(\tau)+n(\tau)]d\tau. \qquad (2)$$

If a Gaussian process is assumed as the input signal, the cross correlation between the quantization noise and the signal is given by [3]

$$R_{XN}(\tau) = 2R_X(\tau)\sum_{n=1}^{\infty}(-1)^n e^{\pi^2 n^2 SQNR/6}. \qquad (3)$$

[1]This work was partially sponsored by the Brazilian Council for Scientific and Technological Research (CNPq).

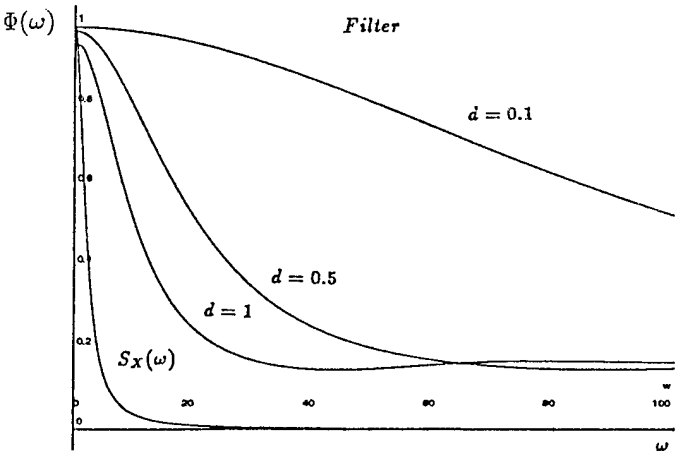


Figure 1: Near optimum filters for selected quantization stepsizes.

For an SQNR above $0dB$, the cross correlation is about eight orders of magnitude smaller than the input signal autocorrelation. This corroborates the assumption of an uncorrelated noise at the output of the quantizer. Therefore, based on the formula for the noise spectrum, one can design a Wiener filter that is near optimum for the above conditions

$$\Phi(\omega) = \frac{S_X(\omega)}{S_X(\omega) + \frac{d^2}{2\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^3}p_{X'}(\frac{\omega d}{2\pi n})}. \qquad (4)$$

where $S_X(\omega)$ is the signal PSD.

Figure 1 depicts the results of application of Formula 4, with the signal obtained by passing white Gaussian noise through a first-order Butterworth lowpass filter [4], for selected values of the stepsize, and shows how the number of levels can influence the design of a filter. The signal spectrum is also drawn in the same figure. It seems clear from this figure that the optimum filter tends to an allpass filter as the stepsize decreases.

References

[1] Marcelo S. Alencar, "A Frequency Domain Approach to the Optimization of Scalar Quantizers", *Proceedings of the IEEE International Symposium on Information Theory*, San Antonio, USA, page 440, January, 1993.

[2] Marcelo S. Alencar, "A New Bound on the Estimation of the Probability Density Function Using Spectral Analysis", *Proceedings of the IEEE International Symposium on Information Theory*, San Antonio, USA, page 190, January, 1993.

[3] B. Lévine, Fondements Théoriques de la Radiotechnique Statistique Éditions de Moscou, Moscow, U.S.S.R, 1973.

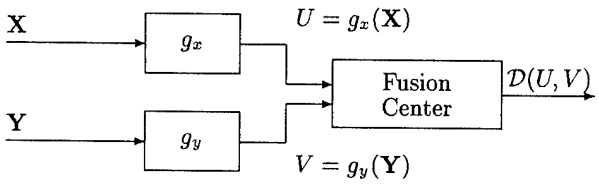
[4] N. S. Jayant and Peter Noll, Digital Coding of Waveforms, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984

Likelihood Ratio Partitions for Distributed Signal Detection in Correlated Gaussian Noise

Po-Ning Chen and Adrian Papamarcou[1]

I. Introduction

A distributed detection system is considered in which two sensors and a fusion center jointly process the output of a random data source (see figure). It is assumed that the null and alternative distributions are spatially correlated Gaussian, differing in the mean; thus the random source is either noise only or a deterministic signal plus noise.



In the presence of spatial dependence, the joint optimization of local quantizers g_x, g_y, and global decision rule $\mathcal{D}$ may yield solutions in which g_x and g_y are *not* based on marginal likelihood ratio tests. This is one instance where distributed detection departs from the traditional statistical framework where likelihood ratios are sufficient for most purposes. This departure was first noted in [1], and was corroborated specifically for the additive Gaussian noise model by means of a counterexample [2] involving two-dimensional vectors $\mathbf{X}$ and $\mathbf{Y}$.

This work is an attempt to characterize noise models for which the optimal system employs marginal likelihood ratio tests. In the setup where each sensor draws one local observation (i.e., $\mathbf{X}$ and $\mathbf{Y}$ are scalars X and Y, respectively), we succeed in obtaining a sufficient condition on the noise mean and covariance under which the optimal binary quantizers are contiguous partitions of the marginal observation space. Since the marginal likelihood ratio is a linear function of the local observation (X or Y), this result implies that g_x and g_y are threshold-type functions of the marginal likelihood ratio. It also reduces the optimization to identifying break points (thresholds) in the marginal observation space.

We also examine whether the sufficient condition discussed previously is also necessary, and find that violation of this condition may in certain—but not all—cases render the contiguous marginal likelihood ratio partition suboptimal. We reach this conclusion by examining the special case where the noise marginals are the same for both sensors; the sufficient condition is then equivalent to positive correlation between X and Y. We find that for values of the correlation coefficient $\rho(X,Y)$ close to -1, local quantizers based on non-contiguous likelihood ratio partitions outperform those based on contiguous likelihood ratio partitions. We were not able to establish the same for $\rho(X,Y)$ close to 0^-.

[1]Po-Ning Chen is with the Computer and Communication Research Laboratories at the Industrial Technology Research Institute, Hsin-Chu, Taiwan ROC.

Adrian Papamarcou is with the Department of Electrical Engineering at the University of Maryland, College Park, USA.

Finally, we consider the following question. Assuming that the sufficient condition discussed previously is satisfied, does symmetry in the signal and noise models (same marginal for both sensors) imply symmetry in the optimal solution, with g_x and g_y being identical contiguous partitions of the real line? We find that this is indeed true, and in such cases, optimal design is further simplified.

II. Statement of Results

The observation statistics are denoted by

$$H_0 : P_{xy} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right)$$

$$H_1 : Q_{xy} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \eta \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}\right).$$

A Bayesian setting is assumed, in which H_0 and H_1 are assigned prior probabilities. Also, quantizers are binary throughout, i.e., $\|g_x\| = \|g_y\| = 2$.

Theorem 1 *If*

$$\sigma_{xy}(\eta\sigma_x^2 - \mu\sigma_{xy})(\mu\sigma_y^2 - \eta\sigma_{xy}) \geq 0, \tag{1}$$

then there exist optimal quantizers of X and Y which are contiguous partitions of the real line.

Counterexample Assume a uniform prior. Let $\sigma_x^2 = \sigma_y^2 = \mu = \eta = 1$ and $\sigma_{xy} = -1$, so that (X,Y) lies on a straight line with probability 1 under each hypothesis. It can be shown that every contiguous binary partition of the real line is outperformed by noncontiguous one.

Remark The above counterexample clearly represents an extreme case where either of the local observations is a sufficient statistic for centralized testing. The same effect, however, can be obtained by choosing $\sigma_{xy} \approx -1$ and applying a continuity argument. A nondegenerate counterexample can then be constructed.

Theorem 2 *Let $\mu = \eta$ and $\sigma_x^2 = \sigma_y^2$. If the local quantizers are constrained to be binary contiguous partitions of the real line, then the optimal quantizer pair employs the same threshold in both quantizers.*

In conjunction with Theorem 1, the above theorem implies the following corollary.

Corollary 1 *Let the signal and noise models be symmetric. If $\sigma_{xy} \geq 0$, then an optimal solution exists in which both quantizers use the same contiguous partition of the observation space.*

References

[1] R. R. Tenney and N. R. Sandell, Jr. "Detection with distributed sensors." *IEEE Trans. Aerospace Electron. Syst.*, AES-17(4):501–510, July 1981.

[2] Po-Ning Chen. "Large deviations approaches to performance analysis of distributed detection systems." *Ph.D. Dissertation*, University of Maryland at College Park, 1994.

RATIONAL MOMENT MAPPING

HC FERREIRA

Cybernetics Laboratory, Rand Afrikaans University, PO Box 524, Auckland Park, 2006, South Africa.
Tel: +27-11-489-2463/2147, Fax: +27-11-489-2357/2054, E-mail: hcf@ing1.rau.ac.za

ABSTRACT - **In this paper we report on some progress towards a solution for the assignment problem in coding, ie the mapping of information words onto codewords at the encoder, and the computationally more difficult problem of mapping codewords onto information words at the decoder.**

I. INTRODUCTION

The assignment problem is still unsolved for many classes of constrained or nonlinear codes [1-7] , including some recording codes developed in recent years [2]. A few previous approaches have been based on applying techniques such as Pascal's triangle [1,2] or mapping block codes onto trellises [3].

Previously, we investigated the assignment problem for constrained codes with short symbol lengths, in order to minimize error extension at the decoder [8]. In this paper, we consider longer codeword lengths and focus on designing a mapping algorithm feasible for implementation when a single lookup table for mapping codewords onto information words, becomes too large to implement.

Traditionally, coding and mapping algorithms have been based on computations which can be modeled as integer manipulations. The mapping algorithm that we propose in this paper, exploits the capability of modern digital circuits to handle rational numbers. It can be implemented with magnitude comparison of integers and two lookup tables, which are for many codes of interest much smaller than the abovementioned single lookup table. For each codeword, a moment with rational weights is computed, similar to the moment computed in the Varshamov-Tenengol'ts construction [7], or when constructing higher-order spectral null codes [2]. If necessary, this computation can be implemented as arithmetic multiplication and division of integers.

II. ALGORITHM

Consider the mapping of codewords onto information words at the decoder of an (n,k) binary block code. An exhaustive lookup table requires a memory of size 2^n, and it may be infeasible to implement. We propose a decoder with memory upperbounded by $2^{k+1}/n$. Our algorithm is thus of interest when decoding codes where a memory of $2^{k+1}/n$ is feasible to implement, while a memory of 2^n is infeasible. For example, many constrained, constant weight, or nonlinear codes of interest have $R = k/n \sim 1/2$ and $k \leq 20$ [2,4,5,6].

When setting up the decoder, we start with the set of 2^k n-bit codewords, $\boldsymbol{x} = (x_1 \ldots\ldots x_n)$, which are ordered using the standard lexicography of n-bit binary numbers. Next, we partition the codebook into $2^k/n$ subsets of consecutive codewords, each with cardinality upperbounded by n. For each subset, we set up a system of n linear equations as follows. For the λ'th codeword, $\boldsymbol{x}^\lambda$, we set

$$\sum_{j=1}^{n} a_j x_j^\lambda = I_\lambda \ , \qquad (1)$$

where I_λ is the integer representation of the information word onto which this codeword is mapped. We now use the set of n linear equations to solve a set of weights $\{a_i\}$ for each subset of codewords. These sets of weights are stored in Table A of dimension $2^k/n$ at the decoder. A second lookup table of dimension $2^k/n$ at the decoder, Table B, is used to store the lexicographically last codeword of each subset.

When mapping a codeword onto an information word, the decoder compares it to the entries in Table B, to determine which entry from Table A should be used to compute the information word, using (1).

While the algorithm is thus conceptually simple, it may present interesting algebraic or combinatorial problems when taking advantage of the structure inherent in many codebooks. For some classes of codes, each codeword's complement is also in the codebook or the codebook can be partitioned into codebooks with words which are identical, except for different prefixes. In these cases, it is possible to reduce the size of the lookup tables at the decoder. The memory requirements may be reduced, or the necessity of accessing the entries in Table B sequentially, may also be precluded, using tree searches.

III. CONCLUSIONS

Many interesting classes of constrained, constant weight, or other nonlinear codes, have previously been constructed in the literature. While the useful properties of these codes have been proved, the difficult problem of mapping codewords onto information words, is often not addressed. With this paper, we hope to contribute towards a general approach to solve this problem.

REFERENCES

[1] K.W. Cattermole, "Principles of pulse code modulation," New York: Elsevier, 1969.

[2] K.A.S. Immink, "Coding techniques for digital recorders," New York: Prentice Hall, 1991.

[3] G. Markarian and B. Honary, "Trellis decoding for the block RLL/ECC codes," Proceedings of the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, June 27-July 1, 1994, p.147.

[4] F.J. MacWilliams and N.J.A. Sloane, "The theory of error correcting codes," New York: North-Holland, 1988.

[5] A.E. Brouwer, J.B. Shearer, N.J.A. Sloane and W.D. Smith, "A new table of constant weight codes," IEEE Transactions on Information Theory, vol. 36, no. 6, pp. 1334-1380, November 1990.

[6] M. Blaum, "A (16,9,6,5,4) error-correcting dc free block code," IEEE Transactions on Information Theory, vol. IT-34, no. 1, pp. 138-141, January 1988.

[7] R.R. Varshamov and G.M. Tenengol'ts, "Correction code for single assymmetrical errors," Avtomika i Telemekhanika, vol. 26, no.2, pp. 288-292, February 1965.

[8] A.S.J. Helberg and H.C. Ferreira, "On the complete decoding of constrained codes," IEEE Transactions on Information Theory, vol. IT-39, no. 1, pp. 228-232 , January 1993.

Real-Time Tracking of Nonstationary Signals Using the Jacobi SVD Algorithm

F. Lorenzelli and K. Yao [1]
Electrical Engineering Dept., UCLA, Los Angeles, CA 90095-1594

Abstract — **In this talk we consider real-time nonparametric algorithms for nonstationarity detection and SVD updating based on Jacobi rotations. We propose two schemes which improve the overall performance when the rate of change of the data is high. In the "variable rotational rate" scheme, the number of Jacobi rotations per update is dynamically determined. In the "variable forgetting factor" approach, the effective width of the observation adjusts to the data nonstationarity.**

I. Introduction

In this talk, we investigate the algorithmic and architectural relationships among the input update rate, the rate of convergence of the Jacobi-SVD algorithm, and the quality of the SVD processed outputs. This approach provides new insights on the selection of forgetting factors needed in adaptive signal processing. We also obtain a real-time, nonparametric nonstationarity indicator of the observed data in terms of their singular value behavior. The proposed algorithm has been applied to problems in DOA estimation, speech segmentation, and linear prediction.

II. The Jacobi SVD Algorithm

Given the computed matrices U_m, Σ_m, V_m

- application of forgetting and vector projection
  $\Sigma'_m \leftarrow \beta\Sigma_m$; $x'_{m+1} \leftarrow x_{m+1}V_m$
  - QR updating
  $$\begin{pmatrix}\Sigma'_m \\ 0\end{pmatrix} \leftarrow Q^{\mathrm{H}}_{m+1}\begin{pmatrix}\Sigma'_m \\ x'_{m+1}\end{pmatrix}$$
  $$U'_{m+1} \leftarrow \begin{pmatrix}U_m & 0 \\ 0 & 1\end{pmatrix} Q_{m+1},\ Q_{m+1} \in \mathbf{C}^{(n+1)\times(n+1)},\ \text{orth.}$$
- Jacobi rotations (rediagonalization)
  $U_{m+1} \leftarrow U'_{m+1}$, $\Sigma_{m+1} \leftarrow \Sigma'_m$, $V_{m+1} \leftarrow V_m$
  for $k = 1, \ldots, \ell$; for $i = 1, \ldots, n-1$
  Apply a Jacobi rotation to rows and columns i and $i+1$ of Σ_{m+1}
  Propagate the rotations to U_{m+1} and V_{m+1}
  end

The QR and the Jacobi rotation steps can be implemented on a parallel/systolic architecture [3].

Variable Rotational Rate Scheme. For sufficiently slowly changing data, a slowly updating implementation of the Jacobi SVD algorithm produces the same (or better) estimates than a higher throughput implementation, for equal computational rate [1]. When the data variation increases, a higher updating rate with no computation rate increase produces computed singular matrices which are far from convergence. The idea we explored is to "decouple" the updating rate from the speed at which rotations are computed ("rotational rate").

[1]This work is supported by NASA/Dryden grant NCC 2-374.

Consider the QR factorization required by the updating algorithm, where Σ'_m is upper triangular. In order to give an estimate to the number of Jacobi rotations needed to diagonalize Σ'_m, what is of interest is the amount of fill-in in the submatrix of Σ'_m This is in turn related to the value $\alpha_{m+1} \equiv \|x_{m+1}V^{\mathrm{n}}_m\|/\|x_{m+1}\|$, $m = 0, 1, \ldots$, where $V_m = (V^{\mathrm{s}}_m, V^{\mathrm{n}}_m)$, and x_{m+1} is the incoming vector. The quantity α_{m+1} represents the degree of nonstationarity of the incoming data and is easily computed in the Jacobi algorithm.

From our analysis of the initial convergence behavior of the Jacobi SVD algorithm, as well as the behavior of the off-norm of Σ'_m in time, we propose the following variable rotational rate scheme, for medium to high SNR, noise power σ^2_{N}, and numerical rank $= r$:

1) Compare the nonstationarity indicator to a threshold μ, function of σ_{N}. 2) If $\alpha_{m+1} \leq \mu$, then choose a value for ℓ not smaller than r. Otherwise choose $\ell \geq n$. 3) Choose a high enough forgetting factor, which guarantees that the diagonal elements of Σ_m are sufficiently large (cf. below).

Variable Forgetting Scheme. We have also studied the relationship between the data variation an the forgetting factor. SVD tracking requires narrower observation windows, as the rate of data variation increases. If it is required that the number of Jacobi rotations which rediagonalize Σ' be kept low, then the amount of fill-in produced by the QRD step has to be limited. This is achieved by setting a minimum value for β. The proposed variable forgetting scheme is summarized as: 1) Determine at every time instant the duration of the stationarity window, N_{w}. 2) Given a threshold b, compute β so that $\beta^{N_{\mathrm{w}}} \leq b$. 3) Make sure that β is not too small, $\beta \geq \beta_{\min}$. Compute β as $\beta = \max\left\{b^{1/N_{\mathrm{w}}}, \beta_{\min}\right\}$.

The proposed SVD updating algorithm can find application in many situations, such as beamforming, adaptive filtering, DOA tracking, speech processing (segmentation, glottal closure detection), adaptive parameter estimation [1]. In all the cases considered, the algorithm promptly detects signal nonstationarities, whether in amplitude or phase. The ability to track data variability is exploited for real-time adaptation of the SVD updating algorithm, thereby producing more accurate estimation of singular values/subspaces. In certain applications, such as speech segmentation, the proposed algorithm can be used with the double function of detecting data transitions (voiced to unvoiced) and computing the desired filter parameters. This algorithm can be implemented on a parallel (systolic) processor with relative ease.

References

[1] F. Lorenzelli and K. Yao. "SVD Updating for Nonstationary Data". *IEEE VLSI Signal Proc.*, 1994.

[2] M. Moonen, P. Van Dooren, and J. Vandewalle. "A SVD Updating Algorithm for Subspace Tracking". *SIAM J. Matr. Anal. and Appl.*, October 1992.

[3] M. Moonen, P. Van Dooren, and J. Vandewalle. "A Systolic Array for SVD Updating". *SIAM J. Matr. Anal. and Appl.*, April 1993.

Improved LMS estimation via structural detection[1]

John Homer[1], Iven Mareels[2], Robert Bitmead[2], Bo Wahlberg[3], Fredrik Gustafsson[4]

1. Elect. & Comp. Eng. Dept., Univ. of Queensland, Brisbane, Qld 4072 Australia
2. Systems Eng. Dept., RSISE, Australian National Univ., Canberra, ACT 0200 Australia
3. S3-Automatic Control, Royal Inst. of Tech., S-100 44 Stockholm, Sweden
4. Automatic Control, Linköping Univ., S-581 83 Linköping, Sweden

Abstract — **We consider the LMS estimation of a channel that may be well approximated by an FIR model with only a few nonzero tap coefficients within a given delay horizon or tap length n. When the number of nonzero tap coefficients m is small compared to the delay horizon n, the performance of the LMS estimator is greatly enhanced when this specific structure is exploited. We propose a consistent algorithm that performs identification of nonzero taps only.**

I. Introduction

In various adaptive estimation applications, the unknown channel is characterized by an impulse response which consists of extended regions of negligible response or 'inactivity'. Examples include circuit echo paths within 4-wire loop telephony networks, which typically show initial inactive regions within their impulse responses, and room acoustic echo paths and mobile radio channels which typically show impulse responses having many inactive regions interspersed by 'active' or nonzero regions. Our aim is to develop a technique which discriminates between the active and inactive regions of such channels and to subsequently LMS estimate only the active regions of the channel.

II. System Description

Assumption 1: Unknown channel is linear, time invariant and is adequately modelled by a discrete-time FIR filter $\Theta(z^{-1})$ with a maximum lag of n sample intervals.
Assumption 2: Only $m < n$ of the taps of $\Theta(z^{-1})$ are nonzero.
Assumption 3: All signals are sampled. At sampling instant k: $u(k)$ is the signal input to the unknown channel and the channel estimator; an additive disturbance, $s(k)$, occurs within the unknown channel; and $y(k) = U(k)^T\theta + s(k)$ is the observed output from the unknown channel, where θ is the n tap unknown channel tap vector and $U(k)$ is the n tuple vector containing the last n input samples.
Assumption 4: (i) The input signal and disturbance signals are zero mean bounded wide sense stationary. (ii) The input and disturbance signals are uncorrelated with each other over time. (iii) The $n \times n$ input signal covariance matrix R is positive definite. (iv) The input signal is uncorrelated over time ('white').

III. Active Tap Detection

The aim is to determine the positions of the m nonzero elements of θ. The approach taken is to minimize the Least Squares cost function $V_N(\hat{\theta}(N)) \triangleq [\sum_{k=1}^{N}(y(k) - U(k)^T\hat{\theta}(N))^2]/N$ under the restriction that all but m elements of $\hat{\theta}$ are zero. In general, this requires the calculation and comparison of $V_N(\hat{\theta}(N))$ for $(n)!/[(m)!(n-m)!]$ different tap combinations. For signals $u(k)$ and $s(k)$ which satisfy the assumptions above, we can show that, for sufficiently large N, the LS cost function $V_N(\hat{\theta}(N))$ can be approximated by a cost function in which the contribution of each tap is decoupled from the rest. This leads to the following result.

Result 1 Subject to the validity of the assumptions 1-4, then, for sufficiently large N, the positions of the m most active taps of the FIR modelled unknown channel are given by the indices corresponding to the m greatest values of:
$X_N(j) \triangleq [\sum_{k=j+1}^{N} y(k)u(k-j)]^2]/[\sum_{k=j+1}^{N} u^2(k-j)]$.

To enable a tap of the unknown channel to be classed as 'active' or 'inactive', rather than just more or less active requires a threshold to be developed for the tap activity measure $X_N(j)$. This is achieved by considering a structurally consistent version of the LS cost function:
$W_N((\hat{\theta}(N)) = V_N(\hat{\theta}(N)) + Cmlog(N)/N$,
where m is the number of active taps to be determined, C is a constant independent of m, N. Through an extension of Result 1 above and application of work by Donoho cited in [1], we obtained the following result.

Result 2 Subject to the validity of the Assumptions 1-4, then, for sufficiently large N, the positions of the active taps of the FIR modelled unknown channel are given by the indices j for which: $X_N(j) > \sigma_y^2 logN$, where σ_y^2 is the variance of $y(k)$.

Simulations demonstrate that this tap activity criterion leads to fast detection of the active taps of the unknown channel

IV. LMS Estimation Via Detection

- Determine at each sample interval k the indices which satisfy the active condition $X_j(k) > \hat{\sigma}_y^2(k)\log(k)$, where $\hat{\sigma}_y^2(k)$ is an estimate of σ_y^2.
- For sample interval k (i) LMS update those taps in the LMS estimator which correspond to the active tap indices (of sample interval k); (ii) apply an exponentially decaying (forgetting) function to the remaining taps (corresponding to the inactive tap indices of sample interval k).

This structural detection LMS algorithm can be easily modified to obtain an NLMS version.

Simulations demonstrate that the structural detection LMS/NLMS algorithms provide considerably better asymptotic/transient performance, respectively, than the standard LMS/NLMS algorithms in which full parametrization is used.

[1]The authors wish to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Government under the Cooperative Research Centres Program

References

[1] H. Cramer and M.R.Leadbetter, *Stationary and Related Stochastic Processes: Sample Function Properties and their Applications*, Wiley, New York, 1967.

The Trellis Structure of Maximal Fixed-Cost Codes

Frank R. Kschischang

Department of Electrical and Computer Engineering, University of Toronto
10 King's College Road, Toronto, Ontario Canada M5S 1A4 (e-mail: frank@comm.utoronto.ca)

Abstract — **We show that the family of maximal fixed-cost codes, with codeword costs defined in a right-cancellative semigroup, have biproper trellis presentations. Examples of maximal fixed-cost codes include such "nonlinear" codes as permutation codes, shells of constant Euclidean norm in the integer lattice, and of course ordinary linear codes over a finite field. The intersection of two codes having biproper trellis presentations is another code with a biproper trellis presentation; therefore "nonlinear" codes such as lattice shells or words of constant weight in a linear code have biproper trellis presentations.**

I. Biproper Trellis Presentations

A proper trellis presentation for a block code C is one in which the set of edges emanating from any trellis vertex are labelled distinctly [1]. A "biproper" trellis is a proper trellis that remains proper when the direction of all trellis edges is reversed. Although not all codes have biproper trellis presentations [2], when a code has a biproper trellis presentation, the fortunate circumstance arises in which the biproper trellis simultaneously minimizes the vertex count at each time index, the trellis presentation is unique (up to relabeling), the trellis is one-to-one, and subtrellises are also biproper. In the language of dynamical systems theory [3], codes with biproper trellises have a well-defined and unique minimal state realization.

Let C be a code of length n, i.e., a set of n-tuples. The following conditions are equivalent:

1. C has a biproper trellis presentation.
2. C forms a rectangular relation [2] at each time index.
3. For fixed partition $(-,-)$ of codewords into "past" and "future" [3], if $(a,d) \in C$ and $(a,c) \in C$ and $(b,c) \in C$ implies $(b,d) \in C$ for all possible (not necessarily distinct) choices of a, b, c, d.
4. Any of six further equivalent conditions of Willems [4] as paraphrased by Forney and Trott [3, p. 1500].

Example. Let C be a group code regarded as a block code of length two, in which the coordinate p (resp., f) of a codeword (p,f) represents the entire past (resp., future) of that codeword. Suppose $c_1 = (a,d)$, $c_2 = (a,c)$, and $c_3 = (b,c)$ are codewords. The combination $c_1 c_2^{-1} c_3 = (a,d)(a^{-1},c^{-1})(b,c) = (b,d)$ must also be a codeword, hence the code satisfies condition (3). Since the split into past and future can be done at an arbitrary time index, the code has a unique minimal state realization at each time index.

The purpose of this paper is to introduce a wider class of codes that also satisfy the equivalent conditions listed above.

II. Maximal Fixed-Cost Codes

Let S be a right-cancellative semigroup, i.e., a semigroup in which $ax = bx$ implies $a = b$ for all $a, b, x \in S$. Let $\mathcal{A}$ be a product $A_1 \times A_2 \times \cdots \times A_n$ of symbol alphabets, and define "cost functions" $\mu_i : A_i \to S$ that associate an element of S with each symbol $a_i \in A_i$. Define the "cost" $\mu(\mathbf{a})$ of an n-tuple $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}$ as the product (in S) of coordinate costs, i.e.,

$$\mu(a_1, a_2, \ldots, a_n) = \mu_1(a_1)\mu_2(a_2)\cdots\mu_n(a_n).$$

Similarly, for a fixed partition of an element of $\mathcal{A}$ into past $p = (a_1, \ldots, a_i)$ and future $f = (a_{i+1}, \ldots, a_n)$, define $\mu(p) = \mu_1(a_1)\cdots\mu_i(a_i)$ and $\mu(f) = \mu_{i+1}(a_{i+1})\cdots\mu_n(a_n)$.

For a fixed cost $s \in S$, define the *maximal fixed-cost code* $M_s = \{\mathbf{a} \in \mathcal{A} : \mu(\mathbf{a}) = s\}$ to be the set of all possible n-tuples from $\mathcal{A}$ having cost s.

Theorem. M_s has a biproper trellis presentation.

Proof: For fixed partition of codewords into past and future, let (a,d), (a,c), and (b,c) be codewords in M_s. Then $\mu(a)\mu(c) = \mu(b)\mu(c)$. By right-cancellation, $\mu(a) = \mu(b)$; therefore, $\mu(a)\mu(d) = \mu(b)\mu(d) = s$. Since $\mu(b,d) = s$, and M_s is maximal, (b,d) is a codeword. Thus M_s satisfies condition (3).

III. Examples

1. Let $A_i = GF(q)$, let $H = [h_1, \ldots, h_n]$ be an $r \times n$ matrix with columns h_i having entries from $GF(q)$, and let $S = GF(q)^r$ be the vector space of r-tuples over $GF(q)$. For $1 \le i \le n$, define $\mu_i(x) = x \cdot h_i$. Then $\mu(\mathbf{a}) = \mathbf{a}H^T$, and M_0 is the linear code with parity-check matrix H.

2. Let $A_i = \{0,1\}$, let $S = \mathbb{N}_0$ be the monoid of non-negative integers under addition, and define μ_i by $\mu_i(0) = 0$; $\mu_i(1) = 1$. Then, for any block length n, M_w is the set of binary n-tuples of Hamming weight w.

3. Let $A_i = \mathbb{Z}$, let $S = \mathbb{N}_0$, and define μ_i by $\mu_i(x) = x^2$. Then, for any block length n, M_w is the set of integer n-tuples of squared norm w, i.e., a shell in the integer lattice $\mathbb{Z}^n$.

4. Let $A_i = \{a_1, \ldots, a_c\}$, let $S = \mathbb{N}_0^c$, the direct product of c copies of $\mathbb{N}_0$. Define μ_i by $\mu_i(a_i) = e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where e_i is the unit vector nonzero only in coordinate i. Then $M_{(m_1,\ldots,m_c)}$ is the permutation code obtained by permuting the vector of "shape" $(a_1^{m_1}, \ldots, a_c^{m_c})$ in all possible ways, where $a_i^{m_i}$ denotes the m_i-tuple $(a_i, a_i, \ldots, a_i)$.

References

[1] D. J. Muder, "Minimal trellises for block codes," *IEEE Trans. on Inform. Theory*, vol. 34, pp. 1049–1053, Sept. 1988.

[2] F. R. Kschischang and V. Sorokine, "On the trellis structure of block codes," *IEEE Trans. on Inform. Theory*. To appear.

[3] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, 1993.

[4] J. C. Willems, "Models for dynamics," in *Dynamics Reported, Volume 2* (U. Kirchgraber and H. O. Walther, eds.), pp. 171–269, John Wiley and Sons, 1989.

On Trellis Complexity of Block Codes: Optimal Sectionalizations

Alec Lafourcade-Jumenbo and **Alexander Vardy**

Coordinated Science Laboratory, University of Illinois, 1308 W. Main Street, Urbana, IL 61801, USA

{lafourca,vardy}@golay.csl.uiuc.edu

Abstract — **We present a polynomial-time algorithm which produces the optimal sectionalization of a given trellis T for a block code C in time $O(n^2)$, where n is the length of C. The algorithm is developed in a general setting of certain operations and functions defined on the set of trellises; it therefore applies to both linear and nonlinear codes, and accommodates a broad range of optimality criteria. The optimality criterion based on minimizing the number of operations required for trellis decoding of C is investigated in detail: several methods for decoding a given trellis are discussed and compared in a number of examples. Finally, analysis of the dynamical properties of optimal sectionalizations is presented.**

I. Introduction

It is now well-known [2, 3] that every linear block code may be represented by a trellis, which can be employed for maximum-likelihood decoding of the code with the Viterbi algorithm or variants thereof. The complexity of a given trellis is usually expressed in terms of parameters such as the number of states and/or branches it contains. While, indeed, these parameters govern the complexity of trellis decoding, in many cases this complexity may be reduced with an appropriate sectionalization of the trellis. By a sectionalization we mean the choice of the symbol alphabet at each time index: for a given order of the time axis $\mathcal{I}$, the sectionalization shrinks $\mathcal{I}$ at the expense of increasing the code alphabet [2]. A wide variety of such granularity adjustments is possible, and each may substantially affect the decoding complexity. For a given code C of length n and a given order of its time axis $\mathcal{I}$, a specific sectionalization of its trellis T is determined by the set $\{h_0, h_1, \ldots h_\nu\} \subset \mathcal{I}$ of section boundaries, where $h_0 = 0 < h_1 < \cdots < h_\nu = n$. Clearly, there are 2^{n-1} possible ways to select the section boundaries, and the *sectionalization problem* consists of finding the optimal choice among the 2^{n-1} possibilities. Examples of specific 'good' sectionalizations for particular codes may be found in [2, 3] among other works. However, at the present time, finding the best sectionalization is more akin to 'art' than to exact science: no systematic method for finding the optimal sectionalization of a given trellis is presently known.

II. The Sectionalization Algorithm

In this work, we present a complete solution to the general sectionalization problem. Namely, we describe a polynomial-time algorithm which produces the optimal sectionalization from a given generator matrix of the code. The algorithm is developed in a general setting of certain operations and functions on the set of trellises. In particular, we generalize to some extent the usual definition of a trellis. We then define the operations of composition and amalgamation of trellises. This enables us to consider a class of functions defined on the set of trellises, that satisfy a certain linearity property with respect to the composition operation. We then seek a sequence of amalgamations and compositions that minimize the value of an arbitrary given function from this class. We show that finding such a sequence is equivalent to finding the minimum-weight path in a certain weighted digraph. This may be accomplished using the well-known Dijkstra algorithm [1]. Thus, to find the sectionalization of a given trellis T which minimizes the value of an arbitrary given function $F(T)$, we construct the corresponding *sectionalization digraph* $\mathcal{G}$, and then apply a variant of Dijkstra's algorithm to $\mathcal{G}$.

III. Examples

The general sectionalization algorithm described above applies to both linear and nonlinear codes and easily accommodates a broad range of optimality criteria. However, herein, we focus on the optimality criterion based on the total number of real additions and comparisons required for decoding the trellis. This criterion conforms to the well-established tradition of counting decoding complexity [2, 3]. For instance, for the $(24, 12, 8)$ binary Golay code C_{24}, we obtain the sectionalization with boundaries at $0, 8, 16, 24$, which coincides with the one given by Forney [2] and proves that this sectionalization is indeed optimal for trellis decoding of C_{24}. The number of decoding operations we obtain for this sectionalization is 1339, which is slightly less than the number reported by Forney in [2], and considerably less than the complexity of the earlier algorithms. Notably, all the previous Golay decoders have been specifically 'tailored' for C_{24}, whereas our decoder is the output of a general-purpose computer program which applies uniformly well to *any* linear code. Other examples include Reed-Muller codes, BCH codes, Shearer codes, and quadratic-residue codes. In particular, for all the primitive BCH codes of length ≤ 64 we improve upon the decoding complexities reported in [3].

IV. Dynamics of Optimal Sectionalizations

Although our algorithm readily produces the optimal sectionalization, it provides little insight as to how it relates to the dynamical properties of the code. Thus, we also investigate, under certain simplifying assumptions, the dynamics of optimal sectionalizations. For instance, we show that if the dimensions of both past and future subcodes change at a given position $i \in \mathcal{I}$, then i is necessarily a section boundary in the optimal sectionalization. Furthermore, in any section of the optimal sectionalization the dimension of either the past or the future or both must be constant.

*This work was supported by the NSF Grant NCR-9501345

References

[1] E.W. Dijkstra, "A note on two problems in connection with graphs," *Numerische Math.*, vol. 1, pp. 269–271, 1959.

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

[3] A. Vardy and Y. Be'ery, "Maximum-likelihood soft decision decoding of BCH codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 546–554, 1994.

On Trellis Complexity of Block Codes: Lower Bounds

Alec Lafourcade-Jumenbo and **Alexander Vardy**

Coordinated Science Laboratory, University of Illinois, 1308 W. Main Street, Urbana, IL 61801, USA

{lafourca,vardy}@golay.csl.uiuc.edu

Abstract — **We present a new lower bound on the state-complexity of linear codes, which includes all the existing bounds as special cases. For a large number of codes this results in a considerable improvement upon the DLP bound. Moreover, we generalize the new bound to nonlinear codes, and introduce several alternative techniques for lower bounding the trellis complexity, based on the distance spectrum and other combinatorial properties of the code. We also show how our techniques may be employed to lower bound the maximum and the total number of branches in the trellis. The asymptotic behavior of the new bound is investigated and shown to improve upon the known asymptotic estimates of trellis complexity.**

I. Introduction

The trellis state-complexity of a linear code C over GF(q) is defined as $s = \max_{i\in\mathcal{I}} \{\log_q |S_i|\}$, where S_i is the set of states at time $i \in \mathcal{I}$ in the minimal trellis for C. Perhaps the earliest known lower bound on s is due to Muder [4]: for an (n,k,d) linear code $s \geq k - \min_{i\in\mathcal{I}} \{K(i,d) + K(n-i,d)\}$, where $K(n,d)$ is the largest possible dimension of a linear code of length n and minimum distance d. This bound was improved by several authors, giving $s \geq k - \min_{i\in\mathcal{I}} \{k(i;C) + k(n-i;C)\}$, where $k(i;C)$ is the dimension-length profile (DLP) of C, i.e., the maximum dimension of any subcode of C of support size i. All these bounds are based on the common idea of dividing the time axis for the code into two sections — the past and the future, and then bounding the dimension of the resulting state-space using any of the known upper bounds on the dimension of the past and future subcodes. In [3] we have recently derived a conceptually different bound $s \geq \lceil k(d-1)/n \rceil$, based on dividing the time axis into $\lceil n/(d-1) \rceil$ sections, and using the fact that there can be no parallel transitions in a trellis section of length less than d.

II. Lower Bound on State Complexity

In this work we present a new lower bound on s, which includes all the existing bounds as special cases. The new bound is obtained by partitioning the time axis for C into several – that is, generally more than two – sections and then selecting the partition which yields the best lower bound.

Theorem 1. *Let $l_1, l_2, \ldots l_L$ be any set of positive integers, with $l_1 + l_2 + \cdots + l_L = n$. Then*

$$s \geq \left\lceil \frac{k - k(l_1;C) - k(l_2;C) - \cdots - k(l_L;C)}{L-1} \right\rceil .$$

For great many codes this results in a substantial improvement upon the DLP bound. We have applied the proposed technique to all the 8128 best known binary linear codes of length ≤ 128, and obtained over 3400 improvements over the DLP bounds. For a complete summary of our results, send e-mail to trellis@golay.csl.uiuc.edu.

*This work was supported by the NSF Grant NCR–9409688

III. Bounds on Branch and Edge Complexity

Trellis branch-complexity was defined by Forney in [1] as $b = \max_{i\in\mathcal{I}} b_i$, where b_i is the logarithm of the number of branches in the trellis section corresponding to time $i \in \mathcal{I}$. The state-complexity bounds of Theorems 1 and 2 can be translated into a lower bound on b, which is often much tighter than the obvious statement $b \geq s$.

Theorem 3. *Let $l_1, l_2, \ldots l_L$ be positive integers, such that $l_1 + l_2 + \cdots + l_L = n - L + 1$. Then*

$$b \geq \left\lceil \frac{k - k(l_1;C) - k(l_2;C) - \cdots - k(l_L;C)}{L-1} \right\rceil .$$

The new bound on branch-complexity was again applied to all the best-known binary linear codes of length ≤ 128, yielding over 3300 improvements over the DLP bound. Notably, in 2621 out of the 3300 cases, the lower bound on b is strictly greater than the lower bound on s. In addition, we derive lower bounds on the *total* number of branches in the trellis — the trellis edge-complexity $E(C)$ as defined in [2]. The bounds follow by solving a nonlinear integer programming problem with linear constraints, which arise from the general relations between the values of b_i derived in the proof of Theorem 3.

IV. Asymptotics

As shown in [3], for a sequence of codes of increasing length n, with rate fixed at R and relative minimum distance fixed at $d/n = \delta$, the state-complexity is bounded by $c_1 n \leq s \leq c_2 n$ for some constants c_1 and c_2 independent of n. The results of [3] establish $c_1 \geq \delta R$, while the work of [5] shows that $c_1 \geq R - R_{\max}(2\delta)$, where $R_{\max}(\cdot)$ is the function describing the JPL upper bound. Herein we prove

Theorem 4. *Let $\varsigma = s/n$. Then for $n \to \infty$ and for all $L = 2, 3, \ldots,$*

$$\varsigma \gtrsim \frac{R - R_{\max}(L\delta)}{L-1}$$

The theorem produces a countably infinite family of lower bounds on ς, and it is easy to see that the apparently dissimilar bounds of [3] and [5] are in fact the extreme members of this family corresponding to $L = 2$ and $L \simeq 1/\delta$, respectively.

References

[1] G.D. Forney, Jr., "Dimension/length profiles and trellis complexity of linear block codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1741–1752, 1994.

[2] A.B. Kiely, S. Dolinar, R.J. McEliece, L. Ekroot, and W. Lin, "Trellis decoding complexity of linear block codes," preprint.

[3] A. Lafourcade and A. Vardy, "Asymptotically good codes have infinite trellis complexity," *IEEE Trans. Inform. Theory*, vol. 41, pp. 555–559, 1995.

[4] D.J. Muder, "Minimal trellises for block codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1049–1053, 1988.

[5] V.V. Zyablov and V.R. Sidorenko, "Bounds on complexity of trellis decoding of linear block codes," *Problemy Peredachi Informatsii*, vol. 29, pp. 3–9, 1993, (in Russian).

Trellis Complexity Versus The Coding Gain Of Lattices.

Vahid Tarokh and I.F. Blake [1]
Elect. & Comp. Eng. Dept., Univ. of Waterloo,
Waterloo, Ontario, Canada

Abstract — **The growth of trellis diagrams of lattices versus their coding gain is studied. It is established that this growth exponentially in terms of the coding gain.**

I. Introduction

The issues of trellis complexity have recently attracted wide attention. In this direction, many authors have studied the relations between trellis complexity, the minimum distance and the dimension of linear block codes. This work reports a parallel development for lattices.

II. Preliminaries

Let $\mathcal{L}$ denote the set of all lattices having a finite trellis diagram. Let $L \in \mathcal{L}$ and n denote the dimension of L. Let $\mathcal{C}(L)$ denote the category of all the finite trellis diagrams for L, then $\mathcal{C}(L)$ is nonempty. Let S and B denote the minimum number of states and branches, respectively, of elements of $\mathcal{C}(L)$. Consider the sum of the cardinality of the label groups for each element of $\mathcal{C}(L)$ and let G denote the minimum of these sums in $\mathcal{C}(L)$. Define $\mathcal{S}(L)$, *the average state trellis complexity* of L, to be $(S-1)/n$ and $\mathcal{B}(L)$, *the average branch trellis complexity* of L to be B/n. Define $\mathcal{G}(L)$, *the average label group complexity* of L to be G/n. For any lattice L, let $\delta(L)$ denote the coding gain of L. Then

$$\begin{aligned} \mathcal{T}_1(\gamma) &= inf\{\mathcal{S}(L) \mid \delta(L) \geq \gamma \text{ and } L \in \mathcal{L}\}, \\ \mathcal{T}_2(\gamma) &= inf\{\mathcal{B}(L) \mid \delta(L) \geq \gamma \text{ and } L \in \mathcal{L}\}, \\ \mathcal{T}_3(\gamma) &= inf\{\mathcal{G}(L) \mid \delta(L) \geq \gamma \text{ and } L \in \mathcal{L}\}, \end{aligned}$$

referred to as the *state trellis complexity*, the *branch trellis complexity* and the *label trellis complexity* functions respectively.

Since $\mathcal{T}_i$, $i = 1, 2, 3$, represent the best trade-off between trellis complexity and gain, it is essential to establish bounds on the behavior of these functions.

In [1] these results are established.

1: $\mathcal{T}_i = 1 + C_i ln(\gamma)$ for C_i, $i = 1, 2, 3$ constants, whenever the coding gain is close to 1.

2: $\mathcal{T}_1$ and $\mathcal{T}_2$ grow exponentially when γ is large.

3: $\mathcal{T}_3$ grows at most linearly.

4: $\mathcal{T}_1(\gamma) \geq (\gamma/\gamma_r)^{r/2}, r = 1, 2, 3, \ldots$, where γ_r denotes the coding gain of the densest lattice in r dimensions.

5- $\mathcal{T}_2(\gamma) \geq \gamma^{(r+1)/2}/\gamma_r^{r/2}, r = 1, 2, 3, \ldots$

6- $\mathcal{T}_1^2$ is bigger than any of the DLP bounds evaluated at γ.

7- A random coding argument was then applied to show that the bounds given above cannot be much improved.

The above results imply that the Viterbi algorithm, applied to the trellis diagrams of lattices, have exponential running time.

[1]This work was supported by the National Sciences and Engineering Research Council of Canada grant number A7382.

Acknowledgements

We are grateful to Dr. G. D. Forney for valuable comments and suggestions.

References

[1] Vahid Tarokh, "Trellis Complexity versus the Coding Gain of Lattice-Based Communication Systems" *Ph.D. Thesis, The University of Waterloo.* Waterloo, Ontario, Canada. 1995.

Rotationally Invariant, Punctured Trellis Coding[1]

Eric J. Rossin, David J. Rowe and Chris Heegard[2]
School of Electrical Engineering, Cornell Univ., Ithaca, New York, USA

Abstract — **In this paper we present a systematic method for combining rotational invariance and punctured codes for use with TCM, and discuss a new perspective on the class of punctured codes.**

I. Introduction

Trellis coded modulation (TCM) based QAM systems, such as those found in the V.32 and V.34 telephone modem standards, incorporate rotational invariance to provide benefits with respect to absolute phase reference and phase noise protection. In binary convolutional code based QPSK systems, the computational savings and code rate flexibility of punctured coding is well known and used to great advantage. However, these systems do not maintain rotational invariance and suffer from the lack of this property.

Until recently, a systematic method of combining both rotational invariance and puncturing in a general framework was unknown. In this paper we present a rotationally invariant encoding/uncoding structure that can use punctured codes. Parts of this work are related to [4, 5].

II. Background

Rotationally invariant (RI) trellis codes are important whenever the modulation signal set has a two-dimensional rotational symmetry and the transmission system can introduce a phase ambiguity at the receiver. A trellis code is RI if the componentwise rotation of a code sequence is always another code sequence in the code (cf., [1]). RI trellis codes with RI encoders/uncoders are highly desirable as a method of handling 90° phase ambiguities as they have the property that the output of the uncoder for any codeword is the same as the output when the codeword is first rotated by 0°, 90°, 180° or 270° before being presented to the uncoder.

A punctured convolutional code is a high-rate code obtained from a lower-rate code by periodically eliminating, i.e., "puncturing" specific symbols from the lower-rate codeword [2, 3]. The resulting punctured code depends on both the original code, and on the number and locations of the symbols to be deleted.

III. Trellis Coding

This work describes a method of encoding, using *any* transparent binary convolutional code (BCC), that results in a RI trellis code for applications to QPSK and QAM modulation. This method incorporates three components: (1) a transparent BCC (2) a 2 dimensional signal space labeling and (3) a precoding/postcoding function.

Transparent codes and transparent encoders/uncoders are highly desirable as a method of handling 180° phase ambiguities. A BCC is said to be *transparent* if the compliment of any codeword is always a codeword. Every transparent code has a transparent encoder/uncoder with the property that even if the codeword is complemented the uncoder will produce the correct sequence.

The QPSK and QAM signal sets need to be labeled in such a way that: (1) the two least significant bits, $(I_j,\ Q_j)$, satisfy $(I_j,\ Q_j) \rightarrow (\bar{Q}_j,\ I_j)$ under 90° rotation and (2) the remaining most significant bits are invariant to 90° rotation.

The following mapping describes the required precoder/ postcoder structure. Precoder equations:

$$\begin{aligned} x_j &= w_j + x_{j-1} + z_j(x_{j-1} + y_{j-1}), \\ y_j &= z_j + w_j + y_{j-1} + z_j(x_{j-1} + y_{j-1}). \end{aligned}$$

Postcoder equations:

$$\begin{aligned} w_j &= x_j + y_{j-1} + (x_j + y_j)(x_{j-1} + y_{j-1}), \\ z_j &= y_j + x_j + y_{j-1} + x_{j-1}. \end{aligned}$$

Note that: (1) the postcoder inverts the precoder, (2) the output of the postcoder is the same under the map $(x_j, y_j) \rightarrow (\bar{y}_j,\ x_j)$ (or any integer power of this map), (3) the postcoder function is feedback free and thus limits error propagation.

In the encoder, the two binary outputs of the precoder are independently encoded with separate transparent BCC encoders. The BCC outputs, along with the remaining uncoded (or parallel edge) information, are combined to select the QAM constellation point to be transmitted. The mapping is such that (1) the BCC outputs independently select the LSB of the I and Q components and (2) parallel edge information is RI.

By using the above encoding method with transparent punctured BCCs, the resulting structure is a punctured, rotationally invariant trellis code. The cost of this technique is a doubling of data memory in the viterbi decoder.

We will also present an alternative view of punctured codes from the perspective of multirate systems.

[1]This work was supported in part by NSF grant NCR-9207331 and by the United States Army Research Office through the Army Center of Excellence for Symbolic Methods in Algorithmic Mathematics (ACSyAM), Mathematical Sciences Institute of Cornell University, Contract DAAL03-91-C-0027.

[2]er14@cornell.edu, rowe@ee.cornell.edu, heegard@ee.cornell.edu

References

[1] 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, Sept. 1984.

[2] J. B. Cain, G. C. Clark Jr., and J. M. Geist, "Punctured convolutional codes of rate (n-1)/n for simplified maximum likelihood decoding," *IEEE Transactions Information Theory*, vol. IT–25, pp. 97–100, January 1979.

[3] D. Haccoun and G. Begin, "High-rate punctured convolutional codes for viterbi and sequential decoding," *IEEE Transactions on Communications*, vol. COM–37, pp. 1113–1125, November 1989.

[4] C. Heegard, S. A. Lery, and W. H. Paik, "Practical coding for QAM transmission of HDTV," *IEEE Journal on Selected Areas in Communications*, vol. SAC–11, January 1993.

[5] J. K. Wolf and E. Zehavi, "p^2 codes: Pragmatic trellis codes utilizing punctured convolutional codes," in *Proc. of the sixth Tirrentia Workshop on Digital Comm.*, (Tirrentia, Italy), Sept. 1993.

Trellises with parallel structure for block codes with constraint on maximum state space dimension

H.T. Moorthy, S. Lin and G. Uehara[1]
Dept. Electrical Eng., 2540 Dole Street, 483,
University of Hawaii, Honolulu, Hawaii, USA

Abstract — **This summary outlines certain results about trellis structures of linear block codes that achieve the highest speed of decoding while satisfying a constraint on the structural complexity of the trellis in terms of the maximum number of states at any particular depth. An upper bound on the number of parallel isomorphic subtrellises in a proper trellis for a code without exceeding the maximum state space dimension of the minimal trellis of the code is derived. The complexity of VLSI implementation of a Viterbi decoder based on an L-section trellis diagram for a code is analyzed and certain descriptive parameters are introduced. It is shown that a VLSI chip Viterbi decoder based on a non-minimal trellis requires less area and is capable of operation at higher speed than one based on the minimal trellis when the commonly used ACS-array architecture is considered.**

I. Introduction

Much effort has been expended on minimizing the number of states in a trellis for a block code by considering all possible permutations of the bits of the code and also on minimizing the number of operations required to decode a received vector using a trellis for the code. If decoding is performed using a stored program that is executed sequentially, then this approach will lead to the fastest speed of decoding. However, if decoding is performed using a VLSI chip, then **the above approach fails** and an alternative approach is more suitable. Given a constraint on the amount of hardware (determined by the number of states and the complexity of branches) in the decoder, **decoding must be done as fast as possible; not with as few computations as possible.** In [3], we have derived properties of the structure of this non-minimal trellis which show that a **non-minimal trellis implementation requires less area in the VLSI chip** than the minimal trellis implementation when the prominent ACS-array architecture is assumed [2].

II. Constrained Parallelism

We show how to build a trellis for a linear block code which is a disjoint union of certain desired number of parallel isomorphic subtrellises. *Although this trellis is not minimal, its state space dimension at every depth is less than or equal to the maximum state space dimension of the minimal trellis.* Let $\{s_0, s_M, \ldots, s_{LM}\}$ denote the state dimension profile (SDP) of the L-section, M bits/section minimal trellis of a (N, K) linear block code $\mathbf{C}$ and $s_{\max,L}(\mathbf{C})$ be the largest among them. Let $\mathbf{G}$ be the trellis oriented generator matrix of an (N, K) linear block code $\mathbf{C}$ [1]. Let $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ be a typical row of $\mathbf{G}$. Then, we define **the span of r**, denoted $span(\mathbf{r})$, to be the smallest interval $[i, j]$, $1 \leq i \leq j \leq N$ which contains all the non-zero elements of $\mathbf{r}$. For a row $\mathbf{r}$ whose span is $[i, j]$ we also define an **active span of r**, denoted $aspan(\mathbf{r})$, as $[i, j-1]$ if $i < j$ and $aspan(\mathbf{r}) = \phi$ if $i = j$. Define the non-empty set,

$$I_{\max}(\mathbf{C}) = \{l : s_l(\mathbf{C}) = s_{\max,L}(\mathbf{C})\} \quad (1)$$

Let $R(\mathbf{C})$ be the following subset of rows of $\mathbf{G}$,

$$R(\mathbf{C}) = \{\mathbf{r} \in \mathbf{G} : aspan(\mathbf{r}) \supseteq I_{\max}(\mathbf{C})\} \quad (2)$$

Let $d = |R(\mathbf{C})|$ where $|Q|$ denotes the cardinality of any finite set Q.

Theorem 1: With $R(\mathbf{C})$ defined as above and $d = |R(\mathbf{C})|$, let $1 \leq d' \leq d$. There exists a subcode $\mathbf{C}'$ of $\mathbf{C}$ such that $s_{\max,L}(\mathbf{C}') = s_{\max,L}(\mathbf{C}) - d'$ and $\dim(\mathbf{C}') = \dim(\mathbf{C}) - d'$ if and only if there exists a subset $R' \subseteq R(\mathbf{C})$ consisting of d' rows of $R(\mathbf{C})$ such that for every l satisfying $s_l(\mathbf{C}) > s_{\max,L}(\mathbf{C}')$, there exist at least $s_l(\mathbf{C}) - s_{\max,L}(\mathbf{C}')$ rows in R' whose active spans contain l. The set of coset representatives $[\mathbf{C}/\mathbf{C}']$ is generated by R'.

The **utility of the above theorem** is that it shows how to choose a subcode $\mathbf{C}'$ of $\mathbf{C}$ with $s_{\max,L}(\mathbf{C}') = s_{\max,L}(\mathbf{C}) - \dim([\mathbf{C}/\mathbf{C}'])$, such that a non-minimal trellis $\mathcal{T}$ for $\mathbf{C}$ with maximum state space dimension $s_{\max,L}(\mathbf{C})$ and which is the union of $2^{\dim[\mathbf{C}/\mathbf{C}']}$ parallel subtrellises $\mathcal{T}_i$ each isomorphic to the minimal trellis for $\mathbf{C}'$ can be built. **Upper Bound on Parallelism:** The smallest such subcode has dimension lower bounded by $\dim(\mathbf{C}) - |R(\mathbf{C})|$. i.e., the maximum number of parallel subtrellises one can obtain with the constraint that the total state space dimension never exceed $s_{\max}(\mathbf{C})$ is upper bounded by $2^{|R(\mathbf{C})|}$ with $R(\mathbf{C})$ as defined above. **Parallelism of the Minimal Trellis:** The logarithm to the base 2 of the number of parallel isomorphic subtrellises in a minimal L-section trellis for a binary (N, K) linear block code is given by the number of rows in its trellis oriented generator matrix whose active span contain the integers $\{M, 2M, \ldots, (L-1)M\}$ where $N = LM$.

[1]This research was supported by NSF Grant NCR-9115400 and NASA Grant NAG 5-931.

References

[1] G.D. Forney,Jr., "Coset codes II: Binary lattices and related codes," *IEEE Transactions on Information Theory*, vol. 34, pp. 1152-1187, 1988.

[2] P.G. Gulak and T. Kailath, "Locally Connected VLSI Architectures for the Viterbi Algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 6, April 1988.

[3] H.T. Moorthy et. al., "Good Non-minimal Trellises for Linear Block Codes," *submitted to IEEE Transactions on Communications*, Feb 1995.

Reconfigurable Trellis Decoding of Linear Block Codes

Alan D. Kot and Cyril Leung

Abstract — A class of methods for soft-decision decoding of linear block codes, referred to as *Reconfigurable Trellis* (RT) decoding, is presented. In RT decoding a reduced trellis (or tree) search is facilitated by carrying out the search on a reconfigured trellis (or tree) that corresponds to an equivalent code. The equivalent code is formed by reordering the received symbols according to their reliabilities. Consequently, the trellis reconfiguration is determined 'on-the-fly', but only a small portion of the trellis needs to be constructed, as guided by the reduced search. The search efficiency improves for channels where the soft-decisions provide a good indication of which symbols are in error. For example, using the M algorithm on an erasure channel, only a single survivor (i.e. $M = 1$) is sufficient to attain maximum-likelihood decoding of maximum-distance codes. For more typical channels, we present simulation results and a detailed assessment of the number of metric and binary-vector operations for the M algorithm.

Summary

We discuss a class of methods for soft-decision decoding of linear block codes that utilize *reconfigured* trellises (or trees). By a *reconfigured* trellis, we mean that the trellis used for decoding corresponds to an equivalent code obtained by a reordering of the symbol positions to exploit their differing reliabilities. Some recent works also utilize reduced searches on specially generated trellises or trees [1][2][3] (see also [4]). Of these works, [1][2] do not reconfigure the code trellis. The work reported herein, while it uses a similar trellis reconfiguration to that of [3], was developed independently [5][6]. Both [2] and [3] focus on ML decoding, while here we emphasize that trellis reconfiguration may facilitate many types of reduced searches, and concentrate in particular on the M algorithm.

The number of branches explored during reduced searches of trellises is decreased by exploring paths that are most likely to be part of the maximum-likelihood path (MLP), while discarding those paths that are unlikely to belong to the MLP as early in the search as possible. The key observation is that few branches would need to be explored if the *rank order* of path metrics rapidly converged with depth to their final values. In other words, a reduced search algorithm could stop any further exploration of a path *relatively early in the search*, without losing the MLP, if the influence of the unexplored branch metrics on the rank order of the path metrics was insignificant. Since reliable symbol-positions have one symbol-hypothesis that is much more likely than its alternatives, and since unreliable symbol-positions have little distinction between alternative symbol-hypotheses, reconfiguring the symbol positions in a most reliable symbol first (MRSF) manner should increase the rate of convergence with depth of the rank order of path metrics. In other words, using MRSF ordering should enable the path exploration to rapidly gather and utilize the most significant branch-likelihood information, regardless of the type of reduced search being used.

Portions of this work were supported by a Natural Sciences and Engineering Research Council (NSERC) Postgraduate Scholarship, a British Columbia Science Council G.R.E.A.T. Scholarship, a British Columbia Advanced Systems Institute Scholarship, and by NSERC Grant OGP–1731.

Alan Kot was with the Dept. of Electrical Engineering, University of British Columbia, Vancouver, Canada. He is now with Sharp Laboratories of America Inc., 5700 N.W. Pacific Rim Blvd., Camas, Washington 98607. Cyril Leung is with the Dept. of Electrical Engineering, University of British Columbia, 2356 Main Mall, Vancouver, B.C., Canada V6T 1Z4.

RT decoding will tend to collect channel errors into a burst in the later depths (tail) of the trellis, thus 'trapping' many errors. That is, many errors can fall in the parity symbol positions in the tail, and for such positions there is only one branch leaving each node, which constrains the search while conveniently ignoring the errors. The number of errors that will be trapped by RT decoding depends on how accurately the soft-decisions indicate the error positions. For example, if we are fortunate enough to have a channel that is extremely well approximated by an erasure channel, then RT decoding will collect all erasures in the tail. In the case of a maximum–distance code on an erasure channel, if there are $n - k$ or fewer erasures they will all be in the final $n - k$ positions of the reconfigured trellis, and these final $n - k = d_{min} - 1$ positions will hold parity symbols. This leaves a correct information set from which the correct codeword will be formed, thus ensuring that *ML decoding can be attained by retaining only one survivor*, regardless of the size of the code. Similarly, for non-maximum–distance codes, we can be assured to correct any pattern of $d_{min} - 1$ errors with only one survivor.

We focus on suboptimal soft-decision decoding to explore the trade-off between the coding gain attained and the computational effort expended. A 'near-ML' decoder can be very efficient while having a loss in decoding performance that is negligible in practice. For example, the extended Golay (24,12) code on an AWGN channel can be RT decoded, using the M algorithm, to within 0.25 dB of ML decoding with only 8 survivors. Tables are presented that summarize the number of metric and binary-vector operations for the M-algorithm.

Finally, we comment that some other important factors contribute to the efficiency of RT decoding. First, the reduced trellis exploration is facilitated by the use of a simplified trellis construction method [7]. Second, the efficiency of the RT-M algorithm is enhanced by the use of a survivor selection method that operates in linear-time [8], needing at most M comparisons at each depth.

References

[1] B. Radosavljevic, E. Arikan, and B. Hajek, "Sequential decoding of low-density parity-check codes by adaptive reordering of parity checks," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 1833–1839, Nov. 1992.

[2] N. C. J. Lous, P. A. H. Bours, and H. C. A. van Tilborg, "On maximum likelihood soft-decision decoding of binary linear codes," *IEEE Trans. Inf. Theory*, vol. IT-39, pp. 197–203, Jan. 1993.

[3] Y. S. Han, C. R. P. Hartmann, and C. C. Chen, "Efficient priority-first search maximum-likelihood soft-decision decoding of linear block codes," *IEEE Trans. Inf. Theory*, vol. IT-39, pp. 1514–1523, Sept. 1993.

[4] Y. S. Han, C. R. P. Hartmann, and C. C. Chen, "Efficient maximum-likelihood soft-decision decoding of linear block codes using algorithm A*," Tech. Rep. SU-CIS-91-42, School of Computer and Information Sciences, Syracuse Univeristy, Dec. 1991.

[5] A. D. Kot, *Reconfigurable Trellis Decoding*. Ph.D. Research Proposal, Dept. of Electrical Eng., University of British Columbia, Dec. 1988.

[6] A. D. Kot, *On the Construction, Dimensionality, and Decoding of Linear Block Code Trellises*. Ph.D. Thesis, Dept. of Electrical Eng., University of British Columbia, Dec. 1992.

[7] A. D. Kot and C. Leung, "On the construction and dimensionality of linear block code trellises," *Submitted for publication*, 1995.

[8] A. D. Kot, "A linear-time method for contender sifting in breadth-first decoding of error control codes," in *Canadian Conference on Electrical and Computer Engineering, Vancouver, Canada*, Sept. 1993.

On the Twisted Squaring Construction, Symmetric-Reversible Designs and Trellis Diagrams of Block Codes

Yuval Berger and Yair Be'ery

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

Abstract - **The structure of the twisted squaring construction is studied. We focus on the subclass of symmetric-reversible codes and show that it includes the extended primitive BCH codes. New results on the trellis complexity of these constructions, and the BCH codes in particular, are derived.**

I. Introduction

Trellis diagrams are primarily used for efficient soft-decision decoding [1]-[3]. The structure of the codes is a fundamental key for investigating the associated trellis diagrams [2],[4]-[7]. The *squaring construction* (SC) was employed by Forney [2] to derive trellis-oriented designs, particularly applied for RM codes and Barnes-Wall lattices. We are interested in the *twisted squaring construction* (TSC), a generalization of the SC [2]. The Nordstrom-Robinson code and a related packing are known examples of nonlinear TSC [2]. We classify several families of the TSC and focus on the *symmetric-reversible codes*. We show that they include the extended primitive BCH codes. The constructions are characterized and new results on the related trellis diagrams are developed.

II. The Twisted Squaring Construction

We follow the notations of [2]. Let S/T denote the partition of a discrete set S into $M=|S/T|$ disjoint subsets T_i, $i=0,1,\cdots,M-1$. The *minimum distance* $d(S)$ is defined as the minimum nonzero *distance* $d(s_1,s_2)$ associated with any pair $(s_1,s_2)\in S$. We also define $d(T)$ as the minimum $d(T_i)$ among the subsets of S. Let T_i^2 denote the set of all pairs (s_1,s_2) where $s_1,s_2\in T_i$. The SC is the union U of the M sets T_i^2, and $d(U)=\min\{d(T),2d(S)\}$ [2]. Let $C(n,k)$ denote a linear code over GF(q) with length n and dimension k. Let D be a subcode of C. The SC is labeled by $|C/D|^2$. It consists of codewords (d_1+b,d_2+b), where $d_1,d_2\in D$ and b belongs to the space $B=[C/D]$ of cosets representatives of D in C. The TSC is the union W of M sets T_iT_j, where i and j cover all values between 0 and M. The lower bound $d(W)\geq\min\{d(T),2d(S)\}$ [2] suggests an improvement of the TSC over the SC. The TSC in terms of linear codes will be labeled by $\|C/D\|^2$. It consists of codewords (d_1+b,d_2+b'), where b and b' run through all elements of B. Let G_C and G_D denote the generator matrices of C and D, respectively. The generator matrix of $\|C/D\|^2$ is equivalent to

$$\begin{pmatrix} G_C & \widetilde{G}_C \\ 0 & G_D \end{pmatrix},$$

where $\widetilde{G}_C$ is obtained from G_C by elementary row operations.

III. Symmetric-Reversible and Related Codes

A code A is called *symmetric* if $(a_1,a_2)\in A$ implies that $(a_2,a_1)\in A$. We show that any symmetric code is a TSC, and $\widetilde{G}_C=EG_C$ such that E is invertible and E^2 is equivalent to the identity matrix. A code is called *reversible* if it contains the reversed version of every codeword. A symmetric-reversible (SR) code is hereby defined as a code that is both symmetric and reversible. We show that part of the above properties are inherited to the subcodes C and D. A code is called *affine-invariant* (AI) if it is invariant under the affine permutation. This class includes the Reed-Muller (RM) and extended primitive BCH codes. We prove that AI codes are iterated SR codes (and obviously iterated TSC), i.e., the subcodes C and D are also SR codes. We characterize the constructions and show that the dual TSC and dual SR designs are, respectively, TSC and SR designs.

IV. Trellis Complexity

A general description of trellis diagrams of block codes is given in [1],[2]. For a given coordinate ordering, the *minimal trellis size* s is defined as the maximal state-space dimension of the *minimal trellis diagram* [2]. The minimal s over any permutation of a code A is labeled by $s(A)$. The general Wolf bound is $s(A)\leq\min\{k,n-k\}$ [1]. Let A be a TSC code $\|C/D\|^2$. A simple recurrence formula for the trellis complexity is given by

$$s(A)\leq s(D)+\dim(C)-\dim(D).$$

Improved bounds are derived for iterated SR codes such as primitive BCH codes. Upper bounds on the decoding complexity are thereby implied in conjunction with results of [3]. Some examples of binary primitive BCH codes are given in Table I. Actual s parameters were numerically obtained using computer program (see also [5],[6]).

TABLE I
UPPER BOUNDS ON $s(A)$ FOR PRIMITIVE BCH CODES

Code	C	D	Wolf Bound	New Bound	s
(16,7,6)	(8,6)	(8,1)	7	6	6
(32,21,6)	(16,15)	(16,6)	11	10	10
(64,45,8)	(32,29)	(32,16)	19	15	14
(64,36,12)	(32,26)	(32,10)	28	21	19

Additional results are developed utilizing the highly structural constructions. Furthermore, the trellis complexity of long BCH codes may be evaluated. The constructions may also be useful for related applications such as the generalized weight hierarchy [8].

References

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

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

[3] Y. Berger and Y. Be'ery, "Soft trellis-based decoder for linear block codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 764-773, 1994.

[4] Y. Berger and Y. Be'ery, "Bounds on the trellis size of linear block codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 203-209, 1993.

[5] T. Kasami, T. Takata, T. Fujiwara and S. Lin, "On the optimum bit orders with respect to the state complexity of trellis diagrams for binary linear codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 242-245, 1993.

[6] A. Vardy and Y. Be'ery, "Maximum-likelihood soft-decision decoding of BCH codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 546-554, 1994.

[7] Y. Berger and Y. Be'ery, "Trellis-oriented decomposition and trellis complexity of composite-length cyclic codes," *IEEE Trans. Inform. Theory*, to appear, July, 1995.

[8] V. K. Wei, "Generalized Hamming weights for linear block codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1412-1418, 1991.

Codes Which Satisfy the Two-Way Chain Condition and Their State Complexities

Sylvia B. Encheva
Høgskolen Stord/Haugesund
Skåregt. 103
5500 Haugesund, Norway

Abstract - **All binary linear codes, satisfing the two-way chain condition with dimension up to 6 are described in terms of their generator matrices. An expression for their state complexity profile is found also. Cases, when such codes are $Z_4 - linear$ are shown.**

I. Introduction

Let C be a binary linear $[n,k,d]$ code. The support of a vector $\mathbf{a} = (a_1, a_2, ..., a_n)$ in $GF(2)^n$ is defined by $\chi(\mathbf{a}) = \{j | a_j \neq 0\}$. The minimum support weight, d_r, of a code C is the size of the smallest support of any r-dimensional subcode of C. In particular $d_1 = d$.
The concept of *the two-way chain condition* was introduced by Forney [3].

Definition 1 *An $[n,k]$ code C satisfies the two-way chain condition if it is equivalent to a code $\tilde{C}$ with the following property: there exist two chains of subcodes of $\tilde{C}$, the* left chain $D_1^{\mathrm{L}} \subset D_2^{\mathrm{L}} \subset \cdots \subset D_k^{\mathrm{L}} = \tilde{C}$, *and the* right chain $D_1^{\mathrm{R}} \subset D_2^{\mathrm{R}} \subset \cdots \subset D_k^{\mathrm{R}} = \tilde{C}$, *where, for $1 \leq r \leq k$, we have $\dim(D_r^{\mathrm{L}}) = \dim(D_r^{\mathrm{R}}) = r$, $\chi(D_r^{\mathrm{L}}) = \{1, 2, \ldots, d_r\}$, and $\chi(D_r^{\mathrm{R}}) = \{n-d_r+1, n-d_r+2, \ldots, n\}$.*

The *state complexity profile* of a linear block code C is $\mathbf{s(C)}$, where $s_i(C) = k - p_i - f_i$ and p_i, f_i are the dimensions of the past and future subcodes [2, 3, 5].
The concept of a binary Z_4 - linear code was introduced by A. R. Hammons, P. V. Kumar, A.R. Calderbank, N. J.A. Sloane, P. Sole [4]. A binary code is Z_4 - linear if its coordinates can be arranged so that it is the image under the Gray map ϕ of a linear block code over Z_4, i.e. an additive subgroup of Z_4^n.

II. Main Results

Lemma 1 *A sufficient condition for an $[n,5]$ code C with $d_1 + d_2 \leq n$ to satisfy the two-way chain condition is that $\tilde{C}$ is generated by a matrix*

$$\begin{pmatrix} \overbrace{1}^{a_1} & \overbrace{1}^{a_2} & \overbrace{0}^{a_3} & \overbrace{0}^{a_4} & \overbrace{0}^{a_5} & \overbrace{0}^{a_4} & \overbrace{0}^{a_3} & \overbrace{0}^{a_2} & \overbrace{0}^{a_1} \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

where

$$a_j \leq a_1, 2 \leq j \leq 5,$$
$$a_j + a_{j+1} \leq a_1 + a_2, 3 \leq j \leq 4,$$
$$a_1 \leq a_j + a_{j+1}, 2 \leq j \leq 4.$$

The state complexity profile for codes from Lemma 1 is

$01^{a_1-1}01^{a_3-1}01^{a_5-1}01^{a_3-1}01^{a_1-1}0$ for $a_2 = a_4 = 0$,
$01^{a_1}2^{a_2-1}1^{a_3}01^{a_5-1}01^{a_3}2^{a_2-1}1^{a_1}0$ for $a_2 \geq 1, a_4 = 0$,
$01^{a_1-1}01^{a_3}2^{a_4-1}1^{a_5+1}2^{a_4-1}1^{a_3}01^{a_1-1}0$ for $a_2 = 0, a_4 \geq 1$,
$01^{a_1}2^{a_2-1}1^{a_3+1}2^{a_4-1}1^{a_5+1}2^{a_4-1}1^{a_3+1}2^{a_2-1}1^{a_1}0$ otherwise.

The state complexity is
1 if $a_4 \leq 1$ and $a_2 \leq 1$,
2 otherwise.

Lemma 2 *Codes described in Lemma 1 are Z_4 - linear if a_5 is an even number.*

Acknowledgements

The author is grateful to Torleiv Kløve for helpful discussions and valuable comments on the manuscript. Thanks are also due to Alexander Vardy.

References

[1] Encheva S.B., Jensen H., Høholdt T., "Subcodes of Z_4- linear codes", *Pros. IEEE International Symposium on Information Theory*, Trondheim, 1994.

[2] G. David Forney, Jr., "Dimension/Length Profiles and Trellis Complexity of Linear Block codes", *IEEE Trans. Inform. Theory*, vol. 40, pp. 1741-1752, November 1994.

[3] G. David Forney, Jr., "Density/Length Profiles and Trellis Complexity of Lattices", *IEEE Trans. Inform. Theory*, vol. 40, pp. 1753-1772, November 1994.

[4] A. R. Hammons, P. V. Kumar, A.R. Calderbank, N. J.A. Sloane, P. Sole, "The Z_4 - Linearity of Kerdock, Preparata, Goethals and Related Codes", *IEEE Trans. on Inform. Theory*, vol. IT-40, No 2, March, 1994.

[5] A.Vardy and Y.Be'ery, "Maximum likelihood soft decision decoding of BCH codes", *IEEE Trans. Inform. Theory*, vol. 40, pp. 546-554, March 1994.

The Trellis Complexity of Convolutional Codes[1]

Robert J. McEliece and Wei Lin
California Institute of Technology

Abstract — **We develop a theory of minimal trellises for convolutional codes, and find that the "standard" trellis need not be the minimal trellis.**

I. Introduction

From a minimal generator matrix $G(D)$ for an (n, k, m) convolutional code, it is possible to construct a "standard" trellis representation for $\mathcal{C}$. This trellis is in principle infinite, but it has a very regular structure, consisting (after a short initial transient) of repeated copies of what we shall call the *trellis module* associated with $G(D)$. The trellis module consists of 2^m "initial states" and 2^m "final states," with each initial state being connected by a directed edge to exactly 2^k final states. Each directed edge is labelled with an n-bit binary vector, namely, the output produced by the encoder in response to the given state transition.

Since the trellis module has 2^{k+m} edges, and each edge has "length" (measured in bits) n, then total edge length of the trellis module is $n \cdot 2^{k+m}$. Since each trellis module represents the encoder's response to k input bits, we are led to define the "standard trellis complexity" of the code as

$$\frac{n}{k} \cdot 2^{m+k} \quad \text{edges per bit.} \tag{1}$$

The standard trellis complexity as defined in (1) is a measure of the effort *per decoded bit* required by Viterbi's algorithm. However, we will see in the next section that this complexity can sometimes be reduced, by the construction of a simplified trellis for the code.

II. Example

Consider the $(8, 4, 3)$, $d_{\text{free}} = 8$, "partial unit memory" convolutional code with minimal generator matrix

$$G(D) = \begin{pmatrix} 11111111 \\ 11101000 \\ 10110100 \\ 10011010 \end{pmatrix} + \begin{pmatrix} 00000000 \\ 11011000 \\ 10101100 \\ 10010110 \end{pmatrix} D \tag{2}$$

(see [3]). According to (1), the "standard" trellis complexity of this code is 256 edges per bit. However, it is quite easy to reduce this number, as follows.

We view the code in (2) as an (infinite-length) block code, with "scalar" generator matrix

$$G_{\text{scalar}} = \begin{bmatrix} G_0 & G_1 & & & \\ & G_0 & G_1 & & \\ & & G_0 & G_1 & \\ & & & G_0 & G_1 \\ & & & & \ddots \end{bmatrix} \tag{3}$$

where $G(D) = G_0 + D \cdot G_1(D)$. From this representation, and using a modification of the now "standard" theory of trellises for block codes [4], one can see that the code has a minimal trellis, built from trellis modules, each of which has 480 edges. Since each module represents four encoded bits, the trellis complexity, as measured in trellis edges per encoded bit, is thereby reduced to 120.

In this example, the trellis complexity can be reduced still further, if we allow column permutations of the original generator matrix $G(D)$ in 2. Indeed, by computer search we have found that one "minimal complexity" column permutation for this particular code is the permutation (01243567), which results in the generator matrix (cf. (2))

$$G(D) = \begin{pmatrix} 11111111 \\ 11110000 \\ 10101100 \\ 10011010 \end{pmatrix} + \begin{pmatrix} 00000000 \\ 11011000 \\ 10110100 \\ 10001110 \end{pmatrix} D. \tag{4}$$

Then after putting the minimal generator matrix of (4) into "minimal span" form, it becomes

$$G(D) = \begin{pmatrix} 11111111 \\ 00001111 \\ 01111111 \\ 00111111 \end{pmatrix} + \begin{pmatrix} 00000000 \\ 11111000 \\ 11111100 \\ 11111110 \end{pmatrix} D. \tag{5}$$

The trellis complexity of the generator matrix in (5) turns out to be 104 edges per encoded bit.

III. General Results

We have found a simple algorithm algorithm for finding a generator matrix $G(D)$ for a convolutional code, for which the corresponding "scalar" generator matrix (cf. (3)) is in "minimal span" form [4]. This generator matrix can then be used to produce the minimal trellis for the convolutional code. In principle, the theory of minimal trellises for convolutional codes can be deduced from the general "Forney-Trott" theory [2], but we believe the observation that the Viterbi decoding complexity of convolutional codes can be thereby systematically reduced is new, as are the details of the algorithms for producing the minimal trellises.

One nice by-product of our theory is that when we apply our techniques to a convolutional code obtained by puncturing [1], we always find a trellis for that code which is as least as simple as the "punctured" trellis. Thus in the new theory, punctured convolutional codes no longer appear as a special class, but simply as high-rate convolutional codes whose trellis complexity turns out to be unexpectedly small.

[1] This work was partially supported by a grant from Pacific Bell.

References

[1] J. B. Cain, G. C. Clark, and J. M. Geist, "P:unctured Convolutional Codes of rate $(n-1)/n$ and simplified maximum likelihood decoding," *IEEE Trans. Infrom. Theory*, vol. IT-25 (January 1979), pp. 97–100.

[2] G. D. Forney, Jr., and M. D. Trott, "The dynamics of group codes: state spaces, trellis diagrams, and canonical encoders." *IEEE Trans. Inform. Theory, vol. IT-39 (1993), pp. 1491–1513.*

[3] G. S. Lauer, "Some optimal partial-unit-memory codes," *IEEE Trans. Inform. Theory*, vol. IT-25 (March 1979), pp. 540–547.

[4] R. J. McEliece, "On the BCJR Trellis," submitted to *IEEE Trans. Inform. Theory.*

If Binary Codes Existed That Exceed Gilbert–Varshamov Bound They Could Not Reach the Cutoff Rate of BSC

Thomas Beth, Dejan E. Lazic

Universität Karlsruhe, Fakultät für Informatik, IAKS, 76 128 Karlsruhe, Germany, e-mail: lazic@ira.uka.de

Abstract — **Binary block codes exceeding the Gilbert–Varshamov bound on minimum Hamming distance (if such codes exist) have their error exponents below the one of the Binary Symmetric Channel (BSC) in the interval (R_{crit}, R_C) and they cannot reach the cutoff rate R_o and thus the capacity R_C of the BSC if a Maximum Likelihood decoder (MLD) is used.**

I. Introduction

In [1] a nonstandard technique for bounding the error exponent of specific families of channel block codes was introduced. Contrary to the standard methods based on ensemble averaging, this technique, called the distance distribution method, enables the unification of three different approaches to the asymptotical analysis of channel codes: channel coding theorems, bounds on the error exponent, and bounds on the minimum distance.

II. The Connection between $\underline{d}_{HGV}(R)$ and R_o

The general lower bound on the code family error exponent was obtained in the following form

$$\underset{\sim}{E}(R)_{\mathscr{B}} = \min_{\substack{\delta_1(\mathscr{B},R)\le d,\\ d\le\delta_L(\mathscr{B},R)}} \left(E_{\mathscr{B}}(d,R) + \underset{\sim}{E}_e(d,R,\mathscr{B}) - R\right) , \qquad (1)$$

where $\mathscr{B}$ is an infinite family of channel block codes $B(R,N)$ over a finite or infinite alphabet, provided by some channel-determined distance measure d between its codewords. $\mathscr{B}$ is characterized by the distance distribution exponents (DDE)

$$E_{\mathscr{B}}(d,R) = \left\{\lim_{N\to\infty} -\tfrac{1}{N}\,\mathrm{ld}\left(\tfrac{\pi_l}{M(M-1)}\right)\right\}_{l=1}^{L} , \qquad (2)$$

where π_l represents the number of ordered pairs of codewords from $B(R,N)$ that are on some fixed distance $d_l > 0$, L the total number of different values of $d > 0$ in $B(R,N)$ (arranged in increasing order), and $M = 2^{RN}$ the number of codewords in $B(R,N)$. The influence of the decoding algorithm and the channel performance is characterized by the error effect exponent (EEE)

$$E_e(d,R) = \left\{\lim_{N\to\infty} -\tfrac{1}{N}\,\mathrm{ld}(e_l)\right\}_{l=1}^{L} , \qquad (3)$$

where $e_l = P[\hat{\boldsymbol{x}} = \boldsymbol{x}_j \mid \boldsymbol{x}_m, m \neq j, d_l = d(\boldsymbol{x}_j, \boldsymbol{x}_m)]$ represents the error effect of the codeword $\boldsymbol{x}_j$ when the codeword $\boldsymbol{x}_m$ is erroneously decoded provided $\boldsymbol{x}_m$ and $\boldsymbol{x}_j$ are on some fixed distance d_l. $\underset{\sim}{E}_e(d,R,\mathscr{B})$ in (1) denotes some lower bound on (3). For each fixed value $R > 0$ of the code rate, chosen from the set of possible family rates $\mathcal{R}$, the code family $\mathscr{B}$ contains an infinite fixed rate sequence of block codes $FRS(R,\mathscr{B}) = (B(R,N_1), B(R,N_2), \ldots)$ where $N_i < N_{i+1}$ and $R = ld\, M_i/N_i = \text{const}$, $i = 1,2,\ldots$ with $M_i = |B(R,N_i)|$. $\delta_1(\mathscr{B},R)$ and $\delta_L(\mathscr{B},R)$ in (1) are asymptotical values of min. and max. distances of codes in $FRS(R,\mathscr{B})$ for each $R \in \mathcal{R}$.

For the family of uniformly distributed binary codes, $\mathscr{B}_{ub}$, whose Hamming distances are binomially distributed

$$\tfrac{\pi_l}{M(M-1)} = 2^{-N}\binom{N}{l}; \qquad l = d_H = 1,2,\ldots,N \qquad (4)$$

for all rates $R \in \mathcal{R} = [0,1]$, the DDE function (2) in the interval $[0, 0.5]$ represents the Gilbert–Varshamov curve $\underline{d}_{HGV}(R)$ when d represents the normalized Hamming distance $\underline{d}_H = d_H/N$, i. e., $E_{\mathscr{B}_{ub}}(\underline{d}_H, R) = 1 - H(\underline{d}_H)$, where $H(x)$ is the binary entropy function [2]. On the other hand, when d represents the normalized Bhattacharyya distance on the BSC (with transition probability p) given by $\underline{d}_B = -\mathrm{ld}\sqrt{4p(1-p)}\underline{d}_H$ the DDE function (2), $E_{\mathscr{B}_{ub}}(\underline{d}_B, R)$, of the family $\mathscr{B}_{ub}$ determines the cutoff rate R_o of the BSC. Under the usual condition of equal prior probabilities of codewords and using the MLD, this fact was shown in [1] by replacing $E_{\mathscr{B}_{ub}}(\underline{d}_B, R)$ in (1) and using a very simple lower bound on the EEE function (3) given by $\underset{\sim}{E}_e(\underline{d}_B, R, \mathscr{B}) = \underset{\sim}{E}_e(\underline{d}_B) = \underline{d}_B$.

III. Sketch of the Proof

Proving that the Hamming distance distribution exponent $E_{FRS^\bullet}(\underline{d}_H, R')$ of a hypothetical fixed rate sequence $FRS^*(R')$ of binary block codes with asymptotical normalized minimum Hamming distance $\underline{\delta}^*_{H1}(R')$ that exceed the Gilbert–Varshamov bound must intersect the Gilbert–Varshamov curve is the first step in the proof of the main statement of this paper. This can be proven by chosing the special value $p = p'_{crit}$ for which $R' = R_{crit}$. Then $E_{FRS^\bullet}(\underline{d}_H, R') > E_{\mathscr{B}_{ub}}(\underline{d}_H, R')$ for $0 \le \underline{d}_H \le 0.5$ contradicts the space-partitioning upper bound on $E(R)_{BSC}$. Furthermore, using the distance distribution method it can be shown that the cutoff rate lower bound $\underset{\sim}{E}_o(R')_{FRS^\bullet}$ on the error exponent of $FRS^*(R')$ on the BSC must be smaller than the cutoff rate lower bound $\underset{\sim}{E}_o(R)$ on the error exponent of the BSC for $R = R'$ and for $p > p'_{crit}$, i. e., when $R_{crit} \le R' < R_C$.

IV. Conclusion

It is shown that the still open famous problem of finding binary codes with minimum distances that exceed the Gilbert–Varshamov bound is irrelevant when MLD is used. In this case the Gilbert–Varshamov bound (curve) uniquely determines the error exponent of asymptotically optimal binary block codes on the BSC in the interval (R_{crit}, R_C). The same conclusion is valid for spherical codes on the AWGN channel that exceed the Shannon lower bound on minimal Euclidean distance.

References

[1] D. E. Lazic and V. Senk: 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 IT–38, no. 5, pp. 1548–1559, Sept. 1992.

[2] Th. Beth, D. E. Lazic, and V. Senk: The Generalized Gilbert–Varshamov Distance of a Code Family and its Influence on the Family's Error Exponent, Proceedings of the ISITA, Sydney, Australia, Vol. 1, pp. 965–970, 1994.

A Simple Proof that Time-Invariant Convolutional Codes Attain Capacity

Nadav Shulman and Meir Feder

Department of Electrical Engineering - Systems, Tel-Aviv University, Tel-Aviv, 68878, ISRAEL

Abstract — **It is well known that time-varying convolutional codes can achieve the capacity of a discrete memoryless channel [1]. The time varying assumption is needed in the proof to assure *pairwise* independency between the codewords. In this work we provide a relatively simple proof that indeed time-*invariant* convolutional codes can achieve the capacity without any restriction (albeit, the error exponent achieved by our proof may not be the optimal).**

I. Overview

We consider the following setting of fixed (time-invariant) convolutional codes with rate $R = b/n$ bits per symbol: At each time instance an information vector $\mathbf{u}_t = \{u_t^1, u_t^2, \ldots, u_t^b\}$ of b bits is pushed into a delay line (register) of length K (i.e. the delay line contains $b \cdot K$ bits). Then $n \cdot q$ bits, $a_{i,j}$, $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, q\}$, which are linear combinations of bits in the register are calculated. These combinations define the specific convolutional code. n output symbols, $\{o_i\}_{i=1}^n$, are produced using a mapping from bits to channel symbols, $\mathcal{M} : \{0,1\}^q \to \{1, \ldots, J\}$, $o_i = \mathcal{M}(a_{i,1}, \ldots, a_{i,q})$. The mapping defines a distribution $Q(k) = 2^{-q}|\{\mathbf{a}; \mathcal{M}(\mathbf{a}) = k\}|$. Note that as $q \to \infty$ any distribution Q can be approximated.

We show that for a given DMC with a transition probability $P(y|x)$, a distribution $Q(x)$ and *any* b and n such that $b/n < I(Q;P)$,where $I(\cdot;\cdot)$ is the mutual information, there exists a sequence of convolutional codes of increasing K such that for $K \to \infty, P_{\text{error}} \to 0$ exponentially where P_{error} is the probability of an error in decoding $N \cdot b$ transmitted bits.

II. Outline of the proof

We analyze the average performance of an ensemble of convolutional codes, defined by a randomly chosen $q \cdot n$ linear combinations (requiring $q \cdot n \cdot b \cdot K$ random bits), and by a random initial value of the register.

Our proof analyzes a sub-optimal decoding procedure in which at each time point t we decode the information symbol $\mathbf{u}_t$ based on a future observed block of size $L_t \cdot n$ symbols. The value L_t will be chosen, as described below, so that Gallager's technique to upper bound the error probability in block coding (see [3, pp 135–150]) can be applied, i.e., that there will be a pairwise independence between the true codeword and any codeword that can cause an error in decoding $\mathbf{u}_t$. If an error will occur at *any* time point, we shall declare that our decoding has failed. We shall show that, on the average, the error probability in decoding $\mathbf{u}_t$ will vanish, exponentially in K. Thus, as long as the information sequence length N is short enough, P_{error} will also vanish exponentially.

Specifically, we first constrain $L_t < K/2$. Then, we use the fact that if A is a binary matrix with rank l and $\mathbf{v}$ is a random binary vector with uniform iid components , then the random vector $A\mathbf{v}$ has l uniform iid components. A lower bound on the number of symbols we can use and still have pairwise independence, $L_t \cdot n$, can be calculated from the current register value. We assume to know the current register value, since otherwise an error has already occurred, and there is no need to calculate the error probability in decoding $\mathbf{u}_t$. It can be shown that taking $L_t = l$, where l is the maximum number such that the rows of the matrix

$$\begin{pmatrix} \mathbf{u}_{t-\frac{K}{2}} & \mathbf{u}_{t-\frac{K}{2}-1} & \cdots & \mathbf{u}_{t-K+1} \\ \mathbf{u}_{t-\frac{K}{2}+1} & \mathbf{u}_{t-\frac{K}{2}} & \cdots & \mathbf{u}_{t-K+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_{t-\frac{K}{2}+l-1} & \mathbf{u}_{t-\frac{K}{2}+l-2} & \cdots & \mathbf{u}_{t-K+l} \end{pmatrix}$$

are still linear independent, will ensure the desired pairwise independence. (This matrix is known to the decoder because it contains only bits that have already been decoded). Now, we analyze the error probability, averaged over a uniform choice over the messages, i.e., under the assumption that $\mathbf{u}$ are uniformly distributed. In this case we have $\Pr\{L_t = l\} \le 2^{b(l-\frac{K}{2})}$. For each value of L_t we face the situation where we observe a block of $L_t \cdot n$ symbols and we try to decide between at most $2^{b \cdot L_t}$ randomly chosen different possible inputs. The error probability in this case can be upper bounded by Gallager's exponential expression for block codes. Using this expression, and taking the expectation with respect to L_t, we get:

$$\begin{aligned} P_e &\le \sum_{l=0}^{K/2} 2^{b(l-\frac{K}{2})} \cdot 2^{blR} \cdot 2^{-nlE_0(\rho,Q)} \\ &\le (1 + K/2) \cdot 2^{-n\frac{K}{2}\min(R, E_0(\rho,Q)-\rho R)} \\ P_{\text{error}} &\le N P_e \end{aligned}$$

For $R < I(Q;P)$ and $\log N = o(K \cdot n)$, the expression above goes to zero exponentially with $K \cdot n$.

III. Discussion and further improvements

The achieved error exponent above is worse than the error exponent for *time varying* convolutional codes [1], and even from the error exponent for block codes [3]. A better exponent was achieved for special cases such as BSC by a slight change in the proof. In [2], it was claimed (without a proof) that for $b \to \infty$, time-invariant convolutional codes can achieve the same exponent as *time varying* codes. This claim was also proved by us with similar technique (but without constraining L_t to be less than $K/2$). The question whether fixed convolutional codes has the same error-exponent as time-varying for any b, is still under investigation.

References

[1] Andrew J. Viterbi and Jim K. Omura. *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[2] K. Sh. Zigangirov. Time-invariant convolutional codes: Reliabilty function. In *Proc. 2nd Joint Soviet-Swedish Workshop Information Theory*, Gränna, Sweden, April 1985.

[3] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

A Construction of Codes with Exponential Error Bounds on Arbitrary Discrete Memoryless Channels

Tomohiko Uyematsu and Eiji Okamoto
School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, 923-12, Japan

Abstract — **This paper proposes an explicit construction of codes achieving Shannon's capacity for arbitrary discrete memoryless channels. The proposed codes are obtained by applying the idea of variable concatenation to a class of concatenated codes with employing algebraic geometry codes as outer codes. Further, we clarify that the error exponent of the proposed code is equal to the error exponent obtained by Forney for concatenated codes.**

I. Introduction

In 1982, P. Delsarte and P. Piret gave an explicit construction of codes achieving the capacity and admitting a simple decoding algorithm[1]. Recently, M. Steiner expanded their results and gave an explicit construction of codes having the decoding error probability bounded by an exponential function of block length at all rate below capacity for any discrete memoryless channel[2]. However, the error exponent of the code is far below Forney's error exponent which gives the performance obtainable with concatenated codes[3].

This paper proposes a new explicit algebraic construction of codes achieving Shannon's capacity for any discrete memoryless channel. The proposed code can be regarded as a generalization of Justesen codes[4] followed by a channel-dependent mapping, with employing an algebraic geometry code[5] as the outer code. The proposed codes are optimum in the sense that they can attain Forney's error exponent.

II. The Ensemble of Inner Encoders

Let us consider a discrete memoryless channel (DMC) with input alphabet X and output alphabet Y. We assume that a set of messages delivered by the information source consists of all k-tuples $a = (a_1, a_2, \cdots, a_k)$ with $a_i \in GF(q)$, for a certain positive integer k.

Let us identify $GF(q)^k$ with any k-dimensional subspace of $GF(q^n)$. To each pair (γ, σ) of elements γ and σ of $GF(q^n)$, we associate the affine encoder $g : GF(q)^k \to GF(q^n)$ given by

$$g(a) = \gamma a + \sigma, \qquad a \in GF(q)^k, \tag{1}$$

and define G_A to be the set of all such encoders. Further, let G be a set of encoders expurgating the encoder with $\gamma = \sigma = 0$ from G_A. Obviously, $G \subset G_A$, $|G_A| = q^{2n}$ and $|G| = q^{2n} - 1$, As is usual, the number $r = k/n$ is referred to as the rate.

III. Code Construction

The outer code is formed by an algebraic geometry code $C_H(N, K)$ constructed from a generalized Hermitian curve[5], which is a linear code over $GF(q^{2m})$ with $q = 2^L$ and the code length $N = q^{2n} - 1$. The inner codes are variable (n, k) codes over $GF(q)$ which belong to G, where $k = 2m$. The overall concatenated code is an (N_o, K_o) code over $GF(q)$, where $N_o = nN$ and $K_o = 2mK$. Let us denote the proposed (channel-independent) code by C.

In order to apply the proposed code C to a channel, the symbols of inner codewords are mapped into channel input symbols by a channel-dependent mapping $\eta : GF(q) \to X$. The mapping η is constructed such that the occurrence probability of $x \in X$ approximates the desired input probability $Q_{max}(x)$ which achieves capacity of the channel[6]. More precisely, for all $x \in X$, let i_x be integers, such that $i_x/q \approx Q_{max}(x)$ and $\sum_{x \in X} i_x = q$. Then, i_x distinct symbols of $GF(q)$ are mapped into the identical channel input symbol x. Hence, for any $\theta > 0$ and appropriately chosen i_x ($x \in X$), we can find a sufficiently large q (or L) such that

$$\left| \frac{Q_{max}(x)}{i_x/q} - 1 \right| \le \theta \qquad \forall x \in X. \tag{2}$$

Let us denote this channel-dependent code by $\tilde{C}$.

IV. Exponential Error Bounds

The next theorem is our main result.

Theorem 1: Let the inner codes be decoded by maximum likelihood decoding and the outer code by GMD decoding. Then, on arbitrary DMC, for arbitrarily fixed $\epsilon > 0$ and sufficiently large m and L, the proposed code $\tilde{C}$ of overall length N_o and overall rate $R_o (= K_o/N_o)$ has the average probability of decoding error $\bar{P}_e$ bounded by

$$\bar{P}_e \le q^{-N_o E_P(R_o, \epsilon, \theta)}, \tag{3}$$

where

$$E_P(R_o, \epsilon, \theta) = \max_{0 < R \le R_o} (1 - R) \left(E\left(\frac{R_o}{R}\right) - \epsilon - \alpha(\theta) \right), \tag{4}$$

$E(r)$ is the Gallager's reliability function[6], and $\alpha(\theta) \to 0$ as $\theta \to 0$. □

By choosing ϵ and θ properly, we can obtain $E(R_o, \epsilon, \theta) > 0$ whenever $R_o < C_o$, where C_o denotes the capacity of the original channel. This implies that the proposed codes achieve Shannon's capacity. Further, the error exponent $E_P(R_o, 0, 0)$ is equal to Forney's error exponent[3].

References

[1] P. Delsarte and P. Piret :"Algebraic Constructions of Shannon Codes for Regular Channel," *IEEE Trans. on Inform. Theory*, vol.IT-28, no.4, pp.593-599, 1982.

[2] M. Steiner : "Constructive Codes for Arbitrary Discrete Memoryless Channels", *IEEE Trans. on Inform. Theory*, vol.IT-40, no.3, pp.929-934, 1994.

[3] G. D. Forney, Jr. : "Concatenated Codes", MIT Press, 1966.

[4] J. Justesen : "A Class of Constructive Asymptotically Good Algebraic Codes", *IEEE Trans. on Inform. Theory*, vol.IT-18, no.5, pp.652-656, 1972.

[5] B.-Z. Shen : "A Justesen Construction of Binary Concatenated Codes that Asymptotically Meet the Zyablov Bound for Low Rate", *IEEE Trans. on Inform. Theory*, vol.IT-39, no.1, pp.239-242, 1993.

[6] R. G. Gallager : "Information Theory and Reliable Communication", Wiley, 1968.

RESTRICTED TWO-WAY CHANNEL: BOUNDS FOR ACHIEVABLE RATES REGION FOR GIVEN ERROR PROBABILITY EXPONENTS

Evgueni A. Haroutunian, Mariam E. Haroutunian and Arthur E. Avetissian

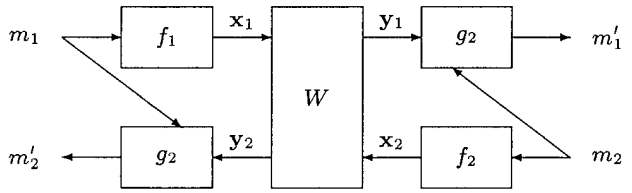
Institute for Informatics and Automations Problems of the
Armenian National Academy of Sciences Yerevan State University
P.Sevak str.1, 375044 Yerevan, Armenia
e-mail: harout @ ipia.armenia.su

Abstract — **Shannon's restricted two-way channel is studied. The outer and the inner bounds are obtained of the region of rates achievable when error probabilities exponentially decrease with given at two outputs exponents.**

A restricted two-way channel (RTWC) is defined by a matrix of transition probabilities

$$W = \{W(y_1, y_2|x_1, x_2), x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2\},$$

where $\mathcal{X}_1$, $\mathcal{X}_2$ are the finite input alphabets and $\mathcal{Y}_1$, $\mathcal{Y}_2$ are the finite output alphabets of the channel. The channel is supposed to be memoryless. RTWC is represented in figure.



There are two terminals. When the symbol $x_1 \in \mathcal{X}_1$ is sent from the first terminal, the corresponding output symbol $y_1 \in \mathcal{Y}_1$ arrives on the terminal 2. At the same time the input symbol x_2 is transmitted from the terminal 2 and the corresponding symbol y_2 arrives on the opposite terminal. Let

$$\mathcal{M}_1 = \{1, 2, \ldots, |\mathcal{M}_1|\} \quad \text{and} \quad \mathcal{M}_2 = \{1, 2, \ldots, |\mathcal{M}_2|\}$$

be the message sets of corresponding sources. The code for RTWC is a collection of mappings (f_1, f_2, g_1, g_2), where $f_1 : \mathcal{M}_1 \to \mathcal{X}_1^n$, $f_2 : \mathcal{M}_2 \to \mathcal{X}_2^n$ are the encodings and $g_1 : \mathcal{M}_2 \times \mathcal{Y}_1^n \to \mathcal{M}_1$, $g_2 : \mathcal{M}_1 \times \mathcal{Y}_2^n \to \mathcal{M}_2$ are the decodings.

The restrictions mentioned in the name of the model, means that in the RTWC there are no connections from decoders to encoders on the same terminal. The average error probabilities of the code $e_i(f_1, f_2, g_1, g_2)$ is considered.

Let $\epsilon = (\epsilon_1, \epsilon_2)$, $0 \le \epsilon_i < 1$, $i = 1, 2$. Nonnegative numbers R_1, R_2 are called ϵ-achievable rates pair for RTWC, if for any $\delta_i > 0$, $i = 1, 2$ there exists a code such that for sufficiently large n

$$\frac{1}{n} \log |\mathcal{M}_i| \ge R_i - \delta_i, \quad i = 1, 2,$$

and

$$e_i(f_1, f_2, g_1, g_2) \le \epsilon_i, \quad i = 1, 2.$$

The set of all ϵ-achievable rates pairs is called the capacity region. For $\epsilon_i = \exp(-nE_i)$, $E_i > 0$, $i = 1, 2$, $E = (E_1, E_2)$ the region of achievable rates we call E-capacity region $C(E)$.

The RTWC was first investigated by Shannon [1], who obtained the capacity region of the RTWC. Important results relating to various models of two-way channels were obtained by Ahlswede [2-4], Zhang, Berger and Schalkwijk [5], Han [6]. Papers of Van der Meulen [7], Gelfand and Prelov [8] and the book of Csiszár and Körner [9] contain the detailed surveys.

In the present paper the outer *sphere packing* and the inner *random coding* bounds for $C(E)$ are constructed. For small E this bounds coincide and when $E_i \to 0$, $i = 1, 2$ we obtain the capacity region of RTWC.

The inner bound is obtained using the Shannon's random coding medhod, and upper bound is constructed by the combinatorial method proposed by Haroutunian [10].

ACKNOWLEDGEMENTS

Authors thank Prof. I. Csiszàr for valuable remarks concerning a version of the paper.

REFERENCES

[1] C. E. Shannon, "Two-way communication channels," *Proc. 4-th Berkeley Symp. Math. Stat. and Prob.*, vol. 1, pp. 611–644, 1961.

[2] R. Ahlswede, "On two-way communication channels and a problem by Zarankiewicz," *Trans. 6-th Prague Conference on Inform. Theory, Statistical Decision Functions, Random Processes.*, pp. 23–37, 1971.

[3] R. Ahlswed, "Multy-way communication channels," *Proc. 2-nd Intern. Symp. Inform. Theory.*, Tsahkadsor, Armenia, 1971, *Publishing House of the Hungarian Academy of Sciences.*, pp. 23–52, 1973.

[4] R. Ahlswede, "The capacity region of a channel with two senders and two receivers," *Ann. Prob.*, vol. 2, no. 2, pp. 805–814, 1974.

[5] Z. Zhang, T. Berger, J. P. M. Schalkwijk, "New outer bounds to capacity regions of two-way channels," *IEEE Trans. on Inform. Theory*, vol. IT-32, no. 3. pp. 383–386, 1986.

[6] T. S. Han, "A general coding scheme for the two-way channels," *IEEE Trans. on Inform. Theory*, vol. IT-30, no. 1, pp. 35–44, 1984.

[7] E. C. Van der Meulen, "A Survey of multi-way channels in information theory: 1961-1976," *IEEE Trans. on Inform. Theory*, vol. IT-23. no 1, pp. 1–37, 1977.

[8] S. I. Gelfand, V. V. Prelov, "Communication with many users (in Russian)," *Itogy nauki i texniki. Prob. theory, math. statistics, technical cybernetics*, vol. 15, M. VINITI, pp. 123–162, 1978.

[9] I. Csiszár, J. Körner, "Information theory. Coding Theorems for Discrete Memoryless Systems," *Budapest:Akademiai Kiado*, 1981.

[10] E. A. Haroutunian, "Combinatorial method of construction of the upper bound for E-capacity," (in Russian), *Mezhvuz. sb. nouch. trudov. Matematika*, vol. 1, Yerevan, pp. 213–220, 1982.

Lattice Codes Can Achieve Capacity on the AWGN Channel

R. Urbanke and B. Rimoldi

Washington University, Dept. of Electrical Engineering
Electronic Systems and Signals Research Laboratory
St. Louis, MO 63130, USA

Abstract — **It is shown that lattice codes (intersection of a sphere with a possibly translated lattice) can achieve capacity on the additive white Gaussian noise channel.**

I. Introduction

Consider the additive white Gaussian noise (AWGN) channel with peak signal-power constraint S. It is well known [1] that the capacity of this channel is $C = \frac{1}{2}\log(1+\frac{S}{N})$, where N is the variance of the i.i.d. Gaussian noise. The proof in [1] is non constructive in nature and, hence, codes that achieve capacity may exhibit little or no structure, making them ill suited for practical applications. An important class of structured codes are *lattice codes* which we define to be the intersection of a possibly translated lattice Λ with a spherical bounding region centered at the origin. The following facts are known: (1) For any rate $R < \frac{1}{2}\log(\frac{S}{N})$ there exists a lattice code which results in an arbitrarily small (maximum) probability of error when used with lattice decoding [4, 5, 6]. (2) If we choose the code as the intersection of a possibly translated lattice with a "thin" spherical shell centered at the origin then rates up to capacity can be achieved with arbitrarily low (average) probability of error under a minimum distance decoding [2, 3]. Further, the rate at which the error probability tends to zero is essentially equal to the optimum one as determined by Shannon [1]. Regarding the second result, in [3] it was pointed out that because of the "thin" spherical bounding region these codes resemble more random codes than lattice codes.

We use [2, 3] to close one of the remaining gaps by showing that lattice codes (where the boundary region is a sphere as opposed to a spherical shell) combined with minimum distance decoding can achieve capacity. This is

Theorem 1 Let S, N and $\epsilon > 0$ be given. If $R < \frac{1}{2}\log(1+\frac{S}{N})$ then there exists a lattice code for the additive white Gaussian noise channel with peak power constraint S and noise variance N with rate lower bounded by R and average probability of error of a minimum distance decoder upper bounded by ϵ.

II. Proof Outline

The result is not surprising since in high dimensions most of the volume within a sphere lies in a thin spherical shell and, hence, by adding the volume of the inner sphere to the bounding region we expect that not too many new lattice points are added. The two new key ingredients which make it possible to extend the proof in [2, 3] to Theorem 1 can be stated as follows.

Let S be the available signal power per dimension and N the noise variance. Let R be given such that $R < \frac{1}{2}\log(1+\frac{S}{N})$. Then there exist numbers R' and S' such that $R < R' = \frac{1}{2}\log\left(1+\frac{S'}{N}\right) < \frac{1}{2}\log\left(1+\frac{S}{N}\right)$. Let T_n be the n-dimensional closed sphere of radius $\sqrt{nS}$ and volume V_n, and let T'_n be the n-dimensional open sphere of radius $\sqrt{nS'}$ and volume V'_n. Further, define $T_n^\Delta = T_n - T'_n$ with volume $V_n^\Delta = V_n - V'_n$. Given a lattice Λ_n with fundamental region P_n and $s \in P_n$, define the lattice code $\mathcal{C}_n = \mathcal{C}_n(\Lambda_n, s) = (\Lambda_n + s) \cap T_n$. Similarly, define the subcodes $\mathcal{C}'_n = \mathcal{C}'_n(\Lambda_n, s) = (\Lambda_n + s) \cap T'_n$, and $\mathcal{C}_n^\Delta = \mathcal{C}_n^\Delta(\Lambda_n, s) = (\Lambda_n + s) \cap T_n^\Delta = \mathcal{C}_n(\Lambda_n, s) \setminus \mathcal{C}'_n(\Lambda_n, s)$. Let $M_n = M_n(\Lambda_n, s)$, $M'_n = M'_n(\Lambda_n, s)$ and $M_n^\Delta = M_n^\Delta(\Lambda_n, s)$ be the cardinalities of these codes.

The first lemma states that adding the lattice points within the inner sphere does not increase the error probability by more than the fraction of these points to the total number of codewords. More precisely, if $P_E^{\mathcal{C}}$ denotes the average error probability of a code $\mathcal{C}$ under minimum distance decoding then we have

Lemma 1 $P_E^{\mathcal{C}_n} \le \frac{M'_n}{M_n} + P_E^{\mathcal{C}_n^\Delta}$.

The second lemma shows that the translation vector of the lattice can indeed be chosen in such a way that there are sufficiently many lattice points within the spherical shell but not too many within the inner sphere.

Lemma 2 Let Λ_n be a lattice with fundamental region P_n and determinant $\det(\Lambda_n)$ and define

$$P_n^\star = \left\{ s \in P_n : M_n^\Delta(\Lambda_n, s) \ge \frac{V_n^\Delta}{4\det(\Lambda_n)}; \frac{M'_n(\Lambda_n, s)}{M_n^\Delta(\Lambda_n, s)} \le 4\frac{V'_n}{V_n^\Delta} \right\}.$$

Then $V_n^\Delta \le 2\int_{P_n^\star} M_n^\Delta(\Lambda_n, s)\, dV(s)$.

These two lemmas are then used together with the methods presented in [2, 3] to prove Theorem 1.

This work was supported by National Science Foundation Grant NCR-9357689 and NCR-9304763.

References

[1] C. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611–656, May 1959.

[2] R. De Buda, "Some optimal codes have structure," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 893–899, August 89.

[3] T. Linder, C. Schlegel, and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Transactions on Information Theory*, vol. 39, pp. 1735–1737, September 93.

[4] R. De Buda, "The upper error bound of a new near-optimal code," *IEEE Transactions on Information Theory*, vol. 21, pp. 441–445, July 75.

[5] H. A. Loeliger, "Averaging bounds for lattices and linear codes." submitted to IEEE Transactions on Information Theory, 1994.

[6] H. Loeliger, "On the basic averaging arguments for linear codes," in *Communications and Cryptography* (R. Blahut, D. Costello, U. Maurer, and T. Mittelholzer, eds.), pp. 251–262, Luwer Academic Publishers, 1994.

Achieving Symmetric Capacity of a L-out-of-K Gaussian Channel using Single-user Codes and Successive Decoding Schemes

Roger S Cheng[1]
Electrical & Electronic Engineering Department
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong

Abstract — **We show that single-user codes and a modified successive decoding scheme can be used to achieve symmetric capacity of a L-out-of-K additive white Gaussian channel in all signal-to-noise ratios.**

A L-out-of-K (LOOK) additive white Gaussian noise (AWGN) channel models a multiuser system with K potential users, but with at most L simultaneously active users. The received signal, when the set of active users is $\mathcal{S}$ ($|\mathcal{S}| \leq L$), is given by

$$Y_{(\mathcal{S})} = \sum_{k \in \mathcal{S}} X_k + N,$$

where N is white Gaussian noise. We assume that neither the transmitters nor the receiver know the set of active users.

The capacity region of this channel is known [1] and the symmetric capacity, defined as the maximum value of R where $(R, \ldots, R)$ is in the capacity region, is given by

$$C_{sym}(L, w) = \frac{1}{2L} \log[1 + Lw],$$

when all users have the same symbol signal-to-noise ratio (SNR), w. The result remains the same even if the users are frame-asynchronous.

It is important to note that this is the same as the symmetric capacity of a L-user AWGN channel. Hence, the fact that the transmitters do not know the set of active users does not cause any degradation in the symmetric capacity.

In low SNR, we show that single-user codes can be used to achieve rate very close to the symmetric capacity. In low SNR (i.e. $w/(1+(L-1)w) << 1$ or $C_{sym} << 1$), binary signaling is close to optimal. If each user uses a low rate convolutional code and a binary scrambler before transmission, the codeword probability of error associated with the single-user soft-decision Viterbi decoder is well approximated by assuming all other users' signals as Gaussian noise. This is mainly because the maximum likelihood codeword decision is based on the sum of many received symbols (at least D where D is the free distance of the convolutional code). When the code rate is low, D is large (for sufficiently large constraint length) and the Central Limit Theorem applies as the scramblers at the transmitters ensure i.i.d. transmit symbols. Hence, the capacity of each user, regardless of which set of L users are active, is closely approximated by

$$C_{suc} = \frac{1}{2} \log \left[1 + \frac{w}{1 + (L-1)w} \right].$$

Defining the symmetric capacity ratio

$$\eta_{suc}(L, w) = \frac{C_{suc}(L, w)}{C_{sym}(L, w)},$$

we find that $\lim_{w \to 0} \eta_{suc}(L, w) = 1$ and $\lim_{w \to \infty} \eta_{suc}(L, w) = 0$. Hence, treating other users' signals as noise is near optimal in low SNR since background noise is the dominating factor.

In high SNR, we propose a modified successive decoding scheme that uses only single-user coding and decoding techniques to achieve the symmetric capacity. In the LOOK channel, since none of the transmitter knows who are the active users, the successive decoding (or onion peeling) scheme used in [2, 3, 4] cannot be applied directly. In this modified approach, we split each user into N sub-users and apply the successive decoding scheme on the sub-users of the users, instead of on the K users themselves. The receiver consists of a N-level successive decoding scheme. In the nth level, the receiver decodes the nth sub-users of all users, treating the remaining interference from all sub-users of all users as noise. Then, it subtracts the re-encoded signals of the nth sub-users from the received signal and passes the difference to the $n+1$th level. Since the sub-users have small signal-to-interference-and-noise ratios, binary signaling is near optimal and the aforementioned argument for the Gaussian approximation in calculating the capacity holds. Hence, the symmetric capacity is closely approximated by

$$C_{suc,sd}(N, L, w) = \max \frac{1}{2} \sum_{n=1}^{N} \log \left[1 + \frac{\alpha_n w}{1 + Lw \sum_{m=n+1}^{N} \alpha_m + (L-1) w \alpha_n} \right],$$

where the maximum is taken over all $\alpha_1, \ldots, \alpha_N$ such that $\sum_{n=1}^{N} \alpha_n = 1$.

Defining similarly the symmetric capacity ratio as

$$\eta_{suc,sd}(N, L, w) = \frac{C_{suc,sd}(N, L, w)}{C_{sym}(L, w)},$$

we have that $\lim_{N \to \infty} \eta_{suc,sd}(N, L, w) = 1$ for all L and w.

Finally, the modified successive decoding strategy can also be extended to include multirate users, where each user uses different subsets of the sub-users depending on its desired rate of transmission.

[1] This work was performed at University of Colorado at Boulder and was supported by NSF Grant NCR-92-9812.

References

[1] A. A. Alsugair and R. S. Cheng, "Symmetric capacity and signal design for L-out-of-K symbol-synchronous CDMA Gaussian channels," to appear in *IEEE Trans. on Info. Theory*, July, 1995.

[2] T. Cover, "Some advances in broadcast channels," in *Advances in Communications Systems Theory and Applications*, pp. 229-260, Academic Press, 1975.

[3] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: capcaity region and multiuser water-filling," *IEEE Trans. on Info. Theory*, Vol. 39, pp. 773-785, May, 1993.

[4] B. Rimoldi and R. Urbanke, "Onion peeling and the Gaussian multiple access channel," submitted to *IEEE Trans. on Info. Theory*, Preprint, Nov., 1994.

Capacity of a Memoryless Quantum Communication Channel

Akio Fujiwara and Hiroshi Nagaoka
Department of Mathematics, Osaka Univ., Osaka 560, Japan
Graduate School of Information Systems, Univ. Electro-Communications, Tokyo 182, Japan

Abstract — **We introduce a proper framework of coding problems for a quantum memoryless channel and derive an asymptotic formula for the channel capacity having an operational significance. Some general lower and upper bounds for the quantum channel capacity are also derived.**

I. Introduction

In order to consider a communication system which is described by quantum mechanics, we must reformulate information (communication) theory in terms of quantum mechanical language. However, most of the previous works [1] seem unsatisfactory since they hastily invoke *a priori* analogy between the classical and the quantum communication systems based on the ostensible similarity of various quantum entropies to the classical ones. One of the reason for the immaturity of quantum information theory lies in the lack of asymptotic approaches, although there are a small number of excellent exceptions such as [2]. The purpose of this paper is to present a proper framework of coding problems for a quantum memoryless channel and to derive an asymptotic formula for the *operational* channel capacity [3].

II. Quantum Channel

We here restrict ourselves to finite dimensional Hilbert spaces and to generalized measurements which take values on finite sets for simplicity. Letting $\mathcal{P}(\mathcal{H}_j)$ be the set of states on Hilbert spaces $\mathcal{H}_j$, a *quantum channel* for an input system $\mathcal{H}_1$ and an output system $\mathcal{H}_2$ is described by an affine map $\Gamma : \mathcal{P}(\mathcal{H}_1) \to \mathcal{P}(\mathcal{H}_2)$. In order to investigate asymptotic properties, we consider the nth extension of the system described by tensor product $\overset{n}{\bigotimes} \mathcal{H} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$. This extension corresponds to the situation where the sender transmits n states $\{\sigma_j\}_{j=1}^n$ successively, which is represented by the state $\sigma_1 \otimes \cdots \otimes \sigma_n$ on $\overset{n}{\bigotimes} \mathcal{H}_1$. The extended quantum channel for extended input and output systems $\overset{n}{\bigotimes} \mathcal{H}_1$ and $\overset{n}{\bigotimes} \mathcal{H}_2$ is defined by an affine map $\Gamma^{(n)} : \mathcal{P}(\overset{n}{\bigotimes} \mathcal{H}_1) \to \mathcal{P}(\overset{n}{\bigotimes} \mathcal{H}_2)$. Now, a channel $\Gamma^{(n)}$ is called *memoryless* if

$$\Gamma^{(n)}(\sigma_1 \otimes \cdots \otimes \sigma_n) = (\Gamma\sigma_1) \otimes \cdots \otimes (\Gamma\sigma_n).$$

Since a memoryless channel $\Gamma^{(n)}$ is thus determined uniquely by Γ, we often drop the superscript (n) for simplicity.

III. Quantum Channel Coding Theorem

We first prepare a finite set of quantum states on $\overset{n}{\bigotimes} \mathcal{H}_1$, called the *quantum codebook*, $\mathcal{C}_n = \{\sigma^{(n)}(1), \cdots, \sigma^{(n)}(M_n)\}$, each element of which is an n-tensor product of states on $\mathcal{H}_1$: $\sigma^{(n)}(k) = \sigma_1(k) \otimes \cdots \otimes \sigma_n(k)$. The transmitter first selects a codeword $\sigma^{(n)} = \sigma_1 \otimes \cdots \otimes \sigma_n$ which corresponds to the message to be transmitted (encoding), and then transmits each signal $\sigma_1, \cdots, \sigma_n$ successively through a memoryless channel Γ. The receiver then receives signals $\Gamma\sigma_1, \cdots, \Gamma\sigma_n$ and, by means of a certain measuring process, he estimates which signal among $\mathcal{C}_n$ has been actually transmitted (decoding). In this case, the decoder is described by a $\mathcal{C}_n$-valued measurement $T^{(n)}$ over $\overset{n}{\bigotimes} \mathcal{H}_2$. By fixing a decoder $T^{(n)}$ arbitrarily, the error probability $P_e(\mathcal{C}_n, T^{(n)})$ averaged over the code becomes well-defined in the classical sense. The average error probability for this codebook $P_e(\mathcal{C}_n)$ is defined as the infimum of that over all possible decoder $T^{(n)}$. Further, the quantity $R_n = \log M_n/n$ is called the *rate* for the code $\mathcal{C}_n$. Consider sequences of codes $\{\mathcal{C}_n\}_n$ which satisfy $\lim_{n\to\infty} P_e(\mathcal{C}_n) = 0$, and denote the supremum of $\lim_{n\to\infty} R_n$ over such sequences by $C(\Gamma)$, which is called the *capacity* of the channel Γ.

We establish the relation between the capacity $C(\Gamma)$ and the mutual information. By fixing arbitrarily a measurement $\Pi^{(n)}$ (the totality of which we denote by $\mathfrak{M}^{(n)}$) on a cetain finite set (not necessarily $\mathcal{C}_n$-valued) over $\overset{n}{\bigotimes} \mathcal{H}_2$, we have the (classical) mutual information

$$I^{(n)}(p^{(n)}, \Pi^{(n)}; \Gamma) \overset{\text{def}}{=} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) D_{\Pi^{(n)}}\left(\Gamma\sigma^{(n)} \middle\| \Gamma\rho^{(n)}\right).$$

Here $p^{(n)}(\sigma^{(n)}) = p^{(n)}(\sigma_1, \cdots, \sigma_n)$ is an arbitrary joint distribution over $\mathcal{P}(\mathcal{H}_1)^n = \mathcal{P}(\mathcal{H}_1) \times \cdots \times \mathcal{P}(\mathcal{H}_1)$ (the totality of which we denote by $\mathfrak{P}^{(n)}$), $D_{\Pi^{(n)}}$ is the Kullback-Leibler divergence between the classical probability distributions $\mathrm{Tr}\,[(\Gamma\sigma^{(n)})\Pi^{(n)}(\cdot)]$ and $\mathrm{Tr}\,[(\Gamma\rho^{(n)})\Pi^{(n)}(\cdot)]$, and $\rho^{(n)} \overset{\text{def}}{=} \sum_{\sigma_1,\cdots,\sigma_n} p^{(n)}(\sigma_1, \cdots, \sigma_n)\, \sigma_1 \otimes \cdots \otimes \sigma_n$. It is shown that, for a memoryless channel Γ, the quantity

$$C^{(n)}(\Gamma) \overset{\text{def}}{=} \sup_{p^{(n)} \in \mathfrak{P}^{(n)}} \sup_{\Pi^{(n)} \in \mathfrak{M}^{(n)}} I^{(n)}(p^{(n)}, \Pi^{(n)}; \Gamma)$$

exhibits the superadditivity $C^{(m+n)}(\Gamma) \geq C^{(m)}(\Gamma) + C^{(n)}(\Gamma)$, which is in remarkable contrast to the classical channel. The following theorem gives the quantum counterpart of channel coding theorem:

Theorem 1 *For a memoryless channel* Γ,

$$C(\Gamma) = \lim_{n\to\infty} \frac{C^{(n)}(\Gamma)}{n} = \sup_n \frac{C^{(n)}(\Gamma)}{n}.$$

The quantum channel capacity $C(\Gamma)$ is compared with other capacity-like quantities to obtain general lower and upper bounds: $C^{\otimes}(\Gamma) = C^{(1)}(\Gamma) \leq C(\Gamma) \leq \tilde{C}(\Gamma)$, where $C^{\otimes}(\Gamma)$ is the capacity when restricted to the recursive decoding, $C^{(1)}(\Gamma)$ the capacity for signals of unit length, and $\tilde{C}(\Gamma)$ the pseudo-capacity defined via formal quantum mutual information [3].

References

[1] For an extensive reference list, see C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481–537 (1994).

[2] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99–114 (1991).

[3] A. Fujiwara and H. Nagaoka, Math. Eng. Tech. Rep. **94–22**, University of Tokyo (1994).

The Effect of an Randomly Time-varying Channel on Mutual Information.

Muriel Medard and Robert G. Gallager
Laboratory for Information and Decision Systems, M.I.T.

Abstract : **We examine the effect of a randomly time-varying channel on mutual information between receiver and sender when the channel is m^{th} order Markov.**

We investigate the effect of a randomly time-varying channel upon the mutual information between sender and receiver. Such channels often occur in mobile communications and can affect the achievable rate. If the channel is perfectly known, then the mutual information between a receiver and an arbitrary number of senders may be found, even if the channel is time-varying [1]. In this paper we consider the case of a single receiver and sender pair to set the framework for the more interesting multiple access case.

We consider a discrete-time matrix model for our channel. Let random variable S[i] denote the input at time i, Y[j] the output at time j, N[j] the additive white Gaussian noise at time j and G[j,i] the multiplicative effect of the channel on the output at time j due to the input at time i (the channel's tap at time j corresponding to a delay of j-i). The channel is assumed to be causal and have finite memory limited to Δ time samples, therefore G[j,i] is zero for $j-i>\Delta$ and $j-i<0$. Let us assume that S[i] for any $i<0$ is zero. Let a subscript on a random variable indicate the vector of random variables from times 1 to k, a double subscript on G the corresponding matrix and a single subscript k on G indicate that we are considering the k^{th} row of $G_{k,k}$. $G_{k,k}$ is block-diagonal and G_i is given by [0, ...,G[i, i-Δ], ..., G[i,i-1], G[i,i], ..., 0]. The effect of the channel is given by

$$Y_k = G_{k,k} S_k + N_k \qquad [1].$$

Let us take the channel to be such that any row of $G_{k,k}$ depends on at most m preceding rows, i.e. that the i^{th} row conditioned on rows i-1 through i-m is independent of row i-m-1. In steady-state, it is equivalent to stating that the i^{th} row conditioned on rows i+1 through i+m is independent of row i+m+1. Such a model is that of a m^{th} order Markov chain. Under some conditions of wide-sense stationarity, we may state that

$$\lim_{k\to\infty}\left(\frac{I\left(Y_k; S_k \mid G_{k,k}\right) - I\left(Y_k; S_k\right)}{k}\right) \le \lim_{i\to\infty}\left(I\left(G_i; S_i \mid Y_i, \left\{G_{i+1}\ldots G_{i+m}\right\}\right)\right) \qquad [2]$$

where both the RHS and LHS reach a limit. The LHS represents the loss incurred by not knowing the channel and the RHS is the information that the input gives about the rate of change of the channel.

Suppose, as a special case, that we can describe the channel by a Gauss-Markov model. We assume that, at any time, the taps of the channel are mutually independent and that the expected energy of the tap corresponding to a given delay does not change in time. Let T_c be the coherence time of the channel, roughly the inverse of the Doppler spread. Let T_s be the time spread of the channel, proportional to Δ. The channel may be modelled as becoming decorrelated in time exponentially with rate inversely proportional to T_c. We may write that:

$$G[j,i] = \alpha\, G[j-1,i-1] + \Xi[j,i] \qquad [3]$$

where α is $1/WT_c$. Since the expected energy remains unchanged, the expected energy of $\Xi[j,i]$ is proportional to $(1-\alpha^2)$. We send a white Gaussian signal.

We may show, for the channel model described above,

$$\lim_{T_c\to\infty}\left\{\lim_{k\to\infty}\left\{\frac{I\left(Y_k; S_k \mid G^{1,k}\right) - I\left(Y_k; S_k\right)}{k}\right\}\right\} = 0 \qquad [4].$$

The smaller Δ, i.e. the less dispersive in time the channel, the faster the LHS of [4] goes to 0. The LHS depends both on the coherence time T_c, i.e. on how fast the channel decorrelates, and on the coherence bandwidth, which is inversely related to T_s.

The above arguments can be extended to the multi-dimensional case. They give some idea of the effect of an imperfectly known channel upon interference cancellation. In (2), for spread-spectrum systems, when the input is perfectly decoded, the effect of the channel measurement error on interference cancellation is bounded. These results should also give some indication as to the usefulness of feedback. When T_C is large, the mutual information can be increased by optimizing the input distribution for the user appropriately.

REFERENCES

(1) M.Medard, R.G. Gallager, *'The Issue of Spreading in Multipath Time-varying Channels'*, presented to IEEE Vehicular Technology Conference, 1995

(2) R.G. Gallager, *'An Inequality on the Capacity Region of Multiaccess Multipath Channels'*, MIT, 1994.

Source coding and transmission of signals over time-varying channels with side information

Masoud Khansari and Martin Vetterli
Department of EECS, 231 Cory Hall, University of California, Berkeley, CA, 94720

Abstract — **We look at the problem of transmitting information over time-varying channels with side information, where for time-varying channels the statistics of the channel change with time and by channel side information we mean the current state of the channel. We show that when this side information is available at both the transmitter and the receiver, then for the power-constrained channel, the power allocation policy that achieves minimum end-to-end distortion is not necessarily the same as the one required for maximum transmission rate.**

I. Introduction

A new challenge in telecommunication is the transmission of information over time-varying channels where the statistics of the channel change with time. Examples of such time-varying channels are wireless links where due to multi-path fading and interference from other users, the received signal strength can vary within a few orders of magnitude. Traditionally, the preferred transmission method has been to make the channel behave or look like a channel with uniformly distributed error - e.g. through use of interleaving. Achieving this, then the problem of communication is no harder than it used to be and all the classical methods and tools can be used. It is well-known that this "average channel" method is inherently sub-optimal [1][2]. However, to achieve higher channel capacity, it is required to provide channel *state* side information to either the transmitter or the receiver.

II. Time-varying channels with side information

We consider the state process with sample space $\mathcal{I}$ where at each time instant the channel is at one of these states and hence has different statistics. For example, consider an AWGN channel, where the noise power is modulated in accordance with the channel state. Based on the availability of the current channel state side information, we can distinguish the following four different cases: (**I**): Informed receiver and transmitter, (**II**): Informed receiver, (**III**): Informed transmitter and (**IV**): Average channel. In this paper, we concentrate on case **I**. Note that providing the current channel state does not imply a knowledge about the distribution of the states. In fact, we assume that neither the receiver nor the transmitter is aware of this distribution. It is well-known that the capacity of the channel is given by $C = \sum_{i\in\mathcal{I}} q_i I(X_i, Y_i)$ where q_i is the probability of the channel being at state i and $I(X_i, Y_i)$ is the mutual information between the channel input and output processes at this state. Note that the policy that achieves this capacity is independent of the channel state distribution (q_i). Also since the distribution of the states is unknown, the capacity of the channel is also not known. By policy, here, we mean the distribution of the input channel alphabets that maximizes $I(X_i, Y_i)$.

We can then show that the minimum end-to-end distortion is given by:

$$D_m = \sum_{i\in\mathcal{I}} q_i D(I(X_i, Y_i)). \qquad (1)$$

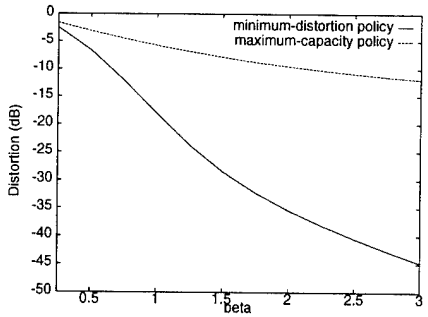


Fig. 1: Performance of minimum distortion and maximum capacity policies vs. β $(D(R) = 2^{-\beta R})$ over a narrow-band Rayleigh fading channel.

Note that had the channel state distribution been also provided to the transmitter then the channel capacity would have been known and $D_m = D\left(\sum_{i\in\mathcal{I}} q_i I(X_i, Y_i)\right)$. In the following section, we look at the power-constrained channels and show that the power allocation policy that achieves minimum end-to-end distortion is not necessarily the same as the one required for maximum transmission rate.

III. Power-constrained Channel

We are considering channels with constraint on the average transmitted power $\bar{S} = \sum_{i\in\mathcal{I}} q_i S_i$ where S_i is the transmission signal power at state i. Moreover, we characterize the channel states based on the received signal to noise ratio (γ). It is then straightforward to show that the following policy results in channel capacity: $S(\gamma)/\bar{S} = 1/\gamma_c - 1/\gamma$ if $\gamma \geq \gamma_c$ and 0 otherwise [2], where γ_c is the cut-off signal to noise ratio which is set so that the constraint on average signal power is met. If we now assume that the source has the distortion rate function $D(R) = 2^{-\beta R}$ then the optimum policy that results in minimum end-to-end distortion is given by:

$$\frac{S(\gamma)}{\bar{S}} = \begin{cases} \left(\frac{1}{\gamma^{\beta}\gamma_c}\right)^{\frac{1}{\beta+1}} - \frac{1}{\gamma} & \gamma \geq \gamma_c \\ 0 & \gamma < \gamma_c \end{cases} \qquad (2)$$

which is dependent on the source through β [3]. In fact, the more convex the distortion-rate function (the higher the value of β) the more dissimilar the above policies become. Figure 1 shows the performance of these two policies over narrow-band Rayleigh fading channel where the received SNR γ has exponential distribution $(f(\gamma) = 1/\gamma_s \exp(-\gamma/\gamma_s))$.

References

[1] C.E. Shannon, "Channels with side information at the transmitter," IBM J.R.D. , Vol. 2, pp. 289-293, 1958.

[2] A. Goldsmith, *Design and Performance of High-Speed Communication Systems over Time-Varying Radio Channels*, Ph.D. dissertation, University of California at Berkeley, 1994.

[3] M. Khansari and M. Vetterli, "Joint source-channel coding for time-varying channels," *IEEE Information theory workshop on information theory, multiple access and queueing, St. Louis*, April 1995.

[4] M. Khansari and M. Vetterli, "Transmission of source with fidelity criterion over time-varying channels with informed transmitter and receiver," *To be submitted.*

One and Two Dimensional Parallel Partial Response for Parallel Readout Optical Memories

Brita H. Olson and Sadik C. Esener[1]

University of California, San Diego
Department of Electrical and Computer Engineering
La Jolla, CA 92093-0407 USA

Abstract -- **We extend Partial Response (PR) precoding [1] to two-dimensions and consider it, as well as parallel one-dimensional (1D) PR, for use in parallel readout optical memory systems. We also develop expressions for optically implementable two-dimensional (2D) zero-forcing equalizers to be used in conjunction with these forms of PR precoding.**

I. SUMMARY

Figure 1 depicts a behavioral model of an array of abutted rectangular pixels being retrieved in parallel from a memory with a coherent imaging readout system. The transfer function, $H(f_x, f_y)$, describes the 2D spatial bandlimiting of the readout system and also includes a frequency description of the shape of the pixels in the memory.

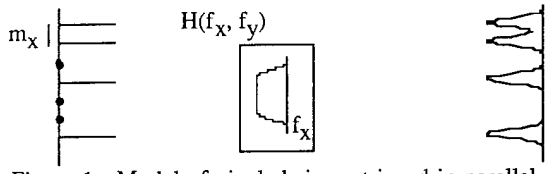


Figure 1: Model of pixels being retrieved in parallel.

Figure 2 shows the reconstruction of an array of binary phase pixels (with values +1 and -1) that have been precoded using what we term 1D *(1+D)* PR precoding. ZERO values are formed by the overlap of two pixels with opposite signs. ONE values are obtained by the overlap of pixels with the same sign. Detection of intensity takes place halfway between the centers of the two pixels used to form the desired data value. 1D strips can be read out in isolation [2] or can comprise the rows or columns of a 2D array.

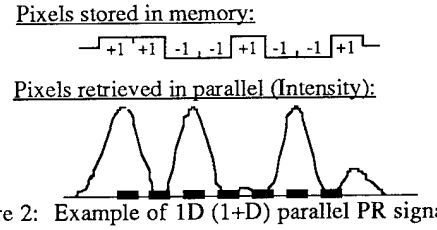


Figure 2: Example of 1D (1+D) parallel PR signaling

2D arrays can also be precoded using 2D PR precoding. With 2D (1+D) PR precoding, each data value is formed at the center of four overlapping reconstructed pixels as illustrated in Figure 3. This form of precoding can be applied to 2D arrays that experience spatial bandlimiting or to 1D arrays read out in succession that are broadened temporally. We introduce two shift operators D_x and D_y to describe this 2D broadening. With this notation, the system polynomial for a reconstructed pixel broadened in two-dimensions can be written as: $\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} a_{ij} D_x^i D_y^j$

To accomplish 2D PR precoding, a 2D array can be thought of as a 1D array, precoded as such, and then returned to its 2D format.

[1]This research was supported by ONR grant #N0014-93-I-0414 and by the AFOSR under grants #F49620-93-I-0057 and #F49620-93-I-0371.

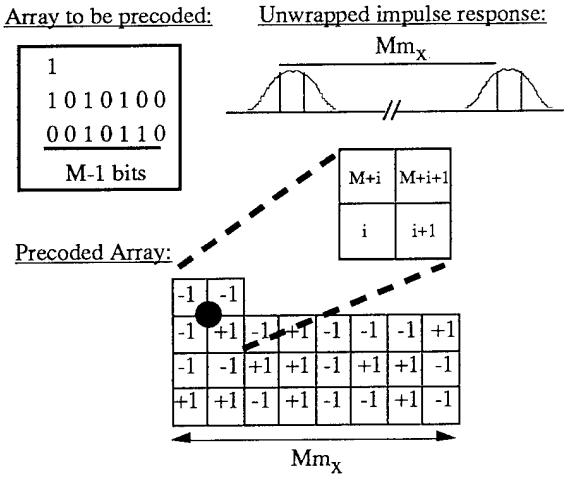


Figure 3: $(1+D_x)(1+D_y)$ PR precoding example.

Towards this end, the 2D system polynomial is made into a 1D system polynomial by substituting D for D_x and D^M for D_y, where M is (N_y-1) plus the number of bits in each row of the 2D array to be precoded. For $(1+D_x)(1+D_y)$ PR precoding performed serially, one would arbitrarily choose the first row of the precoded array and the first bit of the next row. $(1+D+D^M+D^{M+1})$ PR precoding is then applied to bits read from the 2D array to be precoded, as described above. After precoding M-1 bits, an additional arbitrary bit would be inserted in the input bit stream of the precoder to start a new row.

Zero-forcing equalizers for both 1D and 2D PR signaling applied to 2D arrays of binary phase pixels are easily represented in the Fourier domain by extending the work in reference [3]. These equalizers can be implemented as apodizers in the Fourier plane of an optical system. Table 1 lists the equalizers and overall transfer functions for minimum bandwidth one-to-one imaging systems using $(1+D_x)$ and $(1+D_x)(1+D_y)$ signaling.

Signaling	Transfer Function	Equalizer
$(1+D_x)$	$\cos(\pi f_x)\,\mathrm{rect}(f_x)$	$\dfrac{\cos(\pi f_x)\,\mathrm{rect}(f_x)}{\mathrm{sinc}(f_x)\,\mathrm{sinc}(f_y)}$
$(1+D_x)(1+D_y)$	$\cos(\pi f_x)\cos(\pi f_y)$	$\dfrac{\cos(\pi f_x)\cos(\pi f_y)}{\mathrm{sinc}(f_x)\,\mathrm{sinc}(f_y)}$

Table 1: Phase terms are omitted. This is compensated for by detecting between the centers of reconstructed pixels

REFERENCES

[1] A. Lender, "Correlative level encoding for binary data transmission," *IEEE Spectrum*, vol. 3, no. 2, p. 104, 1966.

[2] B. H. Olson and S. E. Esener "Partial response precoding for parallel-readout optical memories," *Opt. Lett.* 19, p. 661, May 1994.

[3] L. C. Barbosa, "Characterization of Minimum Noise Partial Response Channels", *IBM Research Report RJ 6475 (62948)*, Oct. 1988.

MAXIMUM LIKELIHOOD DECODING OF BLOCK CODES ON (1-D) PARTIAL RESPONSE CHANNELS

G.Markarian, B.Honary
Communications and Signal Processing Research Consortium,
Lancaster University, Lancaster LA1 4YR, UK
E-Mail: G.Markarian@cent1.lancs.ac.uk

Abstract- A novel technique for trellis decoding of block both RLL and balanced codes on PR channels is described. The technique allows performance improvement without increament of decoder complexity.

1. INTRODUCTION

Recently, a simple technique for constructing run length limited (RLL) block error control codes (ECC) together with their minimal trellises has been introduced [1,2]. The procedure adapted for the design of such codes is based on taking a linear ECC and incorporating a maximum runlength constraint by carefully modifying the basis code while retaining the minimum distance properties of the parent code. Such codes are particularly suited for magnetic recording applications where the (1-D) partial response (PR) channel provides a good model at low information density rates [3].
In this paper we show that the trellis decoder of these codes has the same trellis structure as the encoder, thus the additional decoding complexity is avoided. We also describe how the trellis diagram of the non-linear balanced codes can be incorporated within the PR channel.

2. DECODING OF LINEAR BLOCK CODES ON PARTIAL RESPONSE CHANNELS

When binary sequences are transmitted over the (1-D) PR channel the received noiseless sequence is ternary and due to the memory of the PR channel contains some additional structure that can be exploited to improve the error performance. For uncoded data, MLD in PR channel can be realised by using a Viterbi decoder, because the memory introduced by the (1-D) channel has a trellis structure [4]. Similarly, for RLL/ECCs MLD can readily be achieved by incorporating the trellis diagram of the code within the decoding trellis of the PR channel. Furthermore, the complexity of the trellis does not increase because the modified RLL codes have an odd number of 1s in the labelling of each branch of the trellis and hence the state of the PR channel is the same for all branches emanating from the same state. Thus all that need to be changed is the branch labels of the RLL/ECC trellis.

3. SIMULATIONS RESULTS AND CONCLUSIONS

The effect of this technique on the decoder performance for some codes has been derived by the comparison of simulation results for the new decoding strategy with the conventional approach. It has been found that the technique provides a performance improvement exceeding 2 dB at error rates of about 10^{-5} .
The technique has also been applied for the balanced codes. Although these codes are non-linear, they possess a regular trellis structure which allows their Viterbi decoding. The technique has been applied to the (16,9,6,5,4) code [5] and simulation results have proved the efficiency of the technique.

REFERENCES

1. Markarian G., Naderi M., Honary B., Popplewell A., O'Reilly J. "Maximum likelihood decoding of RLL -FEC array codes on partial response channels", *Electronics Letters,* vol.29, 1993, No.16, pp.1406-1408.
2. Markarian G., Honary B. "Trellis decoding technique for block RLL/ECC". *IEE Proceedings - Communications,* vol.141, 1994, No.5, pp.297-302.
3. Wolf J.K., Ungerboeck G. "Trellis coding for partial-response channels", *IEEE Transactions on Communications,* vol.34, 1986, No.8, pp.765-773.
4. Kobayashi H., Tang D.T. "Application of partial response channel coding to magnetic recording system." *IBM Journal Research and Development,* vol.14, 1970, pp.368-375.
5. Markarian G., Honary B., Blaum M. "Maximum likelihood trellis decoding technique for balanced codes", *Electronics Letters,* vol.31, 1995, No.6, pp.447-448.

Constrained Block Codes for Class-IV PRML Channels[1]

Khaled A. S. Abdel-Ghaffar* Jos H. Weber**

*University of California, Dept. of Elec. & Comp. Eng., Davis, CA 95616, USA

**Delft University of Technology, Dept. of Elec. Eng., 2600 GA Delft, The Netherlands

Abstract — **We investigate sets of maximal number of fixed-length sequences that can be concatenated without violating certain constraints often required in class-IV PRML channels used in magnetic recording.**

I. Introduction

PRML is a technique that combines partial-response (PR) signaling with maximum-likelihood (ML) sequence estimation in order to combat intersymbol interference and noise, which are common in high density digital magnetic storage channels [3]. One of the most common partial response channels used in magnetic recording is the class-IV channel. Such channel processes independently the even and the odd subsequences of bits with even and odd indices, respectively. Hence, a Viterbi detector can be separately applied to each of the even and the odd output subsequences to obtain maximum-likelihood estimates of the input subsequences.

In order to limit the path memory of the Viterbi detector, the number of consecutive zeros in each of the input subsequences is upper bounded by some positive integer I. Also to maintain clock synchronization, the number of consecutive zeros in the global input sequence is upper bounded by some positive integer G. We say that a binary sequence satisfies the $(0, G/I)$ constraint if it satisfies the two constraints specified by G and I. Notice that if the number of consecutive zeros in each of the even and the odd subsequences of a sequence is upper bounded by I, then the number of consecutive zeros in the sequence itself is upper bounded by $2I$. Hence, we assume in the following that G and I are positive integers such that $G \leq 2I$. Coding schemes are used to map unconstrained sequences of data into $(0, G/I)$ constrained sequences for transmission over the channel [2],[3]. In this paper, we consider schemes based on block codes.

II. $(0, G/I)$ Constrained Block Codes

A $(0, G/I)$ block code is a set of $(0, G/I)$ constrained binary sequences, called codewords, of fixed length such that any juxtaposition of a finite number of codewords is also $(0, G/I)$ constrained. For given G and I, let $\mathcal{M}^{G,I}_{l,r|l_1,l_2;r_1,r_2}(n)$ be the set of all $(0, G/I)$ constrained sequences $(\gamma_1, \gamma_2, \ldots, \gamma_n)$ of length n with at most l, l_1, and l_2 leading zeros at the beginning of the sequence $(\gamma_1, \gamma_2, \ldots, \gamma_n)$, its odd subsequence $(\gamma_1, \gamma_3, \ldots, \gamma_{2\lceil n/2 \rceil - 1})$, and its even subsequence $(\gamma_2, \gamma_4, \ldots, \gamma_{2\lfloor n/2 \rfloor})$, respectively, and at most r, r_1, and r_2 leading zeros at the beginning of the reversed sequence $(\gamma_n, \gamma_{n-1}, \ldots, \gamma_1)$, its odd subsequence $(\gamma_n, \gamma_{n-2}, \ldots, \gamma_{\lfloor n/2 \rfloor - \lceil n/2 \rceil + 2})$, and its even subsequence $(\gamma_{n-1}, \gamma_{n-3}, \ldots, \gamma_{\lceil n/2 \rceil - \lfloor n/2 \rfloor + 1})$, respectively. Let $M^{G,I}_{l,r|l_1,l_2;r_1,r_2}(n)$ be the cardinality of $\mathcal{M}^{G,I}_{l,r|l_1,l_2;r_1,r_2}(n)$. Any $(0, G/I)$ constrained block code of length n is a subset of $\mathcal{M}^{G,I}_{l,G-l|l_1,l_2;I-l_2,I-l_1}(n)$ for some l, l_1, and l_2. Conversely, if n is sufficiently large, any subset of $\mathcal{M}^{G,I}_{l,G-l|l_1,l_2;I-l_2,I-l_1}(n)$ forms a $(0, G/I)$ constrained block code. Hence, to construct efficient $(0, G/I)$ block codes of length n, it is important to determine an option (l, l_1, l_2) that maximizes $M^{G,I}_{l,G-l|l_1,l_2;I-l_2,I-l_1}(n)$. Such option will be called optimal for the given G, I, and n. Two special cases were investigated by Eggenberger and Patel [1]. They determined that the option $(2, 2, 2)$ is optimal in case $G = I = 4$ and $n = 9$, while the option $(1, 3, 3)$ is optimal in case $G = 3$, $I = 6$, and $n = 9$.

III. Results

The main contribution of this paper is presenting general results concerning optimal options for all values of G, I, and n. The results are given in the following three theorems which address the cases $G = 1$, $G \geq 2$ and I is even, and $G \geq 2$ and I is odd, respectively.

Theorem 1 *For $G = 1$, $I \geq 1$, and $n \geq 1$, $(0, 0, \lfloor I/2 \rfloor)$ is an optimal option.*

Theorem 2 *For $2 \leq G \leq 2I$, I is even, and $n \geq 1$, $(\lfloor G/2 \rfloor, I/2, I/2)$ is an optimal option, except in the case $G = I = 2$ and $n = 6$ where the option $(0, 0, 1)$ is optimal.*

Theorem 3 *For $2 \leq G \leq 2I$, I is odd, and $n \geq 1$, at least one of the following three options is optimal:*

$$(\min\{\lfloor G/2 \rfloor, I-1\}, (I-1)/2, (I-1)/2),$$

$$(\min\{\lfloor G/2 \rfloor, I-1\}, (I-1)/2, (I+1)/2),$$

$$(\lfloor G/2 \rfloor, (I+1)/2, (I-1)/2).$$

Theorems 1 and 2 explicitly specify an optimal option in case $G = 1$ or I is even. In particular, the results of Eggenberger and Patel follow as special cases of Theorem 2. In case $G \geq 2$ and I is odd, our results specify three candidates for an optimal option. In general, the optimal options in this case may depend very much on the length n as demonstrated in the following result.

Theorem 4 *For $G = 2$, $I = 1$, and $n \geq 1$, $(1, 1, 0)$ is an optimal option if $n = 1$ or $n \not\equiv 1 \pmod 4$, and $(0, 0, 0)$ is an optimal option if $n \neq 1$ and $n \equiv 1 \pmod 4$.*

References

[1] J. S. Eggenberger and A. M. Patel, "Method and apparatus for implementing optimum PRML codes," U.S. Patent 4,707,681, Nov. 17, 1987.

[2] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Sel. Areas Commun.*, vol. SAC-10, no. 1, pp. 5–37, Jan. 1992.

[3] P. H. Siegel and J. K. Wolf, "Modulation and coding for information storage," *IEEE Commun. Mag.*, pp. 68–86, vol. 29, no. 12, Dec. 1991.

[1] Most of this work was done while the first author was visiting the Dept. of Elec. Eng., Delft Univ. of Tech. The first author was also supported in part by NSF under grant NCR 91-15423.

On Efficient High-Order Spectral-Null Codes*

L. G. Tallini†, S. Al-Bassam‡ & B. Bose†

†Department of Computer Science, Oregon State University, Corvallis OR, 97331.

‡Computer Science Department, King Fahad University of Petroleum & Minerals, Dhahran, Saudi Arabia 31261.

E-mail: tallini@mail.cs.orst.edu, albassam@ccse.kfupm.sa, bose@cs.orst.edu

Abstract — **Let $\mathcal{S}(N,q)$ be the set of all words of length N over the bipolar alphabet $\{-1,+1\}$, having a q-th order spectral-null at zero frequency. Any subset of $\mathcal{S}(N,q)$ is a spectral-null code of length N and order q. In this paper, we give an equivalent formulation of $\mathcal{S}(N,q)$ in terms of codes over the binary alphabet $\{0,1\}$. We show that $\mathcal{S}(N,2)$ is equivalent to a well known class of single error correcting, all unidirectional error detecting (SEC-AUED) codes. We derive an explicit expression for the redundancy of $\mathcal{S}(N,2)$. Further, we give new efficient recursive design methods for second-order spectral-null codes, improving the redundancy of the codes found in the literature.**

Regard the alphabet $\{-1,+1\}$ as a subset of the real field. The following characterization of $\mathcal{S}(N,q)$ is well known [7], [5] (x_i denotes the i-th component of a vector X):

$$\mathcal{S}(N,q)=\left\{X\in\{-1,+1\}^N\colon \sum_{j=1}^{N} x_j j^i=0,\ i=0,\ldots,q-1\right\}.$$

The problem of finding an explicit expression for the redundancy of $\mathcal{S}(N,q)$ was left open in [7]. Using a well known result in number theory (the problem of partitioning a natural number n into w distinct natural numbers less than or equal to a certain bound b), we are able to derive the following explicit expression for the redundancy of $\mathcal{S}(N,2)$:

$$N-\lfloor\log_2|\mathcal{S}(N,2)|\rfloor\simeq 2\log_2 N-1.141,\quad N \text{ multiple of 4.}\qquad(1)$$

Further, by replacing the symbol -1 with 0 and $+1$ with 1 we prove that $\mathcal{S}(N,q)$ is equivalent to the code

$$\mathcal{S}(N,q)=\left\{X\in\{0,1\}^N\colon \sum_{j=1}^{N} x_j j^i=\frac{1}{2}\sum_{j=1}^{N} j^i,\ i=0,\ldots,q-1\right\},$$

where the sums and the products are over the real field. Since $\sum_{j=1}^{N} x_j j^i$ is an integer number, if $S(N,q)\neq\emptyset$ then $\sum_{j=1}^{N} j^i$ must be even for all $i=0,\ldots,q-1$. Note that

$$\mathcal{S}(N,2)=\left\{X\in\{0,1\}^N\colon \sum_{j=1}^{N} x_j=\frac{N}{2}\ \text{and}\ \sum_{j=1}^{N} x_j j=\frac{N(N+1)}{4}\right\}.$$

This is nothing but a particular group theoretic single error correcting and all unidirectional error detecting (SEC-AUED) code over $(\mathbb{Z},+)$ [3]. Clearly, if N is not a multiple of 4 then $S(N,2)=\emptyset$. A binary code $\mathcal{C}$ is a q-th order spectral-null code of length N with k information bits iff 1) $\mathcal{C}$ is a subset of $\mathcal{S}(N,q)$ and 2) $\mathcal{C}$ has 2^k codewords. The authors in [7], presented a recursive method to encode k information bits into a second-order ($q=2$) spectral-null code of length

$$\tilde{N}(k)=\tilde{n}(k)+\tilde{N}(2\lceil\log_2\tilde{n}(k)\rceil-1),\qquad(2)$$

where $\tilde{n}(k)$ is the smallest integer $\tilde{n}$ such that 1)

$$\tilde{n}-k\geq\lceil\log_2\tilde{n}\rceil+1.\qquad(3)$$

and 2) $\tilde{n}$ is a multiple of 4. Here, we give a new efficient recursive method to encode k information bits into a second-order spectral-null code (over the alphabet $\{0,1\}$) of length

$$N(k)=n(k)+N(\lceil\log_2(n(k)\cdot(n(k)-1))\rceil-1),\qquad(4)$$

where $n(k)$ is the smallest integer n such that 1) There exist a first-order spectral-null code of length n with k information bits and 2) n is a multiple of 4. Note that, a first-order spectral-null code is nothing but a balanced code [6]. At present, there exist many efficient balanced code designs which require less than $\log_2 k$ check bits (i.e. $n-k<\log_2 k$) to make a k bit data word balanced [1], [2], [4], [6], [8], and so $\tilde{n}(k)>n(k)$ for infinitely many values of k (see (3)). Comparing relations (2) and (4) it is then clear that, for these k's, we get less redundant codes than those presented in [7]. In our design methods, first, the data word is converted into a balanced word, which in turn is converted into a second-order spectral-null codeword. One of the proposed methods is briefly described here.

Let n be a multiple of 4. Given $X\in\{0,1\}^n$, let $s(X)=\sum_{j=1}^{n} x_j j$ and $w(X)=\sum_{j=1}^{n} x_j$. For $i=0,\ldots,n(n-1)/2$, let $X^{(i)}$ be the binary vector obtained from X by applying the first i exchanges of adjacent components starting from the first component. For example, when $n=4$, $X^{(0)}=X=x_1x_2x_3x_4$, $X^{(1)}=\underline{x_2x_1}x_3x_4$, $X^{(2)}=x_2\underline{x_3x_1}x_4$, $X^{(3)}=x_2x_3\underline{x_4x_1}$, $X^{(4)}=\underline{x_3x_2}x_4x_1$, $X^{(5)}=x_3\underline{x_4x_2}x_1$, $X^{(6)}=\underline{x_4x_3}x_2x_1$. A data word $Y\in\{0,1\}^k$ is encoded as follows.

Encoding Procedure:

1) Balance Y using one of the methods given in [1], [2], [4], [6], [8]. Let X be the codeword of length $n=n(k)$ associated with Y. Note that $w(X)=n/2$.
2) Compute $X^{(i_0)}$, where i_0 is an integer $i\in[0,n(n-1)/2-1]$ (for example the smallest) such that $S(X^{(i)})=n(n+1)/4$.
3) Recursively apply this encoding procedure to the binary representation of i_0. Let $E(i_0)$ be the codeword associated with i_0.
4) Concatenate $E(i_0)$ to $X^{(i_0)}$ to get the codeword $E(Y)=X^{(i_0)}E(i_0)$.

Decoding of the received word $\tilde{X}$ I can be done easily once it is known that $i_0=E^{-1}(I)$. In the paper, we give similar procedures which require only $O(n\log_2 n)$ bit operations.

Example: *Let $k=32$. Using the second construction proposed in [8], it is possible to encode the 32 data bits into a balanced code of length $n=n(32)=36$. In this case, the length of the code is $N(32)=36+N(\lceil\log_2(36\cdot35)\rceil-1)=36+N(10)$. Assume that*

$$X=0110\,1011\,1011\,1100\,0000\,0000\,0000\,1111\,1111$$

is the balanced encoding of a data word $Y\in\{0,1\}^{32}$ that needs to be encoded. Since 28 is the smallest integer i such that $S(X^{(i)})=n(n+1)/4=36\cdot37/2=333$, then Y is encoded as $E(Y)=$

$$X^{(28)}E(28)=1101\,0111\,0111\,1000\,0000\,0000\,0001\underline{10}111\,1111\ E(28).$$

Using a table look-up it is possible to encode 10 information bits into a second-order spectral-null code of length 20 (see (1)). This means that we can encode 32 data bits into a second-order spectral-null code of length $N(32)=36+N(10)=36+20=56$ (instead of 60 which is what we would get using the method proposed in [7]).

References

[1] S. Al-Bassam and B. Bose, *Design of efficient balanced codes*, IEEE Trans. Comput., vol. 43, pp. 362–365, March 1994.

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

[3] B. Bose and T. R. N. Rao, *Theory of unidirectional error correcting/detecting codes*, IEEE Trans. Comput., vol C-31, pp. 521–530, June 1982.

[4] B. Bose, *On unordered codes*, IEEE Trans. Comput., vol 40, pp. 125–131, Feb. 1991.

[5] K. A. S. Immink, *Coding Techniques for Digital Recorders*, London: Prentice-Hall, 1991.

[6] D. E. Knuth, *Efficient balanced codes*, IEEE Trans. Inform. Theory, vol. IT-32, pp. 51–53, Jan. 1986.

[7] R. M. Roth, P. H. Siegel and A. Vardy, *High-order spectral-null codes – constructions and bounds*, IEEE Trans. Inform. Theory, vol. 40, pp. 1826–1840, Nov. 1994.

[8] L. Tallini, R. M. Capocelli and B. Bose, *Design of some new efficient balanced codes*, IEEE Trans. Inform. Theory (To appear).

*This work is supported by the grant from National Science Foundation MIP-9404924. The first author's work is supported by the Italian National Research Council (fellowship CNR 203.01.62).

Viterbi Decoding Considering Insertion/deletion Errors

Takuo Mori Hideki Imai
Institute of Industrial Science, University of Tokyo Tokyo, Japan

Abstract — **In this paper, we propose a Viterbi decoding based on Levenshtein distance. We show that Levenshtein distance is suitable for metric in a channel where both substitution errors and insertion/deletion errors occur. Our proposal makes it possible to continue Viterbi decoding without re-synchronization even if some insertion/deletion errors occur in a channel.**

I. Introduction

Recently, high recoding density is required in digital recording systems. However, as recording density increases, error rate also increases. Especially, it is known that burst errors called a synchronization error caused by insertion/deletion errors occur in optical recording systems more frequently than in other disk systems. In this field, Partial Response Maximum Likelihood (PRML) detection is focused on now. In PRML systems, Viterbi decoding is used in order to realize maximum likelihood decoding. In Viterbi decoding based on Hamming or Euclidean distance, even if just an insertion/deletion error occurs, it is impossible to continue decoding without re-synchronizing, because insertion/deletion errors measured by Hamming or Euclidean distance cause a burst error called synchronization error. In this paper, we propose Viterbi decoding based on Levenshtein distance [1].

II. Channel Model

In this section, we talk about a binary channel model in which not only substitution errors but also insertion/deletion errors occur. In this paper, we call such a channel C_{SID}. Let p be the probability of substitution errors, q_i be the probability of insertion errors and q_d the probability of deletion errors in C_{SID} respectively. In this paper, for convenience, we assume that $q_i = q_d = q$.

III. Levenshtein Distance

Definition 1 *Let x and y be two finite sequences of symbols from a given alphabet. If x can be transformed into y by the substitution of E_i symbols, the insertion of f_i symbols and the deletion of g_i symbols, then the Levenshtein distance (LD) between x and y is defined by*

$$LD(x, y) = \min_i(E_i + f_i + g_i). \quad (1)$$

Notice that Levenshtein distance satisfies three axioms of metric.

Levenshtein distance is computed by using a graph that we call a LD diagram (See Fig1).

This work was presented in part at the IEICE Technical Report, July 15, 1995.

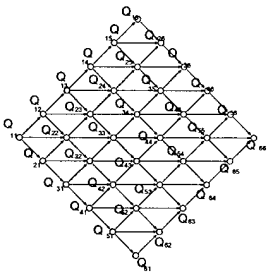


Figure 1: LD diagram for $n = 5$

IV. Conditional Probability

In this section, we consider the conditional probability $P(y_t|w_t)$ in C_{SID}. In Binary Symmetric Channel (BSC), the conditional probability $P(yt|wt)$ is given by the following equation.

$$P(y_t|w_t) = p^E(1-p)^{n-E} \quad (2)$$

where E is the number of substitution errors that occur in BSC. In this case, $-\log P(y_t|w_t)$ is proportional to the number of substitution errors, that is, Hamming distance

In C_{SID}, there are many way that w_t changes y_t by both substitution errors and insertion/deletion errors. The number of ways that w_t changes y_t is given by the number of paths in LD diagram. Thus, the conditional probability $P(y_t|w_t)$ is given by the sum of the probability of each path in LD diagram. Thus,

$$P(y_t|w_t) = \sum_i p^{E_i}(1-p)^{l-E_i}q^{f_i}q^{g_i} \quad (3)$$

where $0 \le i \le {}_{2n}\mathrm{C}_n$, $k = f_i = g_i$, $l = n - k$, and E_i, f_i and g_i are the number of substitution errors, insertion errors and deletion errors in each path, respectively. In this case, what is proportional to $-\log P(y_t|w_t)$? Let d_i be $d_i = E_i + f_i + g_i$, and p_i be a path labeled i, and $P(p_i) = m$ be the probability of p_i. Assume that d_i is the minimum value for all i. Here, we consider $P(p_j)$.

In the case that $E_j = E_i + 1$ and $d_j = d_i + 1$, then $P(p_j) < pm$. In the case that $f_j = f_i + 1, g_j = g_i + 1$ and $d_j = d_i + 2$, then $P(p_j) < q^2m$. Thus, if p, q are relatively small, $P(p_j) \ll P(p_i)$. Thus, it can be said that the value of $P(y_t|w_t)$ in C_{SID} mainly depends on $P(p_i)$. Then, it has been shown that $-\log P(y_t|w_t) \propto \min_i(E_i + f_i + g_i) = d_{LD}(y_t, w_t)$, which shows that it is possible to continue Viterbi decoding appropriately by using Levenshtein distance as metric.

V. Conclusion

In this paper, we have shown that Levenshtein distance is suitable for metric in C_{SID}. This indicates that it is possible to continue Viterbi decoding in C_{SID} by using Levenshtein distance as metric.

References

[1] T. Okuda, E. Tanaka and T. Kasai, "A Method for the Correction of Garbled Words Based on the Levenshtein Metric," *IEEE Trans. Comput.*, vol. c-25, no. 2, Feb. 1976.

Further Results on Cosets of Convolutional Codes with Short Maximum Zero-Run Lengths[1]

Kjell Jørgen Hole and Øyvind Ytrehus
Department of Informatics, University of Bergen, Norway

Abstract — **We study the maximum zero-run length, L_{max}, of cosets of convolutional codes, and show that an associated block subcode to a large extent determines L_{max}.**

I. Introduction

A communication system or storage system may use a coset of a binary convolutional code for both symbol synchronization and error control. To achieve symbol synchronization, the coset must have a short maximum zero-run length, L_{max}. The shortest values of L_{max} can be found in the class of convolutional codes of rate $(n-r)/n$ for which at least one row of a minimal parity check matrix is nonpolynomial [1]. We focus on this class. Each convolutional code $\mathcal{C}$ in the class contains an associated block subcode $\mathcal{C}_B$, consisting of the union of the sets of binary labels in the convolutional code trellis. For any binary vector $\mathbf{p}$ of length n, let $\overline{\mathbf{p}}$ denote the sequence obtained by repeating $\mathbf{p}$ indefinitely. We consider some coset $\mathcal{C}+\overline{\mathbf{p}}$, obtained by adding some vector $\mathbf{p}$ to every binary label in the convolutional code trellis. It is convenient to express L_{max} as $L_{max} = max\{LR, PR\}$, where the label run, LR, is defined as the largest number of intermediate zeros between two ones in any label of $\mathcal{C}+\overline{\mathbf{p}}$ (of Hamming weight at least two), and the path run, PR, is the largest number of consecutive zeros in any sequence $\mathcal{C}+\overline{\mathbf{p}}$ consisting of two or more consecutive coset labels.

II. The Connection between Zero-Run Lengths of Convolutional Code Cosets and Associated Block Codes

We consider an $(n, n-r)$ convolutional code $\mathcal{C}$ defined by a parity check matrix $\mathbf{H}(D)$. The maximum degree of the i-th row of $\mathbf{H}(D)$ is denoted ν_i. Assume that the first r_B rows of $\mathbf{H}(D)$ are nonpolynomial, and that the remaining $r-r_B$ rows are sorted according to increasing row degree. That is, $\nu_1 = \ldots = \nu_{r_B} = 0 < \nu_{r_B+1} \leq \ldots \nu_r$. The associated block subcode $\mathcal{C}_B$ is the $(n, n-r_B)$ block code defined by the submatrix $\mathbf{H}_B$ which consists of the first r_B rows of $\mathbf{H}(D)$. Let PR and LR be the path run and label run, respectively, of the coset $\mathcal{C}+\overline{\mathbf{p}}$. If we view $\mathcal{C}_B$ as a zero constraint length (or one state) convolutional code, we can let PR_B and LR_B be the "path run" and label run of $\mathcal{C}_B+\mathbf{p}$. Then we can show the following results.

Lemma 1 *$PR \leq PR_B$. Further, $PR = PR_B$ if $\nu_{r_B+1} \geq 2$.*

Lemma 2 *$LR = LR_B$.*

[1]This work was supported by the Norwegian Research Council (NFR) under contract numbers 107542/410 and 107623/420.

III. Block Code Zero-Run Lengths

For $1 \leq i \leq r_B$, let λ_i and ρ_i be the first and last position where the i-th row of $\mathbf{H}_B$ is nonzero. Assume, without loss of generality, that the rows of $\mathbf{H}_B$ are sorted according to $\rho_i \leq \rho_{i+1}$, $1 \leq i < n$.

Lemma 3 *For the coset with all-one syndrome, $LR_B \leq \max_{1\leq i\leq r_B+1}\{\rho_i - \max_{0\leq j<i}\{\lambda_j\}\} - 1\}$, where, for convenience, we define $\lambda_0 = 1$ and $\rho_{r_B+1} = n$.*

Lemma 4 *For the coset with all-one syndrome, $PR_B \leq n - \max_{1\leq i\leq r_B}\{\lambda_i\} + \rho_1 - 1\}$.*

IV. Convolutional Code Zero-Run Lengths

Definition 1 Let $\mathcal{D}$ be the class of $(n, n-2, (0, \nu))$, $\nu \geq 2$, binary convolutional codes with the first row of the parity check matrices equal to

$$1 \ldots 1 \quad \overbrace{0 \ldots 0}^{t \text{ zeros}}$$

where t is the number of trailing zeros, $0 \leq t \leq n-2$. □

Theorem 1 *Let $\mathcal{C}$ be a convolutional code in the class $\mathcal{D}$. Any coset $\mathcal{C}+\overline{\mathbf{p}}$ for which the first syndrome sequence is equal to the all-one sequence has the least $L_{max} = 2n-2-t$ for any period n coset representative.*

Definition 2 For $r \geq 3$, let $\mathcal{E}$ be the class of $(n, n-r, (0, 0, \nu_3, \ldots, \nu_r))$ binary convolutional codes for which the two first rows ($\mathbf{H}_B$) of the parity check matrices $\mathbf{H}(D)$ are of the form

$$\begin{array}{ccccccccccc} & & & & & & & & \overbrace{\quad\quad}^{t_1 \text{ zeros}} & & \\ ? & \ldots & ? & 1 & 0 & \ldots & 0 & & & & \\ \underbrace{0 \ldots 0}_{l_2 \text{ zeros}} & 1 & ? \ldots ? & 1 & \underbrace{0 \ldots 0}_{t_2 \text{ zeros}} & & & & & & \end{array}$$

The question marks denote unknown binary digits, t_1 denotes the number of trailing zeros in the first row, and l_2 and t_2 denotes the number of leading and trailing zeros in the second row. It is assumed that $1 \leq l_2 + t_2 \leq n-2$, $l_2 \geq t_2$ and $t_2 \leq t_1 \leq n-2$. □

Theorem 2 *Let $\mathcal{C}$ be a convolutional code in the class $\mathcal{E}$. Any coset $\mathcal{C}+\overline{\mathbf{p}}$ for which the first and second syndrome sequences are equal to the all-one sequence has $L_{max} \leq \max\{n-2-t_2, 2n-2-l_2-t_1\}$.*

We also show, by way of examples, that classes $\mathcal{D}$ and $\mathcal{E}$ contain codes with excellent error-correcting properties.

References

[1] K. J. Hole, "Cosets of convolutional codes with short maximum zero-run lengths," *IEEE Trans. Inform. Theory*, vol. IT-41, July 1995.

A new coding technique: Integer Multiple Mark Modulation (IMMM)

Calvin Menyennett and Hendrik C. Ferreira
Cybernetics Laboratory, Rand Afrikaans University, P O Box 524, Auckland Park, 2006, South Africa.

Abstract—**A new modulation coding technique, called Integer Multiple Mark Modulation (IMMM) is proposed. IMMM codes generate asymmetrical runlength limited sequences with spectral nulls in the power spectrum, whose positions are related to a specific runlength constraint. This coding technique can be used for any channel requiring specific spectral nulls in the code spectrum, such as partial-response optical recording.**

I. Introduction

Numerous applications of digital data transmission and storage systems require the use of runlength limited (RLL) codes with certain defined spectral properties. We investigate a method to furnish codes with spectral nulls in the power spectrum (except DC) of the encoded sequence. An application of such codes is providing a gap for the insertion of auxiliary pilot tones, used for positioning the servo of magnetic or optical disc recorders [1]. In another application, codes with spectral nulls in the power spectrum, which coincide with the nulls of the transfer function of the channel, are used in partial-response optical recording systems [2].

II. The IMMM Coding Technique

The notation for asymmetrical runlength limited (ARLL) sequences, as introduced by Karabed and Siegel [3], will be used. The class of binary, non-return-to-zero (NRZ), ARLL channels can be defined by the 4-tuple (d', k', e', m'), where d' and k' are the minimum and maximum runlength of 0's, respectively, and e' and m' are the minimum and maximum runlength of 1's, respectively. The Integer Multiple Mark Modulation (IMMM) coding technique, which we introduce, requires the runlengths of 1's to be of the form je', $1 \le j \le m'/e'$ (*i.e.* integer multiples of e'). The 1's are usually referred to as written marks in optical storage thus the name "Integer Multiple Mark Modulation".

Even Mark Modulation (EMM) was introduced by Karabed and Siegel [3] to improve the performance of input-restricted partial-response optical recording channels. EMM satisfies the runlength constraint $(d', k', e', m') = (1, \infty, 2, \infty)$ and the requirement that the written marks are of *even* length [3]. The EMM coding technique is therefore a special case of the Integer Multiple Mark Modulation technique.

The IMMM coding technique has the interesting property that it has spectral nulls at rational submultiples of the symbol frequency, the position of which can be chosen merely by adjusting the minimum runlength of 1's:

Proposition 1

An IMMM (d', k', e', m') sequence, with $k' > d' \ge 1$, and $m' > e' \ge 1$, will contain spectral nulls at the frequencies $f = rf_s/e'$, with $r \in \{1, 2, 3, ..., e'-1\}$ and where f_s is the symbol frequency. ■

To calculate the channel capacity for these sequences, the following proposition can be used:

Proposition 2

The noiseless channel capacity for a binary IMMM (d', k', e', m') NRZ input-restricted channel with $k' > d' \ge 1$ and $m' > e' \ge 1$, is given by $H = \log_2 \lambda$ where λ is the largest real root of the characteristic equation:

$$D^{e'+k'+m'+1} - D^{e'+k'+m'} - D^{k'+m'+1} + D^{k'+m'} - D^{k'+m'-d'+1} + D^{k'-d'+1} + D^{m'} - 1 = 0.$$

■

The generating function is a rational function that can be expanded into a power series such that the coefficient of each dummy variable equals the number of possible unique constrained sequences where the length of the sequences is given by the power of the dummy variable. We present the generating function for the number of IMMM (d', k', e', m') sequences of length ϑ, for which an arbitrary concatenation of code words also satisfies the channel constraints.

Proposition 3

The generating function for the number of self-concatenable code words for IMMM (d', k', e', m') sequences of length ϑ, where $\vartheta > d'+e'$, is given by:

$$T(x) = \frac{(x^{e'} - x^{m'+e'})(1-x^{i+1})(x^{d'} - x^{k'-i+1})}{[(1-x)(1-x^{e'}) - (x^{d'} - x^{k'+1})(x^{e'} - x^{m'+e'})](1-x)}$$

if $k' - d' \ge \dfrac{m' - e'}{e'}$, and

$$T(x) = \frac{(x^{d'} - x^{k'+1})(1-x^{je'+e'})(x^{e'} - x^{m'-je'+e'})}{[(1-x)(1-x^{e'}) - (x^{d'} - x^{k'+1})(x^{e'} - x^{m'+e'})](1-x^{e'})}$$

if $k' - d' < \dfrac{m' - e'}{e'}$, and

where $i = \left\lfloor \dfrac{k'-d'}{2} \right\rfloor$ and $j = \left\lfloor \dfrac{m'-e'}{2e'} \right\rfloor$. ■

The above proposition can be used to determine the number of code words when developing IMMM block codes.

References

[1] K. A. S. Immink, *Coding Techniques for Digital Recorders*, Prentice Hall International, Englewood Cliffs, NJ, 1991.

[2] R. Karabed and P.H. Siegel, "Matched spectral-null codes for partial-response channels", *IEEE Trans. on Information Theory*, vol. 37, no. 3, pp. 818-855, May 1991.

[3] R. Karabed and P.H. Siegel, "Even mark modulation for optical recording", *Proceedings of the IEEE ICC '89 Conference*, vol. 3, pp. 1628-1632, Boston, MA, June 11-14, 1989.

A Class of Optimal Fixed-Byte Error Protection Codes — (S+Fb)EC Codes —

Tepparit RITTHONGPITAK, Eiji FUJIWARA, and Masato KITAKAMI
Graduate School of Information Science and Engineering, Tokyo Institute of Technology,
2-12-1 Ookayama, Meguro-Ku, Tokyo 152, JAPAN

I. Introduction

Recently, a new class of unequal error protection codes[1] which protects the fixed-byte in computer words from errors has been proposed[2]. Here, the fixed-byte, which stores valuable and important information such as address in communication messages and pointer in database words, means the clustered information having b-bit length whose position in the word is determined in advance.

This paper proposes an extended class of optimal fixed-byte error protection codes which protects the fixed-byte from single-bit errors outside the fixed-byte as well as any errors within the fixed-byte, occurred simultaneously. This class of codes is called *Single-bit plus Fixed b-bit byte Error Correcting codes*, i.e., (S+Fb)EC codes.

II. Preliminaries

Theorem 1 *A binary linear code, described by the parity check matrix $\boldsymbol{H}$, corrects all single-bit plus fixed-byte errors, if and only if*

(a) $e \cdot \boldsymbol{H}^T \neq 0$ *for* $\forall e \in \{E_1 \cup E_2\}$

(b) $e_i \cdot \boldsymbol{H}^T \neq e_j \cdot \boldsymbol{H}^T$ *for* $\forall e_i, \forall e_j \in E_1, e_i \neq e_j$

(c) $e_p \cdot \boldsymbol{H}^T \neq e_q \cdot \boldsymbol{H}^T$ *for* $\forall e_p, \forall e_q \in E_2. e_p \neq e_q$

(d) $e_i \cdot \boldsymbol{H}^T \neq e_p \cdot \boldsymbol{H}^T$ *for* $\forall e_i \in E_1$ *and* $\forall e_p \in E_2$

(e) $(e_i + e_p) \cdot \boldsymbol{H}^T \neq (e_j + e_q) \cdot \boldsymbol{H}^T$ *for* $\forall e_i, \forall e_j \in E_1$ *and* $\forall e_p, \forall e_q \in E_2$, $e_i \neq e_j, e_p \neq e_q$,

where $\boldsymbol{H}^T$ is the transpose of $\boldsymbol{H}$, E_1 is the error set caused by single-bit errors outside the fixed-byte in the word, and E_2 the error set caused by all possible errors in the fixed-byte.

Theorem 2 *The maximal code length of an $(N, N-r)$ (S+Fb)EC code is shown as*

$$N_{max} = 2^{r-b} + b - 1.$$

Thus, the maximum information-bit length K_{max} can be expressed as $K_{max} = 2^x - x - 1$ where $x = r - b$. Table 1 lists K_{max} for check-bit length $r = b + x$.

III. Code Construction

Without loss of generality, the fixed-byte is assumed to be located at the beginning of the word and the check-bits be located at the end of the word. Here, the $\boldsymbol{H}$ matrix of the code is divided into three submatrices shown in (3-1). The submatrix $\boldsymbol{H}_F$ shows the one corresponding to the fixed-byte having b-bit length, the submatrix $\boldsymbol{I}_r$ the one corresponding to the check-bits having r-bit length, and the intermediate submatrix $\boldsymbol{H}_O$ the remaining one having $(N - b - r)$-bit length.

$$\boldsymbol{H} = [\boldsymbol{H}_F | \boldsymbol{H}_O | \boldsymbol{I}_r] \qquad (3\text{-}1)$$

Table 1: Bounds on information-bit length of (S+Fb)EC codes

$r = b + x$	$b+3$	$b+4$	$b+5$	$b+6$	$b+7$	$b+8$
K_{max}	4	11	26	57	120	247

Theorem 3 *The following $\boldsymbol{H}$ matrix shows the (S+Fb)EC code satisfying the bounds on code length shown in Theorem 2:*

$$\boldsymbol{H} = [\, \boldsymbol{H}_F \mid \boldsymbol{H}_O \mid \boldsymbol{I}_r \,] = \left[\begin{array}{c|c|c} \boldsymbol{I}_b & \boldsymbol{O} & \\ \boldsymbol{P} & \boldsymbol{Q} & \boldsymbol{I}_r \end{array}\right],$$

where $\boldsymbol{H}_F = \left[\begin{array}{c} \boldsymbol{I}_b \\ \boldsymbol{P} \end{array}\right], \boldsymbol{H}_O = \left[\begin{array}{c} \boldsymbol{O} \\ \boldsymbol{Q} \end{array}\right]$,

$\boldsymbol{I}_b(\boldsymbol{I}_r)$: $b \times b$ ($r \times r$) identity matrix $\boldsymbol{O}$ *: zero matrix*
$\boldsymbol{P}$: $(r-b) \times b$ matrix whose b distinct binary column vectors have weight larger than or equal to two.
$\boldsymbol{Q}$: matrix having all possible nonzero $(r-b)$-bit binary columns excluding those in $\boldsymbol{P}$ and weight one columns.

Let the upper b bits of the syndrome S be S_F, and the matrix $\boldsymbol{G}_F$ be defined as $\boldsymbol{G}_F = [\, \boldsymbol{P} \mid \boldsymbol{I}_{r-b} \,]$.

With using S_F and $\boldsymbol{G}_F$, decoding of the (S+Fb)EC code is performed according to Table 2.

IV. Conclusion

This paper has proposed an extended class of optimal fixed-byte error protection codes, and has demonstrated the bounds on code length and the code construction method.

References

[1] B.Masnick and J.Wolf, "On Linear Unequal Error Protection Codes", *IEEE Trans. Information Theory*, vol.IT-3, no.4. pp.600 - 607, Oct. 1967.

[2] E.Fujiwara and M.Kitakami, "A Class of Optimal Fixed-Byte Error Protection Codes for Computer Systems", to appear at the symposium FTCS-25, June 27-30, 1995.

Table 2: Decoding of (S+Fb)EC codes

$S = 0$	error free			
$S \neq 0$	S = one column vector in $\boldsymbol{H}$	corresponding single-bit error correction		
	$S \neq$ one column vector in $\boldsymbol{H}$	$S \cdot G_F^T = 0$	fixed-byte error correction [byte error pattern : S_F]	
		$S \cdot G_F^T \neq 0$	$S \cdot G_F^T$: corresponds to one column vector in $\boldsymbol{Q}$ or $\boldsymbol{I}_{r-b}$ in $\boldsymbol{I}_r$	corresponding one-bit error correction and fixed-byte error correction [byte error pattern : S_F]
			$S \cdot G_F^T$: corresponds to one column vector in $\boldsymbol{P}$, e.g., l-th column vector	l-th ($1 \leq l \leq b$) check-bit error correction and fixed-byte error correction [byte error pattern : $S'_F = S_F + (\overbrace{00\cdots0}^{l-1}\,1\,0\cdots0\,\overbrace{)}^{b}$]

A Forbidden Rate Region for Generalized Cross Constellations

E. A. Gelblum & A. R. Calderbank

Mathematical Sciences Research Center
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

Abstract — **An analysis of the Generalized Cross Constellation (GCC) is presented and a new perspective on its coding algorithm is described. We show how the GCC can be used to address generic sets of symbol points in any multidimensional space through an example based on the matched spectral null coding used in magnetic recording devices. We also prove that there is a forbidden rate region of fractional coding rates that are practically unrealizable using the GCC construction. We introduce the idea of a constellation tree and show how its decomposition can be used to design GCC's matching desired parameters. Following this analysis, an algorithm to design the optimal rate GCC from a restriction on the maximum size of its constellation signal set is given, and a formula for determining the size of the GCC achieving a desired coding rate is derived. We finish with an upper bound on the size of the constellation expansion ratio.**

I. Introduction

The $2N-$dimensional generalized cross constellation (GCC) selects a block of N 2-dimensional points from among a family of simply-defined constituent subconstellations by first choosing a constrained sequence of these subconstellations and second selecting an individual channel symbol from each chosen subconstellation. This construction reduces the multidimensional addressing problem to a series of N 2−dimensional subconstellation mappings [1]. Furthermore, since the sequence constraints select the distinct subconstellations with different probabilities, the GCC also makes it possible to reduce average transmitted signal power. While this addressing technique can be applied to channel constellations of any type, generalized cross constellations have hitherto found application in QAM modems.

Since it is possible to encode a fractional average bitrate into each channel symbol and also possible to use a constellation whose total cardinality is not restricted to an integer power of 2, generalized cross constellations are also powerful tools for maximizing the encoding rate of discrete communications channels. These qualities are especially attractive in coding for channels, other than QAM modems, for which there is a predetermined symbol alphabet of size $2^n < N < 2^{n+1}$.

An example of such a discrete communications channel is the Partial Response Maximum Likelihood (PRML) magnetic recording channel. An elementary approach to coding for the PRML channel involves the construction of DC-free block codes which maintain all of the advantages of the MSN trellis codes without the problems of error propagation and decoder complexity [2]. A DC-free block code is a set of balanced binary $N-$tuples, each of which has an equal number of zeros and ones. Therefore, the codewords making up a DC-free code must be selected from among the $\binom{N}{N/2}$ balanced binary $N-$tuples and should be chosen to allow a simple addressing scheme from binary user data. Difficulties occur because $\binom{N}{N/2}$ is never an integral power of 2. Since a simple look-up-table addressing scheme only works for constellations of size 2^k for $k \in \mathbb{Z}^+$, a DC-free block code must discard $\binom{N}{N/2} - 2^{\left\lfloor \log_2 \binom{N}{N/2} \right\rfloor}$ of the $\binom{N}{N/2}$ possible codewords and encode only $\left\lfloor \log_2 \binom{N}{N/2} \right\rfloor$ user bits per channel symbol.

For example, consider the case $N = 10$. Since $\binom{10}{5} = 252$, the optimal addressing method would encode an average of $\log_2(252) = 7.98$ bits per channel symbol. Unfortunately, due to the integral power-of-two restriction, a simple look-up-table addressing scheme only permits a codebook of size $2^{\lfloor \log_2(252) \rfloor} = 128$ codewords and can therefore encode only 7 bits per symbol. Using a GCC, however, a codebook using 240 of the available 252 DC-free binary 10−tuples is possible, and a rate of $7\frac{3}{4}$ bits per channel symbol can be achieved.

II. Theoretical Results

Theorem .1 *Given a generalized cross constellation, C_β, with average encoding rate $\beta = n + \frac{d}{2^m}$ bits per symbol, the total cardinality of C_β is*

$$|C_\beta| = 2^n \cdot \prod_{p' \in P'} R_{p'} \tag{1}$$

where $R_p = \frac{2^p+1}{2^p}$, $P' = \{m-p | p \in P\}$, and the set P is defined as the ordered set of indices, i, in the binary decomposition of d. □

Theorem .2 *There exists a region of values for parameters d, m and n, $d < 2^m$, for which the associated generalized cross constellation requires more than 2^{n+1} channel symbols.* □

References

[1] L. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. on Information Theory*, vol. 33, pp. 483–501, July 1987.

[2] K. Knudsen, J. Wolf, and L. Milstein, "A concatenated decoding scheme for (1-d) partial response with matched spectral-null coding." Globecom, 1993.

Efficient Multiuser Communication in the Presence of Fading

Gregory W. Wornell[1]
Research Laboratory of Electronics, and
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
gww@allegro.mit.edu

Abstract — **New linear symbol-spreading strategies for efficient single- and multi-user communication in environments subject to fading due to time-varying multipath are introduced. For given power, bandwidth, and delay constraints, these new systems significantly reduce the computation required to achieve a prescribed level of performance. Several aspects of these systems and their performance will be developed.**

I. Spread-Response Precoding

For single-user or frequency-division multiplexed wireless systems, we first develop a technique we refer to as "spread-response precoding," which replaces the interleaving typically used in conjunction with coding in such systems. In traditional bandwidth-limited systems for communication over fading channels, coding is used to combat the effects of both additive receiver noise and fading. Furthermore, achieving high performance generally requires the use of codes with a large number of states. However, the computational requirements inherent in the use of such large codes typically preclude their use in practice. With the new systems described in this paper, much of the burden of combatting fading is shifted to the spread-response precoder, allowing shorter codes to be used for a given level of performance. Since this precoding (and postcoding) is implemented using linear filtering, the net result is a significant reduction in computational complexity in the system.

The precoder is implemented using either linear time-invariant or periodically time-varying filters. The key characteristics of the precoding filters is that they are orthonormal or near-orthonormal transformations of the input symbols, and that their impulse response energy is widely spread in time. This spreading allows each coded symbol to see, in an appropriate sense, the average characteristics of the channel. In fact, from the perspective of the coded symbol stream, spread-response precoding asymptotically transforms an arbitrary Rayleigh fading channel into a nonfading, simple white marginally Gaussian noise channel in which intersymbol interference is transformed into a comparatively more benign form of additive white noise that is uncorrelated with the input.

II. Spread-Signature CDMA

In the multiuser case, spread-response precoding generalizes to a new class of orthogonal code-division multiple-access (CDMA) systems for efficient communication in environments subject to multipath fading phenomena. The key characteristic of these new systems, which we refer to as "spread-signature CDMA" systems, is that the associated signature sequences are significantly longer than the interval between symbols. Using this approach, precoding is embedded into the signature sequences in the system, so that the transmission of each symbol of each user is, in effect, spread over a wide temporal and spectral extent, which is efficiently exploited to combat the effects of fading.

Analogous to the single-user case, spread-signature CDMA systems asymptotically transform the multiuser Rayleigh fading channel into a collection of decoupled quasi-Gaussian channels. Optimizing the signal-to-noise ratio in the resulting quasi-Gaussian channel with respect to the choice of a linear equalizer leads to mininimum mean-square error type equalizers.

An optimum class of spread-signature sets for this application is developed out of multirate system theory, and efficient implementations are described. Estimates of the capacity and uncoded bit-error rate characteristics are computed with these optimized systems and compared with those of more traditional CDMA systems. The performance advantages appear substantial for practical systems. Furthermore, the use of these new systems requires no additional power or bandwidth, and is attractive in terms of computational complexity, robustness, and delay considerations. Some remaining challenges inherent in their use—including managing peak-to-average power requirements and developing suitable timing recovery strategies—are also described.

A detailed development of these results is presented in [1] [2].

Acknowledgements

The author is grateful to N. S. Jayant, C-E. Sundberg, N. Seshadri, J. Kovacevic, M. Sondhi, A. Odlyzko, E. Teletar, A. Wyner, and S. Shamai (Shitz), all at AT&T Bell Laboratories, for many helpful discussions, comments and suggestions regarding this work.

References

[1] G. W. Wornell, "Spread-Response Precoding for Communication over Fading Channels," submitted to *IEEE Trans. Inform. Theory*, Apr. 1994.

[2] G. W. Wornell, "Spread-Signature CDMA: Efficient Multiuser Communication in the Presence of Fading," to appear in *IEEE Trans. Inform. Theory*.

[1]This work has been supported by AT&T Bell Laboratories, where the author was on leave during the 1992-93 academic year, and in part by the Advanced Research Projects Agency monitored by ONR under Contract No. N00014-93-1-0686, the National Science Foundation under Grant No. MIP-9502885, and the Office of Naval Research under Grant No. N0014-95-1-0834.

Information Theoretic Limits on Communication Over Multipath Fading Channels

Richard Buz[1]

Communications Research Centre, 3701 Carling Ave., P.O. Box 11490, Station H
Ottawa, Ontario, Canada, K2H 8S2

Abstract — **Limits on the rate of reliable communication over multipath fading channels are presented. An idealized channel model is considered first in order to determine the loss due to amplitude fading. The requirement of channel estimation is demonstrated through calculation of limits for channels in which the state of the fading process is not completely known. Loss incurred due to the limitation of practical channel estimation schemes is determined. The particular methods of channel estimation considered are pilot tone extraction, differentially coherent detection, and the use of a pilot symbol.**

I. Ideal Fading Channels

The capacity of a discrete-time Rayleigh fading channel has been considered by Ericson [1]. His result is based on the idealistic assumption that the value of the fading process is known at the receiver and is independent with respect to discrete-time symbol intervals. For an ideal Nakagami fading channel and integer values of the Nakagami parameter m, the capacity is

$$C = (\log_2 e)\frac{(-m)^m}{\Gamma(m)}\left[\frac{\bar{E}_s}{N_0}\right]^{-m}\left(\frac{d}{ds}\right)^{m-1}\left[\frac{e^s}{s}Ei(-s)\right] \quad \text{bits}/T$$

where $s = m\left(\frac{\bar{E}_s}{N_0}\right)^{-1}$, $\Gamma(\cdot)$ is the gamma function, $E_i(\cdot)$ is the exponential integral function, and T is the discrete-time signaling interval. When compared to the capacity of an additive white Gaussian noise channel, the maximum loss in average SNR due to Nakagami fading is $me^{-\psi(m)}$ where $\psi(\cdot)$ is Euler's psi function. This expression of loss is valid for any $m > 0$. The capacity of a Nakagami fading channel also represents the capacity of a Rayleigh fading channel when space diversity combining is used. In this case, the Nakagami channel parameter m corresponds to the number of antennae used in the system. In terms of channel capacity, the gain in SNR achievable through the use of antenna diversity is $e^{\mathbf{C}_E} - me^{-\psi(m)}$ where $\mathbf{C}_E$ is Euler's constant.

II. Incomplete CSI

In a Rician fading environment with no CSI, a line-of-sight (LOS) component exists which is normally strong enough to support the transmission of information. In this case, the scattered component is sometimes viewed as an additional source of interference, although it does convey a small amount of information. By using entropy power relations [2], one may determine an upper bound to the average mutual information (AMI) of the form

$$I_U = \log\left[\frac{1+\frac{\bar{E}_s}{N_0}}{1+\frac{\exp(-\Delta_J)}{1+\gamma_R}\frac{\bar{E}_s}{N_0}}\right]$$

where the Rician channel parameter γ_R is the ratio of power in the LOS component to that in the scattered component, and $\Delta_J = \ln E_s - E_{p(x)}\left\{\ln|x|^2\right\}$ is a positive number obtained from Jensen's inequality. As SNR $\to \infty$, I_U approaches a constant value of $\Delta_J \log e + \log(1+\gamma_R)$. For a Rayleigh channel and a Gaussian distributed input, the AMI is bounded to less than 0.83 bits/T.

If a receiver can track variations in the phase of the fading process, then it is reasonable to model the system as having ideal fading phase information but no fading amplitude information. In this case, entropy power relations yield an upper bound on AMI for a Rayleigh channel of the form

$$I_U = \log\left[\frac{1+\frac{\bar{E}_s}{N_0}}{1+\frac{\exp(-\Delta_J)}{2\pi}\frac{\bar{E}_s}{N_0}}\right]$$

which approaches a value of $\Delta_J \log e + \log 2\pi$ as SNR $\to \infty$. With ideal fading phase information and sufficient SNR, data transmitted via the symbol phase can be accomplished with an arbitrarily small probability of error. In addition, a discrete-valued constellation can be used to achieve a higher data rate than a continuous-valued input.

III. Use of Channel Estimation

When CSI is obtained by means of practical estimation methods, the AMI conditioned on knowledge of the channel estimate is a function of both SNR and the Doppler frequency f_D of the fading process. When considering practical signal constellations for coding (i.e. at rates of less than $\log_2 M$ bits/T with an M-point constellation), additional losses are incurred due to the limitations of these estimation schemes. For a Rayleigh fading channel with a normalized Doppler frequency of $f_D T = 0.1$, systems which use pilot tone estimation incur a loss of 1.0-1.5 dB. Under the same conditions, the loss experienced through the use of differentially coherent detection is roughly in the range of 3-4 dB. Systems based on pilot symbol transmission exhibit losses in the range of 4.5-8.5 dB. When using differential detection or pilot symbol transmission, the equivocation of the channel cannot be made arbitrarily small. The magnitude of this remaining uncertainty is strongly affected by the value of $f_D T$.

Acknowledgements

The author would like to thank Dr. Peter McLane for many helpful discussions related to this work.

References

[1] T. Ericson, "A Gaussian Channel with Slow Fading," *IEEE Trans. Inform. Theory*, vol. 16, pp. 353–355, May 1970.

[2] N. M. Blachman, "The Convolution Inequality for Entropy Powers," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 267–271, Apr. 1965.

[1]This work was performed as part of a Ph.D. thesis at Queen's University with support provided by TRIO and NSERC.

Computational Cutoff Rate of BDPSK Signaling Over Correlated Rayleigh Fading Channels

*Wayne C. Dam, †Desmond P. Taylor, *Zhi-Quan Luo

*Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada

†Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand

Abstract — **We derive a compact formulation of the computational cutoff rate of binary differential phase shift keying (BDPSK) over a correlated Rayleigh fading channel. The analysis is more realistic than previous finite state Markov models of a fading channel.**

I. Introduction

Compared to memoryless channel models, there are few capacity and cutoff rate results for channels with memory, and most such models are not very realistic [1]. We present an exact analysis of non-interleaved binary differential phase shift keyed signaling over a correlated Rayleigh fading channel. The lack of interleaving forces the analysis to deal with the channel memory directly. We find that modeling the received channel process as a finite order Markov process allows the sum-over-codewords portion of the computational cutoff rate calculation to be performed combinatorially. This provides for a succinct formulation of the cutoff rate, R_0.

II. Results

On a Rayleigh fading channel, the probability density function of received signal, $\mathbf{y}$, conditioned on the N symbol transmitted vector $\mathbf{x}$, can be written

$$p_N(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^{Nn}\,|R|} e^{-\mathbf{y}^\dagger X R^{-1} X^\dagger \mathbf{y}},$$

where we assume n samples per channel symbol. The total channel correlation matrix, $R \equiv R_f + \sigma^2 I$, is the sum of the fading correlation matrix, R_f, and that of the additive white noise of variance σ^2. The diagonal matrix X takes the vector $\mathbf{x}$ along its diagonal, i.e., $X = \mathrm{diag}(\mathbf{x})$.

The code ensemble average probability of error can be bounded by the combined Union-Bhattacharyya bound [2]. Using the above density, and simplifying for the case of BDPSK signaling, we have the code ensemble bound

$$\overline{P_e} \le \frac{M-1}{2^N} \sum_X \frac{2^{Nn}}{|R^{-1} + X R^{-1} X^\dagger|},$$

where M is the number of codewords in the code and we sum over all binary sequences X. If we assume that the received channel process is auto-regressive of order L, then the inverse channel matrix, R^{-1}, will be Toeplitz banded diagonal in form, except for $L \times L$ sample blocks at each end of the diagonal [3].

The trick here is recognizing that off-diagonal entries of the denominator's matrix will be either zero or a constant non-zero value, depending on whether the phase shift between symbols in X is zero or π phase shift respectively. This, provided we are sampling at at least L samples per symbol. The zero off-diagonal entries will then pinch off the matrix into a block diagonal form. For example,

$$R^{-1} + X R^{-1} X^\dagger \sim \begin{bmatrix} \square\square & & & & \\ \square\square & & & & \\ & \square & & & \\ & & \square\square & & \\ & & \square\square & & \\ & & & \square & \\ & & & & \square\square \\ & & & & \square\square \end{bmatrix}.$$

Summing over all binary sequences of X in the bound is then equivalent to summing over all possible block partitions of the matrix. This then corresponds to summing over all integer partitions of N. If we define $D(m)$ as the determinant of the $m \times m$ symbol band Toeplitz block of R^{-1}, the combinatorics of the partitioning allows us to write

$$\overline{\mathrm{P_e}} \le \frac{2}{2^N} \frac{1}{\mathrm{N!}} \sum_{\mathrm{k=1}}^{\mathrm{N}} \mathrm{k!}\, \mathrm{B_{N,k}}\left(\frac{1!}{\mathrm{D(1)}}, \frac{2!}{\mathrm{D(2)}}, \ldots, \frac{(\mathrm{N-k+1})!}{\mathrm{D(N-k+1)}}\right),$$

where $B_{N,k}(\cdot)$ is the (N,k) Bell polynomial [4]. We define the $m \times m$ symbol matrix $\mathcal{R}(m)$ such that its inverse, $\mathcal{R}^{-1}(m)$, equals the m symbol inverse channel correlation matrix, R^{-1}, however, we extend the inner Toeplitz portion of the matrix into the $L \times L$ sample blocks at either end of the diagonal, overwriting them. Thus, $|\mathcal{R}(m)| = 1/D(m)$.

Using a relation between Bell polynomials and the composition of Taylor series [4], and between the exponential growth of the coefficients of a Taylor's series and its radius of convergence [5], we can then formulate the computational cutoff rate as follows.

Theorem 1 Define the generating function for the determinants of the m symbol Toeplitz extended channel correlation matrices as

$$\begin{aligned} g(t) &= \sum_{m=1}^{\infty} |\mathcal{R}(m)|\, t^m \\ &= |\mathcal{R}(1)|t + |\mathcal{R}(2)|t^2 + |\mathcal{R}(3)|t^3 + \cdots \end{aligned}$$

The computational cutoff rate is then given by

$$R_0 = \log_2(2|t_0|) \quad [\text{bits/symbol}],$$

where t_0 is the smallest magnitude singularity of the function

$$h(t) = \frac{1}{1-g(t)}.$$

References

[1] Robert G. Gallager, *Information Theory and Reliable Communications*, Wiley, New York, 1968.

[2] Andrew J. Viterbi and Jim K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, New York, 1979.

[3] Toby Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1971.

[4] John Riordan, *An Introduction to Combinatorial Analysis*, Wiley, New York, 1958.

[5] Herbert S. Wilf, *Generatingfunctionology*, Academic Press, Boston, 1994.

On the construction of MPSK block codes for fading channels

Jaime Portugheis and Christian D. de Alencar

Decom - Fee - Unicamp, C.P. 6101, 13081-970 Campinas, SP, Brasil; e-mail: jaime@decom.fee.unicamp.br

This paper focuses on the construction of M-PSK block modulation codes for the Rayleigh fading channel. We present some new codes constructed by two different methods. The first method, an exhaustive computer search, is appropriate for short block lengths. Some optimum codes are found. The second method is for multilevel block codes. In this case we use the cutoff rate performance criterion for multistage decoding. Simulation results are presented. From them we conclude that the second method can propose codes which achieve improvement over known codes for low and moderate SNR's.

Codes found by a computer search

The block codes, so far supplied by the literature [1, 2], were constructed using the multilevel coding technique. Good codes can be constructed by the multilevel technique (with the intrinsic advantage of a multistage decoder), but this technique does not always lead to optimum codes. If the block length of the code n is small and $M = 4, 8, 16$, it is possible to generate all M^n sequences of M-PSK symbols and store the subset of sequences with the greatest number of elements that satisfies a specified design criterion.

Our aim was to construct codes based on the following design criterion: "For a given code length n and a given value of the desired minimum Hamming distance d_H, find the code with the greatest rate R (bits/symbol) such that the minimum product distance d_P is not less than γ."

By a computer search, some new codes with lengths $n = 4, 5, 6, 7$ (number of M-PSK symbols) and different minimum Hamming distances were found for 4-PSK and 8-PSK modulation. We will show simulation results for a 4-PSK code with $n = 6, R = 1$ and $d_H = 4$ that cannot be constructed as a multilevel code. This code has a coding gain of about 14.0 dB over the uncoded 2-PSK system, at the bit error probability (P_b) 10^{-3}.

Multilevel block codes

We consider multilevel block codes constructed based on a sequence of binary partitions of the 2^m−PSK modulation. The m-level code consists of the binary component codes $B_0, B_1, \ldots, B_{m-1}$ with rates $R(0), R(1), \ldots, R(m-1)$, respectively. The method for constructing multilevel codes we propose deals with the following question: For a given rate $R = R(0)+R(1)+\cdots+R(m-1)$ of the m-level code and a given SNR of the channel, how can the rates $R(j), 0 \leq j \leq m-1$, be chosen in such a way that the word error probability (P_E) for multistage decoding of the m-level code is minimized? This question is answered in [3] based on the cutoff rate performance criterion. This criterion leads to the rates $R(j), 0 \leq j \leq m-1$, that minimize an upper bound on P_E of multistage decoding.

Assuming a Rayleigh fading channel with ideal interleaving, ideal coherent detection and perfect channel state information, we have obtained the optimum rates for the component codes of 4-PSK, 8-PSK and 16-PSK block modulation codes for different values of SNR's.

Knowing the optimum rates for a given SNR we can construct multilevel block codes with the help of Verhoeff's table. For example, the optimum rates $R_{op}(j)$, $j = 0, 1, 2$, for a fixed rate $R = 2.0$ bits/symbol of the 8-PSK block code and a SNR of 15.0 dB are: $R_{op}(0) = 0.4362$, $R_{op}(1) = 0.7349$ and $R_{op}(2) = 0.8289$. If we choose $n = 16$, we can approximate the optimum rates with the codes: $B_0 = (16, 7, 6)$, $B_1 = (16, 12, 2)$ and $B_2 = (16, 13, 2)$. If we use $B_0 = (16, 7, 6)$, $B_1 = (16, 11, 4)$ and $B_2 = (16, 15, 2)$, we get a code with $R = 2.06$.

Figure 1 shows the behaviour of P_b for two different 8-PSK block codes. Code X is the above mentioned code with $R = 2.06$. Code Y is a code of same rate R and minimun Hamming distance 4 constructed with component codes $B_0 = B_1 = B_2 = (16, 11, 4)$. Code X shows better performance than code Y for low and moderate SNR's. For $P_b = 10^{-3}$, code X has a coding gain of about 1.0 dB over code Y. For SNR's higher than 17.0 dB, the performance of code Y is superior. This behaviour can be explained due to the fact that it has a larger value of d_H, which is the dominanting performance parameter for high SNR's.

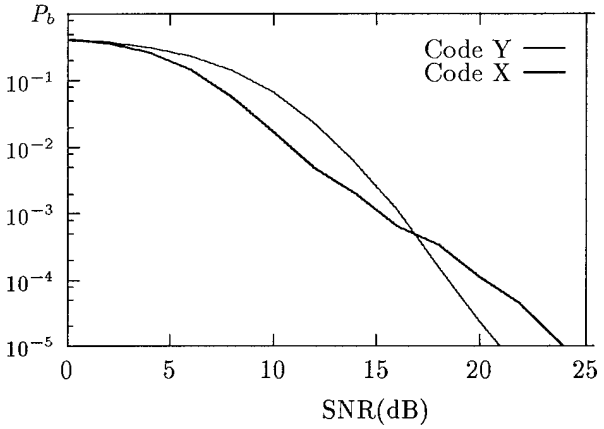


Fig. 1: Performance of 8-PSK block codes X and Y

This work has been partially supported by CNPq under Grant 301045/92-5 and CPqD-Telebras under contract 387/90.

References

[1] L. Zhang, B. Vucetic, *"Bandwidth Efficient Block Codes for Rayleigh Fading Channels"*, Electronic Letters, vol. 26, no. 5, March 1990.

[2] S. Lin, S. Rajpal, D. J. Rhee, *"MPSK Coded Modulations for the Rayleigh Fading Channel"*, Proceedings 1994 IEEE Int. Symp. on Inform. Theory, pp. 451, Norway, June 27-July 1, 1994.

[3] J. Portugheis, *"Cutoff Rate Performance Criterion for Multistage Decoding"*, Proceedings Sixth Joint Swedish-Russian Int. Workshop on Inform. Theory, pp. 56-60, Sweden, August 22-27, 1993.

Multilevel Block Coded 8-PSK Modulation Using Unequal Error Protection Codes for Rayleigh Fading Channels

R.H. Morelos-Zaragoza[♯], T. Kasami[†] and S. Lin[‡]

[♯] 3rd Department, Inst. of Industrial Science, University of Tokyo, 7-22-1 Roppongi, Minatoku, Tokyo 106 Japan

[†] Graduate School of Info. Science, Nara Inst. of Science and Tech., Ikoma, Nara 630-01 Japan

[‡] Dept. of Electrical Engineering, Univ. of Hawaii, Honolulu, Hawaii 96822 U.S.A.

Abstract — **In this paper, new block coded 8-PSK modulations with unequal error protection (UEP) capabilities for Rayleigh fading channels are presented. The proposed codes are based on the multilevel construction of Imai and Hirakawa [1]. It is shown that the use of linear UEP (LUEP) codes [2] as component codes in one or more of the encoding levels provides increased error performance with respect to conventional multilevel codes.**

I. Summary

Previous work on combining LUEP codes and PSK modulation for fading channels is reported in references [3] and [4]. Hagenauer et al. [3] proposed rate-compatible punctured convolutional codes combined with DQPSK modulation to provide UEP by means of their variable rate structure. Reference [4] used Gray labeling of a QPSK signal set to map LUEP codes of even length onto block modulation codes with UEP capabilities. Seshadri and Sundberg [5] studied the UEP capabilities of multilevel codes of length 8 over Rayleigh fading channels. The aim of this research work is to design efficient block coded modulations (BCM) over 8-PSK signal sets *for the specific purpose of UEP* over Rayleigh fading channels. Over a fading channel, the minimum *symbol and product distances* are the parameters that dominate the overall error performance. The symbol distance is closely related to the *Hamming distance* of the component codes. Thus it is natural to consider binary LUEP codes as component codes in the multilevel construction to obtain good BCM for UEP over fading channels.

Let S represent a unit-energy 8-PSK signal set. A label $\ell_k = b_1 + 2b_2 + 4b_3$ represents the signal point $e^{j\ell_k\pi/4}$, for $0 \leq k < 8$, where $j = \sqrt{-1}$, and $b_i \in \{0,1\}$, $1 \leq i \leq 3$. In multilevel block coded modulation [1], codewords of three linear binary block codes of length n, dimension k_i and minimum distance d_i, denoted C_i, are used to select label bits b_i, for $1 \leq i \leq 3$. The set of resulting sequences of n 8-PSK signals is said to be a block modulation code Λ of length n and rate $R = (k_1 + k_2 + k_3)/n$ bits/symbol.

A two-level (n,k) LUEP code is a linear code that it is not spanned by its set of minimum weight vectors. We use UEP(n,k) to denote such a code and refer to its unequal error protection capabilities as follows: *separation vector* $\bar{\mathbf{s}} = (s_1, s_2)$ for the *message space* $\{0,1\}^{k^{(1)}} \times \{0,1\}^{k^{(2)}}$, where $k = k^{(1)} + k^{(2)}$. This means that codewords in correspondence to $k^{(i)}$ information bits are at a Hamming distance at least s_i, $i = 1,2$. Without loss of generality, it is assumed that $s_1 \geq s_2$. Thus an information vector of length k bits can be separated into a most significant part of length $k^{(1)}$ bits (the MSB) and a least significant part of length $k^{(2)}$ bits (the LSB). The proposed multilevel construction uses an (n, k_2, d_2) linear code, or a UEP(n,k_2) code, C_2 in the second encoding level and a UEP(n,k_3) code C_3 in the third encoding level.

As an example, let C_1, C_2 and C_3 be $(8,4,4)$, $(8,7,2)$ and $(8,7,2)$ linear codes, respectively. The Imai-Hirakawa multilevel construction results in a block modulation code $\mathbf{\Lambda_1}$ of length 8, rate $R = 2.25$ bits/symbol, minimum symbol distance $\delta_H = 2$ and minimum product distance $\Delta_p^2 = 4$. [5]. By letting C_3 be a binary optimal LUEP code, UEP$(8,5)$, from [6] with separation vector $\bar{\mathbf{s}} = (3,2)$ for the message space $\{0,1\}^4 \times \{0,1\}$, a block modulation code $\mathbf{\Lambda_2}$ is obtained. $\mathbf{\Lambda_2}$ has length 8, rate $R = 2$ bits/symbol, $\delta_H = 2$ and $\Delta_p^2 = 4$. In addition, 25% of the information bits (the 4 MSB encoded by UEP$(8,5)$) have corresponding symbol and product distances equal to 3 and 64, respectively. That is, a subset of the coded sequences, those corresponding to the MSB encoded by the LUEP code, has increased symbol and product distances. It follows that, with *no bandwidth expansion* over uncoded QPSK, *higher error performance* is achieved. In the presentation, simulation results will be presented showing an increase in both overall coding gain and that for the most important message part. At a bit error rate (BER) of 10^{-3}, the coding gain in the third level is at least 13 dB for $\mathbf{\Lambda_2}$, compared to about 8.5 dB for $\mathbf{\Lambda_1}$. In addition, at a BER of 10^{-3}, an advantage of 2 dB in overall coding gain is achieved.

[1] This work was supported in part by NASA under grant NAG 5-931, by the NSF under grants NCR-88813480 and NCR-9115400, and by the Japanese Society for the Promotion of Science (JSPS) under fellowship no. 93157.

References

[1] H. Imai and S. Hirakawa, "A New Multilevel Coding Method Using Error Correcting Codes," *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, pp. 371-376, May 1977.

[2] B. Masnick and J. Wolf, "On Linear Unequal Error Protection Codes," *IEEE Trans. Info. Theory*, vol. IT-13, no. 4, pp. 600-607, Oct. 1967.

[3] J. Hagenauer, N. Seshadri and C.-E. W. Sundberg, "The Performance of Rate-Compatible Punctured Convolutional Codes for Digital Mobile Radio," *IEEE Trans. Communications*, vol. 38, no. 7, pp. 966-980, July 1990.

[4] R.H. Morelos-Zaragoza and S. Lin, "Block QPSK Modulation Codes With Two Levels of Error Protection," *Proceedings of the Fifth IEEE International Symposium on Personal, Indoor and Mobile Communications (PIMRC'94)*, vol. II, pp. 548-552, The Hague, The Netherlands, Sept. 1994.

[5] N. Seshadri and C.-E. W. Sundberg, "Coded Modulation with Time Diversity, Unequal Error Protection, and Low Delay for the Rayleigh Fading Channel," *1st. Conference on Universal Personal Communications (ICUPC '92)*, Conf. Rec., pp. 283-287, Dallas, Texas, Sept. 1992.

[6] W.J. Van Gils, "Two Topics on Linear Unequal Error Protection Codes: Bounds on Their Length and Cyclic Code Classes," *IEEE Trans. Info. Theory*, vol. IT-29, no. 6, pp. 866-876, Nov. 1983.

A Change-Detection Approach to Monitoring Fading Channel Bandwidth

Steven D. Blostein[1] and Yong Liu
Dept. Elect. & Comp. Eng. Queen's University,
Kingston, Ontario, Canada K7L 3N6 Email: sdb@ee.queensu.ca

Abstract — **The statistics of mobile communications over frequency nonselective fading channels are determined largely by fading bandwidth, which is related to vehicle speed. On-line estimation of fading bandwidth can be used to optimize coherent signal transmission, as well as improve handoff algorithms. In the following, level crossing rates of received signal amplitude, combined with recent on-line change detection techniques [1] are used to estimate fading bandwidth. The proposed estimator takes AGWN into account and has low complexity and processing delay. Applied to adaptive on-line tracking of fast fading channel parameters as performed in [2], significant BER reduction is demonstrated, particularly in situations where vehicle speed increases abruptly.**

I. The Model

Signal x_k is transmitted over a frequency-nonselective Rician fading channel. The received low-pass equivalent discrete-time signal is $y_k = x_k c_k + n_k$ where c_k is the channel gain. Let mean $a = E\{c_k\}$ and covariance function

$$r_n = r_0 J_0(2\pi f_m nT) = r_0 \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j2\pi f_m nT\sin\theta} d\theta \qquad (1)$$

where $J_0()$ is the 0th order Bessel function, T is the symbol period and $f_m = \frac{v}{\lambda}$ is the maximum Doppler frequency (fading channel bandwidth) with v and λ defined as mobile vehicle speed and transmission wavelength, respectively.

II. Monitoring Vehicle Speed by Measuring Level Crossing Rate

In [3], it is shown that the number of crossings at voltage level, A, is

$$\bar{n}(|c_k - a| = A) = \bar{n}(R) = \sqrt{2\pi} f_m R e^{-R^2} \qquad (2)$$

and the average fade duration is

$$\bar{t}(|c_k - a| = A) = \bar{t}(R) = \frac{1}{\sqrt{2\pi} f_m}\frac{1}{R}(e^{R^2} - 1) \qquad (3)$$

with $R = \frac{A}{\sqrt{r_0}}$, and $r_0 = Var(c_k) = E(c_k^2) - |a|^2$. Therefore measuring the level crossing rate yields an estimate of v and f_m. By choosing $R = -5$ dB, i.e., $A = 0.5623\sqrt{r_0}$ as in [3],

$$v \approx \frac{1.6579\lambda}{2\pi\bar{t}} = 6.1154\frac{\bar{n}\lambda}{2\pi}. \qquad (4)$$

[1]This research was supported by NSERC Research Grant OGP0041731

III. Measuring Level Crossing Rate Using Change Detectors

Since y_k contains noise, rather than count level crossings directly, we view the problem as a sequential change detector as follows: Letting z_i denote the power of y_i, it can be shown that for reasonably large SNR

$$z_i \approx |c_i|^2 + c_i^* x_i^* n_i + c_i x_i n_i^* \qquad (5)$$

Since c_i has small bandwidth relative to n_i, c_i can be treated as locally deterministic. Conditioned on x_i, it can be shown that z_i is Gaussian with mean $E\{z_i\} = |c_i|^2$ and variance

$$Var\{z_i\} = 2|c_i|^2\sigma_n^2 \qquad (6)$$

where σ_n^2 is the variance of n_i. The channel energy $|c_i|^2$ equals $A_0^2 = r_0 + |a|^2$ and nominally and drops to below $A_1^2 = 0.5623^2 r_0 + |a|^2$ during a fade. ¿From the above, the problem can be transformed into one of quickest detection of a change from

$$H_0: \quad z_i \sim f_0(z_i) = \frac{1}{\sqrt{4\pi A_0^2\sigma_n^2}} e^{-\frac{(z_i - A_0^2)^2}{4A_0^2\sigma_n^2}}$$

$$H_1: \quad z_i \sim f_1(z_i) f_1(z_i) = \frac{1}{\sqrt{4\pi A_1^2\sigma_n^2}} e^{-\frac{(z_i - A_1^2)^2}{4A_1^2\sigma_n^2}}$$

and vice-versa. We have investigated a two-sided Page's cumulative-sum (CUSUM) statistic, as well as an alternative change-detection procedure [1] that is well-suited to cases where the change-time is known to be finite. The resulting fading bandwidth estiamtor has been applied to the adaptive fading channel tracker, DFALP, described in [2] as follows: the optimal DFALP linear prediction and LPF filter parameters are first recorded off-line for a set of fading bandwidths in constant speed conditions. In on-line use, the DFALP parameters are then adjusted adaptively in steps of 20 km/hour. The BER performance of differential quadrature phase-shift keying DQPSK detection is used as a reference. From simulations, it is shown that when the vehicle speed increases from 60 to 100 km/hour (in Rician fading with $a^2/r_0 = 4dB$), a 5 dB gain is observed over both DQPSK and DFALP (without parameter adaptation) at BER 6.30×10^{-3} and SNR $= 20\ dB$. Noticeable gains are also observed if the SNR is greater than 12 dB.

References

[1] Liu, Y. and S.D. Blostein, "Quickest detection of an abrupt change in a random sequence with finite change-time." *IEEE Transactions on Information Theory*, November, pp. 1985-1993, 1994.

[2] Liu, Y. and S.D. Blostein, "Identification of Frequency Nonselective Fading Channels Using Decision Feedback and Adaptive Linear Prediction," *IEEE Transactions on Communications.*, vol. 43, no. 4, pp 1484-1492, 1995.

[3] Lee, W.Y.C., *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.

On the Correlation and Scattering Functions of Mobile Uncorrelated Scattering Channels

J. S. Sadowsky and V. G. Kafedziski
Dept. of Electrical Engineering, Arizona State University
Tempe, AZ USA 85287-7206
e-mail: sadowsky@asu.edu

Abstract — **In this paper we consider impulse response statistics of the wide sense stationary – uncorrelated scattering (WSSUS) multi-path channel that results from a stationary scattering field with either the transmitter or receiver in motion, but not both.**

Summary

Let $h(\tau;t)$ denote the time varying impulse response of the channel. By *delay uncorrelated scattering* we mean

$$\mathrm{E}[h(\tau_a;t)^* h(\tau_b;t+\Delta t)] \quad = \quad 2\,\phi_h(\tau_a;\Delta t)\,\delta(\tau_b-\tau_a).$$

In this paper we derive general formulas for the *correlation function* $\phi_h(\tau;\Delta t)$ and the *scattering function* $S(\tau;\lambda) = \int \phi_h(\tau;\Delta t) e^{-j2\pi\lambda\Delta t}\, d\Delta t$ where λ is the Doppler frequency shift variable.

This general family of channels has been the subject of a great deal of research. The commonly cited result is the correlation function $\phi_h(\tau;\Delta t) \propto J_0(2\pi\lambda_m\Delta t)$ where λ_m is the maximal doppler frequency, due to Jakes [2]. Here we show that, for arbitrary scattering fields, Jakes' result is actually just the 0th term in the series

$$\phi_h(\tau;\Delta t) \quad = \quad 2\pi \sum_{n=-\infty}^{\infty} \psi_n(\tau) e^{jn(\theta_0+\pi/2)} J_n(2\pi\lambda_m\Delta t)$$

where θ_0 is the velocity vector angle relative to the base-to-mobile baseline. The series coefficients $\psi_n(\tau)$ are determined from the spatial distribution of scatterers and propagation path loss factors.

For the case of a spatially uniform scattering field with $1/r^2$ propagation loss factors, we obtain

$$\psi_n(\tau) \quad = \quad \frac{2\,c\,\phi_\beta}{c\tau\,((c\tau)^2 + r_0^2)\,\sqrt{1-a(\tau)^2}} \times \left(\frac{a(\tau)}{1+\sqrt{1-a(\tau)^2}}\right)^{|n|}$$

where ϕ_β is the spatial scattering intensity, r_0 is the base-to-mobile baseline length, c is the speed of propagation, and $a(\tau) = 2c\tau r_0/[(c\tau)^2 + r_0^2]$. As opposed to uniformly distributed scatterers, Jakes' derivation emphasizes scattering near the mobile unit.

In addition to the classical 2-D mobile problem discussed above, we also derive results for some 3-D problems.

Our analysis is based on the theory of *generalized stochastic processes* [1]. A generalized process is a continuous random linear functional on a topological vector space of test functions. This theory is a direct extension of the well known theory of generalized functions (also known as "distribution theory"). We assume a spatially uncorrelated scattering field with spatial scattering intensity function. The field may be either diffuse (white Gaussian) or specular (white Poisson). Our results are obtained by application of an elliptical change of variables to the appropriate spatial test function integral.

The channel impulse response $h(\tau;t)$ is a time-varying *linear* transformation of the spatial scattering field. Thus, *the channel's 2nd order moments are completely determined by the spatial scattering field's 2nd order moments and the propagation geometry.* The common method of analysis first considers discrete elemental scatterers and then passes to a limit of increasingly dense but vanishingly small scatterers. This limiting method is quite cumbersome, it is difficult to identify the corresponding spatial scattering intensity, and, in our analysis, the limiting method is entirely unnecessary.

References

[1] L. Arnold, *Stochastic Differential Equations: Theory and Application.* New York: Wiley, 1974.

[2] W. C. Jakes, *Microwave Mobile Communications.* New York: Wiley, 1974.

High Diversity Lattices for Fading Channels

Joseph Boutros, Emanuele Viterbo
Ecole Nationale Supérieure des Télécommunications, Paris, France
Politecnico di Torino, Torino, Italy

Abstract — **We show that particular versions of the densest lattice packings present very good performance over the Rayleigh fading channel. These versions not only have high diversity, but also may be decoded efficiently since they are binary lattices.**

I. Introduction

The practical interest in lattice constellations presenting good performance over fading channels rises from the need to transmit information at high bit rates over terrestrial radiomobile links. Constellations matched to the fading channel are effective because of their high degree of *diversity*. By diversity we intend the number of different component values of any two distinct points in the constellation. The signal constellations for Gaussian channels are usually very bad when used over Rayleigh fading channels since they have small diversity. We constructed signal constellations with high spectral efficiency matched to the Rayleigh fading channel using algebraic number theory [1]. The signal constellations are derived from the densest lattices (D_4, E_6, E_8, K_{12}, Λ_{16}, Λ_{24}) and their diversity order is half the lattice dimension.

II. System model

Consider the following model. A mapper associates an m-uple of input bits to a signal point $\mathbf{x} = (x_1, x_2, \ldots x_n)$ in the n-dimensional Euclidean space $\mathbf{R}^n$. Let $M = 2^m$ be the total number of points in the constellation. The points are transmitted over a Rayleigh channel giving $\mathbf{r} = \alpha * \mathbf{x} + \mathbf{n}$, where $\mathbf{r}$ is the received point. $\mathbf{n} = (n_1, n_2, \ldots n_n)$ is a noise vector, whose real components n_i are zero mean, N_0 variance Gaussian distributed independent random variables. $\alpha = (\alpha_1, \alpha_2, \ldots \alpha_n)$ are the independent random fading coefficients with unit second moment and $*$ represents the componentwise product.

The signal points $\mathbf{x}$ are chosen from a constellation which is carved from a lattice Λ. The spectral efficiency is measured in number of bits per two dimensions $s = 2m/n$, and the signal-to-noise ratio per bit is given by $SNR = E_b/N_0$, where E_b is the narrow band average energy per bit and $N_0/2$ is the narrow band noise power spectral density.

III. New constellations

An accurate analysis of the symbol error probability shows that the most important feature of a good constellation for the fading channel is its diversity L. The following theorem enables us to evaluate the diversity L of any lattice constructed from an algebraic number field.

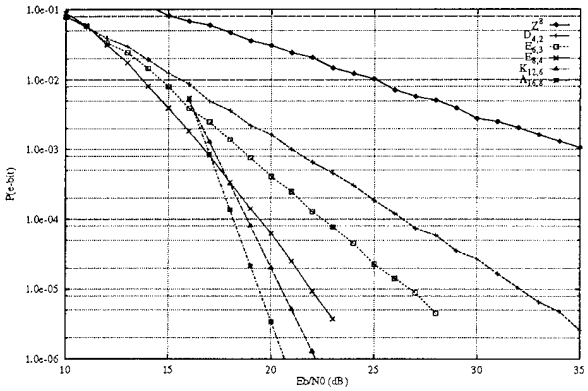
Theorem. *The lattices obtained from the canonical embedding of an algebraic number field with signature (r_1, r_2) exhibit a diversity $L = r_1 + r_2$.*

Since totally complex cyclotomic fields have a signature $(0, n/2)$ the diversity of the corresponding lattices is $L = n/2$. We use Craig's work [2, 3], who showed how to construct the lattices E_6, E_8, Λ_{24} (Leech lattice) from the totally complex cyclotomic fields $K = \mathbf{Q}(e^{i2\pi/N})$ for $N = 9, 20, 39$. We applied the same procedure and we found D_4 (Schlafli lattice), K_{12} (Coxeter-Todd's lattice) and Λ_{16} (Barnes-Wall's lattice) from the $8th, 21st$ and the $40th$ root of unity. These lattices are obtained by applying the canonical embedding to particular integral ideals of the above cyclotomic fields. The ideals are given in the table below. The lattices are indicated with $\Lambda_{n,L}$. Two generators for each ideal are given in the last column.

Lattice	N	Ideal
$D_{4,2}$	8	$(2, \theta+1)$
$E_{6,3}$	9	$(3, (\theta+1)^2)$
$E_{8,4}$	20	$(5, \theta-2)$
$K_{12,6}$	21	$(7, \theta+3)$
$\Lambda_{16,8}$	40	$(2, \theta^4+\theta^3+\theta^2+\theta+1)(5, \theta^2+2)$
$\Lambda_{24,12}$	39	$(3, \theta^3+\theta^2-1)(13, \theta-3)$ $(3, \theta^3+\theta^2+\theta+1)$

IV. Results

The figure below shows the performance over the Rayleigh fading channel of the rotated versions of the lattices D_4, E_6, E_8, K_{12} and Λ_{16}. Simulations were made up to dimension 8, while for higher dimensions we have plotted upper bounds. The bit error probability is given as a function of E_b/N_0 for $s = 4$ bits/symbol. The slopes of the curves asympotically correspond to the diversity order which is 2, 3, 4, 6 and 8 respectively. At 10^{-3} the gain over Z^8 is about 17dB and it exceeds 25dB at 10^{-5}. It is important to notice that



the maximal diversity reached with a reasonable trellis coded modulation does not exceed 6. The diversity of the rotated Leech lattice $\Lambda_{24,12}$ is 12. This is equivalent to a trellis coded QAM with 2^{44} states or a trellis coded PAM with 2^{22} states at 4 bits per symbol.

References

[1] J. H. Conway, N. J. Sloane: *Sphere packings, lattices and groups*, 2nd ed., 1993, Springer-Verlag, New York.

[2] M. Craig: "Extreme forms and cyclotomy," *Mathematika*, vol. 25, pp. 44-56, 1978.

[3] M. Craig: "A cyclotomic construction for Leech's lattice," *Mathematika*, vol. 25, pp. 236-241, 1978.

Real-Number DFT Codes on a Fading Channel

Jiun Shiu and Ja-Ling Wu
Dept. of Comp. Sci. & Info. Eng., Nat'l Taiwan Univ.
Taipei, 10764, Taiwan, Republic of China

Abstract —

The utilization of real-number DFT codes for a multiplicative channel is introduced in this paper. By the proposed encoding procedure, some redundancies can be added into the transmitted data. With these redundancies, syndromes for the parameters of a fading channel can be obtained from the received data. The decoding algorithm for real-number DFT codes can be used to calculate the fading parameters with these syndromes.

I. Introduction

In 1981, Marshall first defined error control codes for real or complex data and suggested that real-number codes could have applications similar to those of Reed-Solomon codes. Wolf, with a different view, took real-number codes as a new technique for solving signal processing problems such as impulsive noise cancellation in information transmission.

A common feature of previous studies is that the channel error model is assumed to be additive. In this paper, the real-number decoding method for multiplicative channel error model (which corresponds to the situation of transmitting over a fading channel in practice) will be investigated.

II. Encoding and Decoding Scheme for a Fading Channel

Usually the effect of a fading channel is modeled by a slowly varying component multiplying the transmitted signal, that is

$$r_i = y_i \cdot e_i + n_i \tag{1}$$

where y_i is the transmitted signal, e_i the multiplicative parameters of a fading channel, n_i the background noise, and r_i the received signal. In a block coding scheme, we can also assume that the index i is in the range of $0, 1, 2, \ldots, N-1$.

A multiplication can be transformed into an addition by taking logarithm. However, since the signals under consideration are assumed to be complex, complex logarithm function are required. It can be easily derived from eqn. (1) that

$$\log_c r_i = \log_c y_i + \log_c e_i + \hat{n}_i \tag{2}$$

where $\hat{n}_i = \log_c(1 + \frac{n_i}{y_i \cdot e_i})$. It should noted that when $n_i \ll y_i \cdot e_i$, $\hat{n}_i$ will approach 0. Since e_i is slowly varying, both e_i and $\log_c e_i$ can be viewed as a lowpass signal. Therefore, it is reasonable to assume that $\log_c e_i$ can be obtained from the sum of some unknown low frequency components E_k, that is

$$\log_c e_i = \sum_{l=1}^{\nu} E_{k_l} \cdot e^{2\pi \frac{ik_l}{N}} \tag{3}$$

where k_l is the location for a nonzero frequency components, and E_{k_l} is the magnitude of that component. Now suppose that y_i is encoded as

$$y_i = \begin{cases} 1 & i = 0, 1, \ldots, N-K-1 \\ x_i & i = N-K, N-K+1, \ldots, N-1 \end{cases} \tag{4}$$

The first $N-K$ equations in eqn. (2) become the desired syndromes

$$S_i = \log_c r_i = \log_c e_i + \hat{n}_i \tag{5}$$

These noisy syndromes can readily be input to some decoding algorithms for the DFT codes. to compute E_k, provided that the number of nonzero terms of E_k in eqn. (5). After E_k is computed, an estimation of y_i can then be derived. In this way, one can estimate out the channel parameters e_i and, at the same time, the transmitted data x_i.

Good k/(k+1) Time-Varying Convolutional Encoders from Time-Invariant Convolutional codes

Kazuhiko YAMAGUCHI† and Hideki IMAI‡

† Dept. of Comp. Sci. and Info. Math., The Univ. of Electro-Comm., Cho-fu, Tokyo, 182, Japan,
‡ 3rd Dept., Inst. of Industrial Sci., Univ. of Tokyo, Roppongi, Minato-ku, Tokyo, 106, Japan

Abstract — **This paper studies the construction of good time-varying convolutional codes from the relation between time-varying and time invariant encodes.**

I. Introduction

In this study, we discuss construction of good time-varying convolutional encoders, i.e. punctured convolutional codes, from time-invariant convolutional encoder.

In general, an k/n time-varying convolutional code can be described by p $n \times k$ generator matrices, where p is the period of time varying. Such an k/n time-varying convolutional code may have better error protection capability than the best convolutional codes which has time-invariant encoder with the same number of states.

The $k/k+1$ time-varying convolutional encoders, which are discussed here, are realized as $k/(k+1)$ punctured convolutional codes from $1/(k+1)$ convolutional codes. Any of the encoders may not have better error protection capability than the best time-invariant convolutional codes, since such a $k/(k+1)$ time-varying convolutional encoder can translate into time-invariant encoder with the same or less number of states.

However such a time-varying convolutional encoder has benefit in applications. The Viterbi decoder of a time-varying convolutional encoder has smaller complexity than an ordinal one.

The time-varying convolutional encoder have been independently studied from time-invariant codes. We show transformation $k/(k+1)$ time-varying convolutional encoders from *equivalent* $k/(k+1)$ time-invariant encoders, and discuss good time-varying convolutional encoder. Some of the time-varying convolutional encoder which derived from the best known $k/(k+1)$ time-invariant codes, has the same free distance as the known time-varying/punctured encodes.

II. Transformation between time-varying and time-invariant convolutional encodes

The $k/(k+1)$ time-varying convolutional encoder discussed here (i.e. $k/(k+1)$ punctured code) has period k, k generator matrices. only one matrix is 1×2 and other $k-1$ matrices are 1×1. If 1×2 generator matrix is used in i-th interval, the $k/(k+1)$ time-varying convolutional encoder can be wrote with $k+1$ polynomials as

$$\{g_1(D),\ g_2(D),\ \cdots,\ (g_i(D)\ g_0(D)),\ g_{i+1}(D),\ \cdots,\ g_k(D)\},$$

where $g_j(D) = g_j^0 + g_j^1 D + g_j^2 D^2 + \cdots$. Let the generator matrix of corresponding $k/(k+1)$ time-invariant convolutional encoder as

$$\begin{pmatrix} g*_{1,1} & g*_{1,2} & \cdots & g*_{1,k+1} \\ g*_{2,1} & g*_{2,2} & \cdots & g*_{2,k+1} \\ \vdots & & & \vdots \\ g*_{k,1} & g*_{k,2} & \cdots & g*_{k,k+1} \end{pmatrix},$$

where $g*_{i,j}(D) = g*^0_{i,j} + g*^1_{i,j} D + g*^2_{i,j} D^2 + \cdots$.

The $k/(k+1)$ time-varying and time-invariant convolutional encoders are equivalent if $g^x_p = g^y_{z,p}$, $g^x_0 = g^y_{z,k+1}$ where

$$\begin{aligned} x &= ky - z + p, \\ y &= x + k - p, \\ z &= \lfloor (k-x) \bmod k \rfloor + p, \\ 1 &\le z \le k, \\ 1 &\le p \le k, \\ g*^0_{j,p} &= 0 \quad (\text{for } p < j \le k), \qquad (1) \\ g*^0_{j,k+1} &= 0 \quad (\text{for } i < j \le k), \qquad (2) \end{aligned}$$

are satisfied. proof and detail discussions are omitted here.

Most of best time-invariant convolutional codes satisfy the conditions (1) and (2) with permutation the inputs and outputs. Let us consider a 2^M-states time-invariant convolutional encoder which satisfies the conditions. The encoder is translated into 2^M, 2^{M+1} $\cdots$ or 2^{M+p}-states time-varying convolutional encoder. The exact number of states is given by the discussion from the equations above (omitted here). We can easily find 2^M or 2^{M+1}-states time-varying convolutional encoder from best or good 2^M-states time-invariant convolutional encoder, where we call good code which has maximum free distance for given number of states.

III. Examples

Here we show the two examples of transformation. The best 32-state 2/3 time-invariant convolutional code

$$\begin{pmatrix} 1+D & D+D^2 & 1+D+D^2 \\ D^2 & 1 & 1+D+D^2+D^3 \end{pmatrix},$$

is translated as 64-state 2/3 time varying code;

$$\{1+D^2+D^3,\ (1+D^3+D^5\ \ 1+D+D^2+D^3+D^4+D^5+D^6)\}.$$

the best 64-state 3/4 time-invariant convolutional code

$$\begin{pmatrix} D+D^2 & 1 & D^2 & 1+D \\ D+ & D+D^2 & 1+D+D^2 & 1+D^2 \\ 1+D+D^2 & 1+D & 1+D & 1 \end{pmatrix}$$

is translated as 64-state 2/3 time varying code

$$\{1+D+D^2+D^3+^4+D^6,\ \ D+D^2+D^3+D^4+D^6,$$
$$(1+D^2+D^4+D^5+D^6+D^7\ \ 1+D+D^2+D^3+D^7)\}.$$

The use of the conditions (1) and (2) with free distance bound for convolutional codes gives an efficient algorithm to find good time-varying convolutional codes.

Cascaded Convolutional Codes

Lance C. Perez Daniel J. Costello, Jr.
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Abstract — **The construction and performance of cascaded convolutional codes is investigated. An interleaver is used between the inner and outer codes to redistribute errors out of the inner decoder. In addition, the structure of the interleaver is exploited to improve the distance properties of the overall cascaded code. This configuration is shown to have a performance advantage compared to a single complex convolutional code with the same rate and decoder complexity.**

I. Introduction

In this paper, the design and performance of cascaded convolutional codes [1] for the additive white Gaussian noise channel is investigated. A cascaded convolutional code is the serial concatenation of two binary convolutional codes. They are decoded using the serial concatenation of the decoders corresponding to the two convolutional codes. In order to realize the full performance potential of cascaded convolutional codes, it is necessary to pass soft information from the inner decoder to the outer decoder [2]. In this work, the *maximum a posteriori* (MAP) algorithm developed and described in [3] is used to decode the inner code.

II. A Simple Example

A block diagram of a simple cascaded convolutional coding scheme is shown in Figure 1. The outer convolutional code is a maximal free distance (MFD) rate $k_1/n_1 = 2/3$ code with total encoder memory $\nu_1 = 3$ and free distance $d_{free_1} = 4$ The generator matrix of this $(3, 2, 3)$ code in nonsystematic feedforward form is given by

$$G_1(D) = \begin{bmatrix} 1 & D & 1+D \\ D^2 & 1 & 1+D+D^2 \end{bmatrix}.$$

The inner code is a maximal free distance rate $k_2/n_2 = 3/4$ code with total encoder memory $\nu_1 = 3$ and free distance $d_{free_2} = 4$. The generator matrix of this $(4, 3, 3)$ code in nonsystematic feedforward form is given by

$$G_2(D) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1+D & D & 1 \\ 0 & D & 1+D^2 & 1+D^2 \end{bmatrix}.$$

The inner and outer convolutional codes will be referred to as the *component codes*. The overall cascaded code has rate

$$R = \frac{k_1}{n_1} \times \frac{k_2}{n_2} = \frac{2}{4} = \frac{1}{2}.$$

If the generator matrices of the two codes in this example are multiplied, ignoring the effect of the interleaver, the resulting generator matrix, $G(D)$, is given by

$$\begin{bmatrix} 1 & 1 & D+D^3 & D^2+D^3 \\ D^2 & 1+D^3 & 1+D^2+D^3+D^4 & D+D^2+D^3+D^4 \end{bmatrix}$$

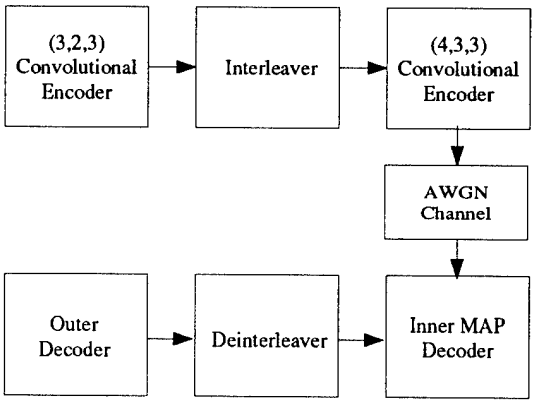


Fig. 1: Block diagram of a cascaded convolutional coding scheme.

and the resulting code is called the *composite code*. The generator matrix $G(D)$ realizes a $(4, 2, 7)$ code with $d_{free} = 6$.

Note that the constraint length of the composite code is greater than the sum of the constraint lengths of the component codes. (It may be that the $(4, 2, 7)$ code is not in minimal form.) Unlike concatenated block codes and product codes, the free distance of the overall code is not the product of the free distances of the component codes. (Thus, using MFD codes for the component codes is not necessarily optimal.) However, by carefully designing the interleaver, the free distance of the cascaded code may be increased and made larger than that of a single convolutional code of the same complexity. In addition, cascaded convolutional codes tend to have less dense distance spectra than a comparable single code. As the code complexity increases, the sparse distance spectra of cascaded convolutional codes improves their performance at low and moderate signal to noise ratios.

III. Conclusion

Cascaded convolutional codes appear to be a reasonable alternative to complex convolutional codes. The combination of soft-output decoding and interleaving enables cascaded codes to outperform a single code of the same complexity. Cascaded convolutional codes also lend themselves to a form of iterative decoding [4].

References

[1] F. Pollara and D. Divsalar, "Cascaded convolutional codes," *TDA Progress Report*, pp. 202-205, August 1992.

[2] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *Proc. 1993 IEEE GLOBECOM '89*, Dallas, Texas, pp. 47.1.1–47.1.7, 1989.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, **IT-20**, pp. 284–287, 1974.

[4] C. Berrou, A. Glaieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: TURBO codes," *Proc. ICC'93*, Geneva, Switzerland, 1993, pp. 1064–1070.

An Algorithm for Identifying Rate $(n-1)/n$ Catastrophic Punctured Convolutional Encoders

Feng-Wen Sun[†]
A. J. Han Vinck[‡]
Elec. Eng. Dept. McGill Univ. Montreal H3A 2A7 Canada[†]
Institute of Exper. Maths Essen Univ., 45326 Essen Germany[‡]

Abstract — **An algorithm is presented to identify catastrophic encoders when the original rate $1/b$ encoder is antipodal. The key technique is to use the syndrome former to determine the constraint length of the dual code. The major part of the algorithm solves a linear equation of $\nu+1$ variables, where ν is the constraint length of the original rate $1/b$ code.**

I. Introduction

Both Viterbi decoding and sequential decoding of high-rate convolutional codes are greatly simplified by employing the class of punctured convolutional codes, which are obtained by periodically deleting a part of the bits of a low-rate code. The simple structure of the low-rate code can be utilized to encode and decode the high-rate code.

Good punctured convolutional codes are generally obtained by computer searches. During the searching procedure, catastrophic encoders, which result in infinite number of decoded errors from finite channel errors, must be identified. This appears to be a nontrivial problem since some deleting maps may result in catastrophic encoders even if the original code is noncatastrophic. Therefore, in order to speed up the search for good punctured codes, an efficient algorithm to identify catastrophic encoders is highly desirable.

In this work, we propose an algorithm to identify catastrophic encoders of rate $(n-1)/n$ punctured codes when the original encoder is antipodal. The algorithm is computationally efficient for both large and small constraint lengthes.

From [2], we know that a punctured convolutional encoder obtained from an antipodal encoder is noncatastrophic if and only if it is minimum. The algorithm to be presented first finds a nonzero codeword of the dual of the punctured rate $(n-1)/n$ code. Since the dual code is a rate $1/n$ code, its minimum encoder can be easily found from any nonzero codeword. Thus, the overall constraint length of a minimum encoder of the dual code is determined. The constraint length of a minimum encoder is always equal to that of the minimum encoder of its dual. In this way, the minimality of the punctured encoder, thus the catastrophic property, is determined.

II. The Algorithm

For a fixed deleting matrix and a finite weight sequence x, $ext(x)$ is defined to be a sequence with the property that $ext(x)$ redueces to x after puncturing by applying the deleting matrix, and those deleting positions are equal to zero. If we further define dual of a convolutional code as anti-Laurant sequences orthogonal to all the codewords of the convolutional codes. We have the following two lemmas.

Lemma 1 *A finite-weight sequence $\underline{x}$ of n-dimensional vectors is in the dual of the punctured convolutional code if and only if* $\mathrm{ext}(\underline{x})$ *is in the dual of the original rate $1/b$ code.*

Lemma 2 *For any state of the syndrome former of the dual of an antipodal rate $1/b$ convolutional code , there exist two n-dimensional vectors, say $x, x\prime$, such that when the syndrome former starts from this state, the input* $\mathrm{ext}(x)$ $(\mathrm{ext}(x\prime))$ *causes the syndrome former to transfer to another state with the* **all-zero output**. *Any one of the two vectors can be found in no more than $n(\nu+1)$ binary operations.*

By these lemmas, we can establish the following algorithm to identify catastrophic encoders.

1. Initialize the adjoint-obvious realization [4] of $D^{\nu}G^{T}(D^{-1})$ as the all-zero state.
2. Find
   a sequence of n-dimensional vectors $(x_1, x_2, \cdots, x_{\nu+1})$ such that $x_1 \neq 0$ and $(ext(x_1), ext(x_2), \cdots, ext(x_{\nu+1}))$ is a valid input sequence of the syndrome former with state transitions $(0, S_1, S_2, \cdots, S_{\nu+1})$.
3. Find a nontrivial solution $(b_1^*, \cdots, b_{\nu+1}^*)$ of the equation
$$\sum_{i=1}^{\nu+1} b_i S_i = 0.$$
4. Calculate the sum
$$\underline{y} = \sum_{i=1}^{\nu+1} b_i^*(0, \cdots, 0, x_1, \cdots, x_i).$$
5. Represent $\underline{y}$ in n polynomials.
6. If all the degrees of the n polynomials are less than ν or the degree of their greatest common divisor is larger than one, the punctured convolutional encoder is catastrophic.

END

This algorithm significantly redueces the computational complexity of all known algorithms [1, 3].

References

[1] K. J. Hole, "An algorithm for determining if a rate $(n-1)/n$ punctured convolutional encoder is catastrophic," *IEEE Transactions on Communications*, vol. COM-39, pp. 386–389, Mar. 1991.

[2] K. J. Hole, "Rate $k/(k+1)$ minimum punctured convolutional encoders," *IEEE Transactions on Information Theory*, vol. IT-37, pp. 653–655, May 1991.

[3] J. L. Massey and M. K. Sain, "Inverses of linear sequential circuits," *IEEE Trans. Comput.*, vol. C-17, pp. 330–337, Apr. 1968.

[4] G. D. Forney Jr., "Structural analysis of convolutional codes via dual codes," *IEEE Transactions on Information Theory*, vol. IT-19, pp. 512–518, July 1973.

Generalized Hamming Weights of Convolutional Codes[1]

Joachim Rosenthal[2] Eric Von York
Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556-5683
e-mail: Joachim.Rosenthal@nd.edu, Eric.York@nd.edu

Abstract — **Motivated by applications in cryptology K. Wei introduced in 1991 the concept of a generalized Hamming weight for a linear block code. In this paper we define generalized Hamming weights for the class of convolutional codes and we derive several of their basic properties.**

I. Introduction

An important set of code parameters defined for a linear block code are the so called generalized Hamming weights first introduced by Wei in [1]. By definition the r-th generalized Hamming weight $d_r(C)$ of a linear block code C is equal to the smallest support of any r-dimensional subcode of C. In particular $d_0(C) = 0$ and $d_1(C)$ is equal to the distance of C.

In this way every $[n, k]$ linear block code has associated a whole weight hierarchy

$$0 = d_0(C) \leq d_1(C) < \cdots < d_k(C) \leq n. \qquad (1)$$

In this correspondence we will study the weight hierarchy of a convolutional code. After formally introducing this concept we will derive in the next section several of the basic properties. In particular we will show that the generalized Hamming weights form an infinite strictly increasing sequence $d_i(C)$ of positive integers. The main result (Theorem 4) is a generalized Griesmer bound.

II. Definitions

Let $\mathbf{F}_q$ be the Galois field of q elements, $\mathbf{F}_q[D]$ be the polynomial ring over $\mathbf{F}_q$ and $\mathbf{F}_q(D)$ the ring of rational functions. In the following it will be convenient to view elements of $\mathbf{F}_q(D)$ as infinite (periodic) power series of the form $\sum_{i=0}^{\infty} x_i D^i, x_i \in \mathbf{F}_q$. Let C be a rate k/n convolutional code represented through a non-catastrophic encoder $G(D)$. Without loss of generality we will assume that the matrix $G(D)$ which is defined over $\mathbf{F}_q[D]$ is in row proper form, in other words we will assume that the "high order coefficient matrix" has full row rank. We also will assume that $G(D)$ has ordered row (Kronecker) indices

$$\nu_1 \geq \cdots \geq \nu_k$$

where the indices ν_i are formally defined through:

$$\nu_i = \max\{\deg(g_{ij}) \mid 1 \leq j \leq n\}\ ,\ i = 1, \ldots, k.$$

We will denote the memory, complexity and constraint length of a convolutional code by m, c, and η respectively. In terms of the Kronecker indices we have: $m = \nu_1$, $c = \sum_{i=1}^{k} \nu_i$ and $\eta = n(\nu_1 + 1)$.

In an obvious way we can view C also as an (infinite dimensional) linear $\mathbf{F}_q$ vector space. Let

$$\{u_1(D), \ldots, u_r(D)\}$$

be r vectors in $\mathbf{F}_q^k(D)$, that are linearly independent over $\mathbf{F}_q$. Since $G(D)$ has by assumption linearly independent rows it follows that

$$\{u_1(D)G(D), \ldots, u_r(D)G(D)\} \subset C \subset \mathbf{F}_q^n(D)$$

defines an r-dimensional subspace of C and clearly every r-dimensional subspace $U \subset C$ is of this form.

Definition 1 *Let $U \subset C$ be a linear subspace of C. Then*

$$\chi(U) := \{(i,j) \mid \exists \left(\sum x_{1j} D^j, \ldots, \sum x_{nj} D^j\right) \in U, x_{ij} \neq 0\}$$

is called the support *of U and*

$$d_r(C) := \min\{|\chi(U)| \mid U \subset C \text{ and} \dim U = r\}$$

is called the rth generalized Hamming weight *of C.*

Note that the generalized Hamming weights are well defined for any positive integer r and not just for $r = 0, \ldots, k$ as it is the case for block codes. Also note that if U is one dimensional and $u \in U$ is any nonzero codeword then $|\chi(U)|$ is nothing else than the usual Hamming weight $w(u)$ of the codeword u. In particular it follows in analogy to the block code case that $d_1(C)$ is equal to the free distance of C.

III. Basic properties.

Lemma 2 *Let C be a convolutional code of rate $\frac{k}{n}$ and memory m. In order to compute $d_i(C)$ it is enough to consider subspaces of the form*

$$U = span\{u_1(D)G(D), \ldots, u_r(D)G(D)\}$$

where $u_i(D) \subset \mathbf{F}_q^k[D]$ and the $\deg(u_i(D)) < (m^2 + mr)n$.

The following Lemma is a natural generalization of Wei's monotonicity theorem [1, Theorem 1] for block codes.

Theorem 3 *The generalized Hamming weights of a convolutional code form a (strictly) increasing set of positive integers*

$$0 = d_0(C) < d_1(C) < d_2(C) < \cdots$$

Theorem 4 *Let C be a binary convolutional code of rate k/n and having a basic encoder $G(D)$ with Kronecker indices $\nu = (\nu_1, \ldots, \nu_k)$. Let γ be a positive integer and let $\mathcal{K} = \sum_{i=1}^{k} \max(\gamma - \nu_i + 1, 0)$. Then the rth generalized Hamming weight of C satisfies*

$$d_r(C) + \sum_{i=1}^{\mathcal{K}-r} \left\lceil \frac{d_r(C)}{2^i(2^r - 1)} \right\rceil \leq n\gamma + n. \qquad (2)$$

Example 5 *Let C be the rate $\frac{1}{2}$, $m = 2$, $\eta = 6$ code with generator matrix $G(D) = (D^2 + D + 1, D^2 + 1)$. Then one can verify that $d_1(C) = 5$ and $d_i(C) = 2(i-1) + \eta$, $\forall i > 1$.*

References

[1] V. K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 5, pp. 1412–1418, 1991.

[1] An extended version of this paper has appeared as a report: CWI Report BS-R9507, Amsterdam, The Netherlands, 1995.

[2] This author was supported by NSF grant DMS-9400965

Upper Bounds on the Probability of the Correct Path Loss for List Decoding of Fixed Convolutional Codes [1]

Rolf Johannesson
Dept. of Information Theory
Lund University
P.O. Box 118
S-221 00 LUND, Sweden

Kamil Sh. Zigangirov
Dept. of Telecommunication Theory
Lund University
P.O. Box 118
S-221 00 LUND, Sweden

Abstract — **In list decoding (M-algorithm) the decoder state space is typically much smaller than the encoder state space. Hence, it can happen that the correct path is lost. This is a serious kind of error event that is typical for list decoding. In this paper two upper bounds on the probability of correct path loss for list decoding are given. For fixed convolutional codes counterparts to Viterbi's upper bounds for maximum-likelihood decoding of fixed convolutional codes are proved. Finally, it is shown that there exists a fixed convolutional code whose probability of correct path loss when decoded by list decoding satisfies a simple expurgated bound.**

I. Introduction

Viterbi decoding is an example of a non-backtracking decoding method that at each time instant examines the total encoder state space. The error correcting capability of the code is fully exploited.

In *list decoding* (M-algorithm) we first limit the resources of the *decoder*, then we choose an encoding matrix with a state space that is larger than the *decoder* state space. Thus, assuming the same decoder complexity, we use a more powerful code with list decoding than with Viterbi decoding. A list decoder is a very powerful non-backtracking decoding method that does not fully exploit the error correcting capability of the code.

List decoding is a breadth-first search of the code tree. At each depth only the L most promising subpaths are extended, not all, as is the case with Viterbi decoding. These subpaths form a *list* of size L.

Since the search is breadth-first, all subpaths on the list are of the same length and finding the L best extensions reduces to choosing the L extensions with the largest values of the cumulative Viterbi metric.

II. The correct path loss problem

Since only the L best extensions are kept it can happen that the correct path is lost. This is a very severe event that causes many bit errors. If the decoder cannot recover a lost correct path it is of course a "catastrophe", i.e., a situation similar to the catastrophic error propagation that can occur when a catastrophic encoding matrix is used to encode the information sequence.

The list decoder's ability to recover a lost correct path depends heavily on the type of *encoder* that is used. A systematic encoder supports a spontaneous recovery.

[1] This work was supported in part by the Swedish Research Council for Engineering Sciences under Grants 92-661 and 94-83.

III. Upper bounds on the probability of correct path loss

The correct path loss on the ith step of a list decoding algorithm is a random event $\mathcal{E}_i$ which consists of deleting at the ith step the correct codeword from the list of the L most likely codewords.

To upper bound $P(\mathcal{E}_i)$ we introduce the *l-list generating function for the path weights* $T_l(D)$. Consider the trellis for a rate $R = b/c$ and memory m fixed convolutional code. At a given depth consider the set of 2^{bm} paths of least weight leading to the 2^{bm} states. Order these paths according to increasing weights and let w_j denote the weight of the jth path ($w_0 = 0$). Introducing

$$T_l(D) = \sum_{j=l}^{2^{bm}-1} D^{w_j},$$

the l-list generating function of the path weights, we can prove the following

Theorem 1 *For the BSC with crossover probability ε and fixed convolutional codes with l-list generating function $T_l(D)$ the probability of correct path loss is upper bounded by*

$$P(\mathcal{E}_i) \le \min_{1\le l\le L} \frac{T_l(D)\,|_{D=\sqrt{4\varepsilon(1-\varepsilon)}}}{L-l+1}.$$

□

For the Gaussian channel we have the corresponding bound:

Theorem 2 *For the channel with additive white Gaussian noise (AWGN) with signal-to-noise ratio E_b/N_0 and fixed convolutional codes of rate R with l-list generating function $T_l(D)$ the probability of correct path loss is upper bounded by*

$$P(\mathcal{E}_i) \le \min_{1\le l\le L} \frac{T_l(D)\,|_{D=e^{-RE_b/N_0}}}{L-l+1}.$$

□

Furthermore, we can prove

Theorem 3 *There exists a fixed convolutional code satisfying the following expurgated bound:*

$$P(\mathcal{E}_i) \le L^{-\frac{\log_2\sqrt{4\varepsilon(1-\varepsilon)}}{\log_2(2^{1-R}-1)}} \cdot O(1).$$

□

References

[1] Kamil Sh. Zigangirov and Harro Osthoff: "Analysis of Global-list Decoding for Convolutional Codes". European Transaction on Telecommunications, No. 2, 1993.

Minimal, Minimal-Basic, and Locally Invertible Convolutional Encoders

Ajay Dholakia, Donald L. Bitzer, Havish Koorapaty[1], and Mladen A. Vouk
Dept. of Comp. Sci., North Carolina State Univ., Raleigh, NC 27695-8206, USA

Abstract — **Rate-k/n locally invertible convolutional encoders are defined. It is shown that a basic locally invertible encoder is minimal-basic. Local invertibility is used to re-derive Forney's [1] upper and lower bounds on the maximum number of consecutive all-zero branches in a convolutional codeword. A time-domain test for minimality [2] of an encoder is given.**

I. Introduction: Time-Domain Approach

A rate-k/n convolutional encoder is characterized in the time-domain by a discrete semi-infinite generator matrix $\mathbf{G}$ [3]. Consider a finite section $\mathbf{G}_{[m,\, m+\beta]}$ of $\mathbf{G}$ given by

$$\mathbf{G}_{[m,\, m+\beta]} = \begin{bmatrix} \mathbf{G}_m & & & \\ \mathbf{G}_{m-1} & \mathbf{G}_m & & \\ \vdots & \vdots & \ddots & \\ \mathbf{G}_0 & \mathbf{G}_1 & & \mathbf{G}_m \\ & \mathbf{G}_0 & & \vdots \\ & & \ddots & \mathbf{G}_0 \end{bmatrix}. \qquad (1)$$

This matrix represents a mapping between a $k(m+\beta+1)$-bit information subsequence $\mathbf{u}_{[t-m,\, t+\beta]}$ and an $n(\beta+1)$-bit encoded subsequence $\mathbf{v}_{[t,\, t+\beta]}$, given by $\mathbf{v}_{[t,\, t+\beta]} = \mathbf{u}_{[t-m,\, t+\beta]}\mathbf{G}_{[m,\, m+\beta]}$, where $t \geq 0$ is the time index and $\mathbf{u}_{[-m,\, -1]}$ is the the starting state of the encoder.

A time-domain approach for analyzing rate-k/n convolutional encoders has recently been presented in [4, 5, 6]. This approach is based on performing elementary column operations on a finite section $\mathbf{G}_{[m,\, m+\nu]}$ of $\mathbf{G}$, corresponding to $(\nu+1)$ output branches, to obtain its column canonical form $\mathbf{G}^c_{[m,\, m+\nu]}$. A matrix is in *column canonical* form if 1) All all-zero columns appear as the left-most columns of the matrix, and 2) The last nonzero element in a column is the only nonzero element in its row, is a 1, and appears above the last nonzero element in succeeding columns. The last nonzero element in each column is called a *pivot* if it is the only nonzero element in the column.

II. Main Results

Definition 1 *A rate-k/n convolutional encoder is locally invertible if $\mathbf{G}^c_{[m,\, m+\nu]}$ has a pivot in every nonzero row, i.e., if all the nonzero rows in $\mathbf{G}_{[m,\, m+\nu]}$ are linearly independent.*

The time domain test for a rate-k/n encoder being basic is the existence of k pivots in the last k rows of $\mathbf{G}^c_{[m,\, m+\nu]}$.

Theorem 1 *A basic rate-k/n convolutional encoder is minimal-basic if and only if it is locally invertible.*

A fast time-domain algorithm for testing whether a rate-k/n convolutional encoder is minimal-basic is as follows: 1) Compute $\mathbf{G}^c_{[m,\, m+\nu]}$, and 2) Inspect $\mathbf{G}^c_{[m,\, m+\nu]}$ to ascertain that all the nonzero rows have a pivot and that all the last k rows have pivots. In [7], it is shown that the test for minimal-basicity of an encoder requires a smaller section of $\mathbf{G}$, corresponding to only ν output branches.

Upper and lower bounds on the number of consecutive all-zero outputs of a rate-k/n minimal-basic encoder starting in a nonzero state, given in [1], may also be derived using the property of local invertibility. If a basic encoder is locally invertible at length $(\beta+1)$, the rank of $\mathbf{G}_{[m,\, m+\beta]}$ is equal to the number of nonzero rows in it. For such an encoder, an all-zero encoded subsequence $\mathbf{v}_{[t,\, t+\beta]}$ cannot be produced by a nonzero information subsequence $\mathbf{u}_{[t-m,\, t+\beta]}$ since there is a one-to-one mapping between the information and encoded subsequences at length $(\beta+1)$ [7]. Therefore, the required bounds on the number of consecutive all-zero outputs coincide with the bounds on the parameter β at which an encoder may achieve local invertibility. These bounds are derived in [7] and are shown to coincide with Forney's original bounds.

A rate-k/n encoder is minimal if and only if it has a polynomial inverse in D and a polynomial inverse in D^{-1} [2]. The time-domain test for minimality is given by the following theorem [7]:

Theorem 2 *A rate-k/n convolutional encoder is minimal if and only if $\mathbf{G}^c_{[m,\, m+\nu]}$ contains k pivots in the band of k rows operating on the information block $\mathbf{u}_t$.*

[1] H. Koorapaty is with Dept. of Elec. & Comp. Eng., NCSU, Raleigh, NC 27695-7911, USA.

References

[1] G.D. Forney, Jr. Structural analysis of convolutional codes via dual codes. *IEEE Trans. Inf. Theory*, IT-19(4):512–518, 1973.

[2] R. Johannesson and Z. x. Wan, "A linear algebra approach to minimal convolutional encoders", *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1219–1233, 1993.

[3] A. Dholakia. *Introduction to Convolutional Codes with Applications.* Kluwer Academic Publishers, Boston, 1994.

[4] H. Koorapaty, A. Dholakia, D.L. Bitzer, and M.A. Vouk. Rate-k/n convolutional encoders: A time-domain analysis. Tech. Rep. TR-95-04, Dept. Comp. Sci., NCSU, Raleigh, NC, 1995. Submitted to *IEEE Trans. Info. Theory.*

[5] H. Koorapaty, A. Dholakia, D.L. Bitzer, and M.A. Vouk. A time-domain approach to existence of polynomial inverses of convolutional encoders. To be presented at the 1995 Canadian Workshop on Information Theory, 1995.

[6] H. Koorapaty, A. Dholakia, D.L. Bitzer, and M.A. Vouk. Determination of polynomial inverses and syndrome formers of a convolutional encoder in time-domain. To be presented at the 1995 Information Theory Workshop, 1995.

[7] A. Dholakia, H. Koorapaty, D.L. Bitzer, and M.A. Vouk. A time-domain analysis of minimal convolutional encoders. Tech. Rep. TR-95-05, Dept. Comp. Sci., NCSU, Raleigh, NC, 1995. Submitted to *IEEE Trans. Info. Theory.*

First Order Representations for Convolutional Encoders

Joachim Rosenthal[1] Eric Von York

Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556-5683
e-mail: Joachim.Rosenthal@nd.edu, Eric.York@nd.edu

Abstract — **It is well known that convolutional codes are discrete time linear systems defined over a finite field. In this short correspondence we report about some important first order representations recently considered in the systems literature. Using this description we derive a new factorization of the well known "sliding block" parity check matrix often encountered in the coding literature.**

I. Generalized First Order Systems

Let $\mathbb{F}_q = \mathbb{F}$ be the Galois field with q elements and consider a $n \times k$ matrix $G(\mathcal{D})$ defined over the polynomial ring $\mathbb{F}[\mathcal{D}]$. $G(\mathcal{D})$ generates a $[n,k]$ convolutional code through:

$$C := \{w(\mathcal{D}) \mid w(\mathcal{D}) = G(\mathcal{D})\ell(\mathcal{D})\} \tag{1}$$

Note that we follow the convention in systems theory by writing all vectors as column vectors. From the point of view of systems theory (1) defines an MA-representation, the k-vectors $\ell(D)$ describe the set of latent variables and the set of n-vectors $w(D)$ describe the so called behavior, i.e. the code words. In the sequel we will assume that $G(\mathcal{D})$ is in column proper form having column indices $\mu_1, \ldots, \mu_k$ and overall constraint length $c := \sum_{i=0}^{k} \mu_i$. Then one has the following equivalent first order description.

Theorem 1 *There exist $(c+n-k) \times c$ matrices K, L and a $(c+n-k) \times n$ matrix M (all defined over $\mathbb{F}$) such that (1) is equivalently described through*

$$C := \{w(\mathcal{D}) \mid \exists x(\mathcal{D}) : Kx_{t+1} + Lx_t = Mw_t\}. \tag{2}$$

where $w(\mathcal{D}) = \sum w_t \mathcal{D}^t \in \mathbb{F}^n[\mathcal{D}]$, and $x(\mathcal{D}) = \sum x_t \mathcal{D}^t \in \mathbb{F}^c[\mathcal{D}]$. In addition the following minimality conditions are satisfied:

M1: K has full column rank.

M2: The full size minors of $[\mathcal{D}K + L \;\; M]$ are coprime.

Remark 2 If $G(\mathcal{D})$ is in addition a minimal encoder, then one can show (compare with [1, 3]) that the $c \times c$ full size minors of the pencil $\mathcal{D}K + L$ are coprime.

II. Duality

Let $H(\mathcal{D})$ be a $(n-k) \times n$ full rank polynomial matrix having the property that $H(\mathcal{D})G(\mathcal{D}) = 0$. $H(\mathcal{D})$ describes a parity check matrix for the convolutional code C introduced in (1) through:

$$C = \{w(\mathcal{D}) \mid H(\mathcal{D})w(\mathcal{D}) = 0\}. \tag{3}$$

Theorem 3 *There exist $c \times (c+k)$ matrices P, Q and a $n \times (c+k)$ matrix R (all defined over $\mathbb{F}$) such that (3) is equivalently described through*

$$\{w(\mathcal{D}) \mid \exists z(\mathcal{D}) : w_t = Rz_t, Pz_{t+1} = Qz_t\}. \tag{4}$$

In addition the following minimality conditions are satisfied:

M1': P has full row rank.

M2': The full size minors of $\begin{bmatrix} \mathcal{D}P+Q \\ R \end{bmatrix}$ are coprime.

The minimality conditions (M1') and (M2') guarantee that that after a possible permutation of the external variables the matrices P, Q, R in (4) have an equivalent description of the form:

$$P = (I \;\; 0) \quad Q = (A \;\; B) \quad R = \begin{pmatrix} C & D \\ 0 & I \end{pmatrix} \tag{5}$$

which in turn is equivalent to the representation:

$$x_{t+1} = Ax_t + Bu_t, \quad y_t = Cx_t + Du_t, \tag{6}$$

a well known description [2].

III. Factorization of the Sliding Block Matrix

One way of studying convolutional codes is usually through the use of the so called 'sliding block matrix' induced through the parity check matrix $H(\mathcal{D})$. In the sequel we provide a factorization of this matrix. Let K, L, M be defined as in (2) and define:

$$S = \begin{bmatrix} K & 0 & \ldots & 0 \\ L & K & \ddots & \vdots \\ 0 & L & \ddots & 0 \\ \vdots & \ddots & \ddots & K \\ 0 & \ldots & 0 & L \end{bmatrix} \quad T = \begin{bmatrix} M & 0 & \ldots & 0 \\ 0 & M & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & M \end{bmatrix}, \tag{7}$$

where we assume that both S and T consists of $s+1$ vertical blocks. Let U be a matrix with the property, that

$$\text{Ker } U = \text{Im } S.$$

Theorem 4 $w(\mathcal{D}) := \sum_{t=0}^{s} w_t \mathcal{D}^t \in C$ *if and only if*

$$UT\left(w_0^t, w_1^t, \ldots, w_s^t\right)^t = 0,$$

i.e. UT represents a factorization of the sliding block matrix of order s.

References

[1] M. Kuijper. *First-Order Representations of Linear Systems.* Birkhäuser, Boston, 1994.

[2] J. L. Massey and M. K. Sain. Codes, automata, and continuous systems: Explicit interconnections. *IEEE Trans. Automat. Contr.*, AC-12(6):644–650, 1967.

[3] M. S. Ravi and J. Rosenthal. A general realization theory for higher order linear differential equations. *Systems & Control Letters*, 1995. To appear.

[4] J. Rosenthal and E. York. Linear systems defined over a finite field, dynamic programming and convolutional codes. Preprint, May 1995.

[2]This author was supported by NSF grant DMS-9400965

Some Remarks on Convolutional Codes

Zhe-xian Wan
Dept. of Information Theory, Lund University,
P.O. Box 118, S-221 00 LUND, Sweden

I. On the Definition of Convolutional Codes

Let $1 \leq k \leq n$, F be a finite field, $F(D)$ be the field of rational function in D over F, and $F((D))$ be the field of formal Laurent series in D over F.

Definition 1 (Massey [1]) A rate k/n convolutional code over F is a k-dimensional subspace of the n-dimensional (row) vector space $F(D)^n$.

Definition 2 (Forney [2]) A rate k/n convolutional encoder over F is a k-input n-output constant linear causal finite-state sequential circuit. And a rate k/n convolutional code C over F is the set of outputs of the sequential circuit.

An equivalent formulation of Definition 2 (cf. [3]) is

Definition 2' A rate k/n convolutional code over F is a k-dimensional subspace of the n-dimensional (row) vector space $F((D))^n$ with a basis consisting of n-tuples of polynomials (or rational functions).

Clearly, a convolutional code in the sense of Definition 1 is a subcode of a convolutional code in the sense of Definition 2.

Definition 3 (Dholakia [4]) A rate k/n convolutional code over F is a k-dimensional subspace of the n-dimensional (row) vector space $F((D))^n$.

Clearly, a convolutional code in the sense of Definition 2' is a convolutional code in the sense of Definition 3. However, we have

Proposition 1 *There exist convolutional codes in the sense of Definition 3 which is not a convolutional code in the sense of Definition 2'.*

Proof: Let $f(D)$ be a formal Laurent series in D which is not ultimately periodic, and let C be the 1-dimensional subspace

$$F((D))(1, f(D))$$

of $F((D))^2$. Then C is a rate 1/2 convolutional code in the sense of Definition 3 but not a convolutional code in the sense of Definition 2'. □

Corollary 2 *There exist convolutional codes in the sense of Definition 3 which can not be realized by constant linear causal finite-state sequential circuit.*

II. On the Dual Code of a Convolutional Code

Let C be a rate k/n convolutional code in the sense of Definition 2'. Define

$$C^{\perp} = \{\underline{v}(D) \in F((D))^n \mid \underline{v}(D)\,\underline{c}(D)^{\mathsf{T}} = 0 \,\forall\, c(D) \in C\}.$$

Proposition 3 *Let C be a rate k/n convolutional code in the sense of Definition 2'. Then $C^{\perp}$ is a rate $(n-k)/n$ convolutional code in the sense of Definition 2'.*

Using invariant factor theorem Forney [2] actually proved this proposition. A simple elementary proof can be given by using Definition 1.

III. A Minimality Criterion of Encoding Matrices

Let $G(D)$ be a $k \times n$ matrix of full rank with entries in $F(D)$. If $G(D)$ is realizable and delayfree then $G(D)$ is called an encoding matrix of the convolutional code

$$C = \{\underline{v}(D) = \underline{u}(D)G(D) \mid \underline{u}(D) \in F((D))^k\}$$

in the sense of Definition 2. For any $\underline{u}(D) \in F((D))^k$, write

$$\underline{u}(D) = \underline{u}_{-m}D^{-m} + \ldots + \underline{u}_{-1}D^{-1} + \underline{u}_0 + \underline{u}_1 D + \underline{u}_2 D^2 + \ldots,$$

where $\underline{u}_i \in F^k$. Define

$$\begin{aligned} \underline{u}(D)P &= \underline{u}_{-m}D^{-m} + \ldots + \underline{u}_{-1}D^{-1}, \\ \underline{u}(D)Q &= \underline{u}_0 + \underline{u}_1(D) + \underline{u}_2 D^2 + \ldots. \end{aligned}$$

The set

$$\{\underline{u}(D)PG(D)Q \mid \underline{u}(D) \in F((D))^k\}$$

is called the abstract state space of C relative to the encoding matrix $G(D)$. If its cardinal attains the minimum, $G(D)$ is called a minimal encoding matrix (cf. [3]).

Proposition 4 *Let $G(D)$ be an encoding matrix. Then the following statements are equivalent.*

- *(a) $G(D)$ is a minimal encoding matrix.*
- *(d) $G(D)$ has a polynomial right inverse in D and a polynomial right inverse in D^{-1}.*
- *(e) For any $\underline{v}(D) = \underline{u}(D)G(D)$ where $\underline{u}(D) \in F((D))^k$, if $\underline{v}(D)$ is polynomial in D then so is $\underline{u}(D)$, and if $\underline{v}(D)$ is polynomial in D^{-1} then so is $\underline{u}(D)$.*

The equivalence of (a) and (d) was proved in [3]. Now the equivalence of (d) and (e) is proved.

References

[1] J. L. Massey, Coding theory, in *Handbook of Applicable Mathematics* (ed. by W. Ledermann and S. Vajda), Vol. V, Part B, Chapter 16, 623–676, Wiley, New York, 1985.

[2] G. D. Forney, Jr., Convolutional codes I: Algebraic structure. *IEEE Trans. Inform. Theory*, **16** (1990), 720–738.

[3] R. Johannesson and Z.-x. Wan, A linear algebra approach to minimal convolutional encoders, *IEEE Trans. Inform. Theory*, **39** (1993), 1219–1233.

[4] A. Dholakia, *Introduction to Convolutional Codes with Applications*, Kluwer Academic Publishers, Boston, 1994.

Improved Union Bound for Viterbi Decoder of Convolutional Codes

Marat V. Burnashev

Institute for Problems of Information Transmission, Russian Academy of Sciences
19 Ermolovoy str., 101447, Moscow, RUSSIA; e-mail: burn@ippi.ac.msk.su

Abstract — **Some improved version of the union bounds, expressed in the same terms is proposed.**

Transmission of binary information sequence over the BSC with crossover probability $0 < p < 1/2$ is considered. It is assumed that a noncatastrophical time-invariant convolutional encoder and Viterbi decoder are used. There are two types of performance characteristics that are usually used to describe the probabilistic behavior of such communication system. The first type characteristics describe the stationary behavior of the system (e.g. bit-error probability, averaged decoding delay, etc.). Usually they are of the main interest. The second type characteristics describe the behavior of the system at initial moment (e.g. first-error event probability). The most commonly used "union bounds" to upperbound any of mentioned above characteristics do not take into account some essential difference between these two types of characteristics [1,2]. We show that standard "union bounds" for stationary characteristics can be considerably improved preserving the same form and terms.

Denote by P_e the conditional probability that at any given moment the edge will be decoded incorrectly provided that all preceding semi-infinite information sequence was decoded correctly.

Theorem 1. *Conditional first-error event probability P_e satisfies the inequality*

$$P_e \le \sum_l \sum_w \frac{a(w,l)A_w}{1-A_w}\,(1-P_e)^l, \qquad (1)$$

where $a(w,l)$ is the number of codepaths of weight w and length l, and A_w is the error probability when testing two codewords of weights 0 and w [1].

Remarks. 1) Inequality (1) differs from a "standard" union bound by presence of factors $(1-P_e)^l$ in the right-hand side of (1). As a result it gives a nontrivial (i.e. $P_e < 1$) upper bound for any crossover probability $p < 1/2$ and this bound is always tighter then the "standard" union bound (which works only for some small p). 2) Inequality (1) can be expressed in terms of the generating function $T(D,L)$ with $L = 1-P_e$. 3) Inequality (1) remains also valid for some other channels (e.g. gaussian).

In the case of bit-error probability P_b we limit ourselves here only to the folowing result.

Theorem 2. *There exists some critical value p_0 such that if $p \le p_0$, then $P_b \le B$, where B is defined from the following system of equations*

$$E = \sum_i \sum_l \sum_w \frac{a(w,l)A_w}{1-A_w}\,(1-E)^l, \qquad (2)$$

$$B = \sum_i \sum_l \sum_w \frac{ia(w,l,i)A_w}{1-A_w}\,(1-E)^l, \qquad (3)$$

where $a(w,l,i)$ is the number of codepaths of weight w, length l and information weight i.

Remarks. 1) It is possible to evaluate the critical value p_0 . 2) If $p > p_0$, then the equation (2) will be replaced by some similar equation. 3) Both theorems are based on some reccurent relations and on a certain inequality from [3].

References

[1] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication*, New York: McGraw-Hill, 1979.

[2] M. V. Burnashev and D. L. Cohn, "On Bit-Error Probability for Convolutional Codes," *Probl. of Inform. Trans.*, vol. 26, no. 4, pp. 3–15, 1990.

[3] M. V. Burnashev, "On Extremal Property of Hamming Halfspaces," *Probl. of Inform. Trans.*, vol. 29, no. 3, pp. 3–5, 1993.

Assessing Generalization of Feedforward Neural Networks

Michael J. Turmon and Terrence L. Fine
School of Electrical Engineering, Cornell University, Ithaca, NY 14853

I. Introduction

Neural networks have been used to tackle what might be termed 'empirical regression' problems. Given independent samples of input/output pairs (x_i, y_i), we wish to estimate $f(x) = E[Y \mid X = x]$. The approach taken is to choose an approximating class of networks $\mathcal{N} = \{\eta(x; w)\}_{w \in \mathcal{W}}$ and within that class, by an often complex procedure, choose an approximating network $\eta(\cdot; w^*)$. The distance (in mean squared error) of this network from f can be separated into two terms: one for approximation or bias — choosing $\mathcal{N}$ large enough so that some $\eta(\cdot; w^0)$, say, models f well — and one for estimation or variance — how well the chosen $\eta(\cdot; w^*)$ performs relative to $\eta(\cdot; w^0)$. We address the latter term.

II. Problem Statement

Networks are parameterized by weight vectors $w \in \mathcal{W} \subseteq R^d$ and take inputs $x \in R^k$. In classification, network output is restricted to $\{0, 1\}$ while for regression it may be any real number. The complexity of the architecture $\mathcal{N}$ may be measured by the number of weights d or by its Vapnik-Chervonenkis (VC) dimension v. Performance of a network is measured by $\mathcal{E}(w) = E\,(\eta(x; w) - y)^2$ and the optimal net w^0 minimizes this. In practice, the law P is unknown so weights $w^* \in \mathcal{W}$ are chosen using the training set $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^n$ by minimizing $\nu_{\mathcal{T}}(w) = \frac{1}{n} \sum_{i=1}^n \big(\eta(x_i; w) - y_i\big)^2$.

The question of determining the relation between architecture complexity, estimation error, and training set size comes down to finding n large enough so that for a given d (or v), $\mathcal{E}(w^*) - \mathcal{E}(w^0) < \epsilon$ with high probability. We adopt this as a definition of reliable generalization. We can avoid dealing directly with the stochastically chosen network w^* by noting the triangle equality implies

$$0 \le \mathcal{E}(w^*) - \mathcal{E}(w^0) \le 2 \sup_{w \in \mathcal{W}} |\nu_{\mathcal{T}}(w) - \mathcal{E}(w)| \quad .$$

Vapnik [1] shows that $n = (9.2v/\epsilon^2) \log(8/\epsilon)$ is sufficient for reliable generalization. In cases where $\nu_{\mathcal{T}}(w^*) = 0$, this can be lowered [2] to $n = (5.8v/\epsilon) \log(12/\epsilon)$, but both are orders of magnitude higher than practice indicates.

III. Approximations via Poisson Clumping

For the large n we anticipate, the central limit theorem leads us to replace the original empirical process $\nu_{\mathcal{T}}(w) - \mathcal{E}(w)$ with the corresponding zero-mean Gaussian process $Z(w)$:

$$P(\| \, |\nu_{\mathcal{T}}(w) - \mathcal{E}(w)| \, \| > \epsilon) \simeq P(\| \, |Z(w)| \, \| > b)$$

where we have set $b = \epsilon\sqrt{n}$ and used the notation $\| \cdot \|$ for supremum over weights.

The Poisson clumping heuristic (PCH) [3] is a recently introduced tool for finding such exceedance probabilities. The PCH tells us that the region of weight space where $Z(w)$ exceeds level b is a group of clumps. The clump centers fall according to a Poisson process and the size $C_b(w)$ of a clump centered at w is chosen independently of all other clumps. The PCH leads to

$$P(\|Z(w)\| > b) \simeq \int_{\mathcal{W}} \frac{\bar{\Phi}(b/\sigma(w))}{E C_b(w)} \, dw \qquad (1)$$

where $\bar{\Phi}$ is the complementary cdf of $N(0,1)$ and $\sigma^2(w)$ is the variance of $Z(w)$. Loosely, the overall exceedance probability is a sum (integral) of the point exceedance probabilities, each scaled according to the number of weights that have exceedances with it.

This provides a means to get accurate approximations for the exceedance probabilities when the level b is large. For example, if network activation functions are twice differentiable and the variance has a unique maximum $\bar{\sigma}^2$ at $\bar{w} \in \mathcal{W}$, then $n = d\bar{\sigma}^2 K/\epsilon^2$ samples are sufficient for reliable generalization, where K is determined by P and $\mathcal{N}$. Explicit results for the problems of recognizing rectangles and halfspaces in R^k can also be obtained. These are again of order d/ϵ^2 but with constants far lower than previous upper bounds.

IV. Lower Bounds

These PCH-based estimates are of theoretical interest, but in practice evaluation of the constants is not possible due to ignorance of P. Now consider the following related tool for obtaining rigorous lower bounds to exceedance probabilities of $Z(w)$, where for simplicity we normalize $Z(w)$ by its standard deviation $\sigma = \sigma(w)$.

$$\begin{aligned} P(\|Z(w)/\sigma(w)\| > b) &= \int_{\mathcal{W}} \frac{\bar{\Phi}(b)}{E[D_b^{-1} | Z(w)/\sigma > b]^{-1}} \, dw \\ &\ge \bar{\Phi}(b) \int_{\mathcal{W}} \frac{1}{E[D_b | Z(w)/\sigma > b]} \, dw \end{aligned}$$

where D_b is the volume of $\{w : Z(w)/\sigma(w) > b\}$. Simple computations link this to the correlation $\rho = \rho(w, w')$ via

$$E[D_b | Z(w)/\sigma > b] \simeq \int_{\mathcal{W}} \bar{\Phi}((b/\sigma)\,\zeta) \, dw' \qquad (2)$$

with $\zeta = \zeta(w, w') = \big((1-\rho)/(1+\rho)\big)^{1/2}$.

This link provides the basis for estimating the exceedance probability empirically, without knowledge of P. Using the training set, compute $(y_i - \eta(x_i; w))^2$ at w and w' for all n points. This yields an estimate of ρ and in turn an estimate of ζ which can be used to compute the integral (2). Simulations for the examples of recognizing rectangles and halfspaces show that reasonable estimates of sample size can be obtained in the absence of analytical information about P and $\mathcal{N}$.

References

[1] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. Springer, 1982.

[2] A. Blumer et al., "Learnability and the Vapnik-Chervonenkis dimension," *Jour. Assoc. Comp. Mach.*, 36(4):929–965, 1989.

[3] D. Aldous, *Probability Approximations via the Poisson Clumping Heuristic*. Springer, 1989.

Optimal Stopping and Effective Machine Complexity in Learning

Changfeng Wang[1], Santosh S. Venkatesh[1], and J. Stephen Judd[2]

I. Introduction

We study learning in a general class of machines which return a (variable) linear form of a (fixed) set of nonlinear transformations of points in an input space. A fixed machine in this class accepts inputs X from an arbitrary input space and produces scalar outputs

$$Y = \sum_{i=1}^{d} \psi_i(X) w_i^* + \xi = \psi(X)' w^* + \xi. \qquad (1)$$

Here, $w^* = (w_1^*, \ldots, w_d^*)'$ is a fixed vector of real weights representing the *target concept* to be learned, for each i, $\psi_i(X)$ is a fixed real function of the inputs, with $\psi(X) = \big(\psi_1(X), \ldots, \psi_d(X)\big)'$ the corresponding vector of functions, and ξ is a random noise term.

We suppose that the learner receives an i.i.d., random sample of examples $(X_1, Y_1), \ldots, (X_n, Y_n)$ generated according to the joint distribution on input-output pairs (X, Y) induced through the medium of the (unknown) relation (1) and a fixed (unknown) distribution on input-noise pairs (X, ξ). The goal of the learner is to infer a hypothesis $w = (w_1, \ldots, w_d)'$ with small (mean-square) generalisation error $\mathbb{E}\big(Y - \psi(X)'w\big)^2$ on future random examples (X, Y) generated independently of the training sample from the same underlying distribution. Here $\mathbb{E}$ denotes expectation with respect to the underlying probability distribution generating the examples. Note that, as expected,

$$w^* = \arg\min_w \mathbb{E}\big(Y - \psi(X)'w\big)^2.$$

II. Results

We develop a rigourous characterisation of the time-dynamics of generalisation in this class of machines when a finite sample of examples is available and training is carried out by minimisation of the empirical (or training) error $\mathbb{E}_n\big(Y - \psi(X)'w\big)^2$ via gradient descent, where $\mathbb{E}_n$ denotes expectation with respect to the empirical distribution which puts equal mass $\frac{1}{n}$ on each of the n random examples which constitutes the sample. More specifically, given the sample, the batch-mode gradient descent algorithm provides an iterative refinement $\{ w(t), t \geq 0 \}$ of a hypothesis weight vector $w(t)$ representing the true concept w^*. The sequence of weight vector updates is specified recursively according to the usual gradient formulation:

$$w(0) \text{ is an arbitrary initial hypothesis in } \mathbb{R}^d;$$
$$w(t) = w(t-1) - \tfrac{1}{2}\epsilon \nabla \mathbb{E}_n\big(w(t-1)\big) \qquad (t \geq 1).$$

In the recursion, the integer parameter t denotes the update epoch and the positive parameter ϵ controls the rate of learning.

The empirical minimum mean-square estimate,

$$\widehat{w} = \arg\min_w \mathbb{E}_n\big(Y - \psi(X)'w\big)^2,$$

which corresponds to the estimate obtained in the limit of training over an infinity of time steps, is unbiased and consistent. Should we then carry training out to its limit? Surprisingly, perhaps, the answer is "No." Indeed, we show analytically that as training progresses in time three distinct phases in generalisation dynamics are evidenced. In the first phase, the generalisation error $\mathbb{E}\big(Y - \psi(X)'w\big)^2$ (where $\mathbb{E}$ denotes expectation with respect to the unknown underlying distribution generating the examples) decreases monotonically (keeping pace with a corresponding decrease in the training error); this phase is completed in $O(\log n)$ time steps where n is the number of examples. The behaviour grows more interesting in the second phase where the generalisation error exhibits complex dynamics and an *optimal stopping time* t_{opt} is evidenced at which the smallest generalisation error obtains; the second phase is also ephemeral and takes only $O(\log n)$ time steps. Finally, in the third phase, the generalisation error increases monotonically to a limiting value of error; this phase takes the rest of time. *Thus, best generalisation occurs not at the limit of training when the global minimum of the training error is achieved, but rather after a finite number of steps of the order of the logarithm of the sample size.*

One of the key concepts that emerges from our analysis is the formal notion of the *effective size* of a network. This is a time-varying, algorithm-dependent quantity which, in the limit of training over an infinity of time steps, coincides with the VC-dimension of the machine. As is well known, a salient characteristic of neural networks is that they often involve a very large number of adjustable parameters as compared to traditional statistical models (such as classification and regression models) with a resulting large VC-dimension. For a given (small) sample of fixed size, how then does one explain empirical claims reporting valid generalisation? Our results shed light on this puzzle: stopping learning at the optimal time results in a network with small complexity in the sense that its effective size at that time is typically substantially smaller than its effective size in the limit of training (the VC-dimension). *More generally, we show that the generalisation error of the machine during the training process is determined at each training epoch by the effective size of the machine at that epoch rather than its VC-dimension.* Our analysis provides a formal framework within which optimal stopping can be viewed as dynamically fitting machine complexity to the sample wherein best generalisation obtains when effective machine size best fits the sample size. Thus we rescue the prevailing intuition (Occam's razor) from its impending dilemma.

The study of generalisation dynamics leads naturally to a *new network size selection criterion* which can be viewed as a generalisation of Akaike's information criterion to cover not just network complexity (the effective machine size in the limit of training) but the time evolution of the learning process as well.

[1]Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104. The work of the first two authors was supported by the Air Force Office of Scientific Research under grant F49620-93-1-0120.

[2]Siemens Corporate Research, Princeton, NJ 08540.

On Batch Learning in a Binary Weight Setting

Shao C. Fang and Santosh S. Venkatesh[1]
Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract — **We consider the problem of inferring a finite binary sequence $w^* \in \{-1,1\}^n$ from a random sequence of half-space data $\{u^{(t)} \in \{-1,1\}^n : \langle w^*, u^{(t)}\rangle \geq 0, t \geq 1\}$. In this context, we show that a previously proposed randomised on-line learning algorithm dubbed Directed Drift [1] has minimal space complexity but an expected mistake bound exponential in n. We show that batch incarnations of the algorithm allow of massive improvements in running time. In particular, using a batch of $\frac{1}{2}\pi n \log n$ examples at each update epoch reduces the expected mistake bound to $O(n)$ in a single bit update mode, while using a batch of $\pi n \log n$ examples at each update epoch in a multiple bit update mode lead to convergence to w^* with a constant (independent of n) expected mistake bound.**

I. Introduction

Write $\mathbb{B} \triangleq \{-1,1\}^n$ for simplicity and let $\mathbb{B}^n \triangleq \{-1,1\}^n$ denote the vertices of the binary n-cube. Let $w^* \in \mathbb{B}^n$ be some fixed (but unknown) vertex. Suppose we are provided with a random labelled sequence of positive examples $\{u^{(t)}, t \geq 1\}$ of w^* obtained by independent sampling from the uniform distribution on the positive half-space of vertices

$$\mathbb{B}^n_+(w^*) \triangleq \{u \in \mathbb{B}^n : \langle w, u\rangle \geq 0\}.$$

Our goal is to infer the finite binary sequence w^* in an efficient (on-line) manner from the sample $\{u^{(t)}\}$.

II. Directed Drift

Directed Drift[1] is a randomised, on-line learning algorithm with minimal space complexity.

Algorithm D (*Directed Drift*). Given a confidence parameter $\delta > 0$ and a sample of positive examples $\{u^{(t)}, t \geq 1\}$ generated by independent sampling from the uniform distribution on $\mathbb{B}^n_+(w^*)$, the algorithm generates a hypothesis w which, with confidence in excess of $1-\delta$, coincides with the concept w^*.

D1. [Initialise.] Set epoch $t \leftarrow 1$, confidence counter $T \leftarrow 0$, and select an arbitrary initial hypothesis $w \in \mathbb{B}^n$.

D2. [Is the hypothesis consistent on the example?] Set $Y \leftarrow \langle w, u^{(t)}\rangle$.

D3. [If it ain't broke, don't fix it.] If $Y \geq 0$, increment the confidence counter $T \leftarrow T+1$: if $T \geq \sqrt{\frac{\pi n}{2}} \log \delta^{-1}$, output the hypothesis w and terminate the algorithm; else go to step D5.

D4. [Update hypothesis.] Else (if $Y < 0$) set $T \leftarrow 0$, $J \leftarrow \{j : w_j \neq u_j^{(t)}\}$ and pick a random index j from the uniform distribution on J. Set $w_j \leftarrow -w_j$ and leave the other components of w unchanged.

D5. [Increment time and iterate.] Set $t \leftarrow t+1$ and go back to step D2.

By a consideration of the equilibrium probability distribution of the states of the finite Markov chain which represents the system we show that the algorithm has minimal space complexity $2n$ and exponential time complexity $\Omega\left(e^{0.139n}\right)$.[2]

Massive improvements in running time result if the algorithm is modified to run in batch mode. In a single bit update batch mode, Step D4 is replaced by

D4′ [Update hypothesis.] Else (if $Y < 0$) set $T \leftarrow 0$ and call an additional $m-1$ examples $u^{(t+1)}, \ldots, u^{(t+m-1)}$. Define the indicator functions

$$I_k^{(s)} = \begin{cases} 1 & \text{if } w_k \neq u_k^{(s)}, \\ 0 & \text{if } w_k = u_k^{(s)}, \end{cases}$$

and select the index j garnering the most votes: $j \leftarrow \arg\max_k \sum_{s=t}^{t+m-1} I_k^{(s)}$. Set $w_j \leftarrow -w_j$ and leave the other components of w unchanged. Set $t \leftarrow t+m-1$.

In a multiple bit update batch mode, Step D4 is replaced by

D4″ [Update hypothesis.] Else (if $Y < 0$) set $T \leftarrow 0$ and call an additional $m-1$ examples $u^{(t+1)}, \ldots, u^{(t+m-1)}$. Define the indicator functions

$$I_k^{(s)} = \begin{cases} 1 & \text{if } w_k \neq u_k^{(s)}, \\ 0 & \text{if } w_k = u_k^{(s)}. \end{cases}$$

Tally the votes $b_k = \sum_{s=t}^{t+m-1} I_k^{(s)}$ and order the indices such that $b_{j_1} \geq b_{j_2} \geq \cdots \geq b_{j_n}$. Set $w_j \leftarrow -w_j$ if $j \in \{j_1, \ldots, j_{\lfloor (1-Y)/2 \rfloor}\}$ and leave the other components of w unchanged. Set $t \leftarrow t+m-1$.

Relatively small batch sizes m are needed. We show that in a single bit update batch mode, a batch size of $m = \frac{1}{2}\pi n \log n$ reduces the time complexity of the algorithm to $O(n)$ while in a multiple bit update batch mode, a batch size of $m = \pi n \log n$ reduces the time complexity of the algorithm to $O(1)$, independent of n.

References

[1] S. S. Venkatesh, "Directed Drift: a new linear threshold algorithm for learning binary weights on-line," *J. Comp. Sys. Sciences*, vol. 46, pp. 198–217, 1993.

[1]This work was supported by the Air Force Office of Scientific Research under grants F49620-93-1-0120 and F49620–92–J–0344.

[2]We use the number of bits of buffer memory needed as a measure of space complexity and the expected mistake bound of the algorithm, i.e., the expected number of epochs when an example is misclassified by the current hypothesis, as a measure of the algorithm's time complexity.

PATTERN RECOGNITION VIA MATCH BETWEEN CODED PATTERNS AND FEATURE VECTORS

Luan L. Lee* and Blanca R. M. Sosa+
*DECOM/FEE/UNICAMP, C.P. 6101, 13081-970 Campinas, SP, Brazil
+CT/UFPa, C.P. 8619, 66075-900 Belém, PA, Brazil

Abstract — **This paper describes a novel approach for pattern recognition based on the matching between coded patterns to feature vectors. Our intent is to integrate three individual steps (data acquisition, feature extraction and decision making) of a pattern recognition problem and to solve them simultaneously as a unique problem. The proposed pattern recognition method was explicitly illustrated by a numerical character recognition problem.**

Coded patterns matched to feature vectors in a pattern recognition system is conceptually analogous with the matching between a group and a set of signal in a digital communication system. In order to get a set of signals matched to a group it is necessary to set up a correspondence between the linearity and the distance measure. Such arrangement allows us to replace the Hamming distance measure by the Euclidean distance measure [1]. Now we define formally the matching of a set of signals to a group (Definition 1) and the transitive group (Definition 2).

Definition 1 [1]: A signal set S is matched to a group G if there exists a mapping h from G onto S such that, for any g_1 and g_2 in G, $d(h(g_1), h(g_2)) = d(h(g_1^{-1} * g_2), h(e))$, where e denotes unit of G. A mapping h satisfying this condition will be called a matched mapping. Moreover, if h is one-to-one then its inverse, h^{-1}, will be called a matched labeling.

Definition 2 [1]: Let S be a set of signals and $f : S \rightarrow S$ be an isometry. If Δ is a group of transformations of S and s is an element of S, then *orbit* of s under Δ is the set $\Delta(s) = \{f(s) : f \geq \Delta\}$. The transformation group Δ is called transitive of $\Delta(s) = S$ for some $s \in S$ (therefore, for all $s \in S$).

Next we consider only the case of set of signals with order 2^n. The first Sylow's Theorem which guarantees the existence of a group of order 2^n is as follows.

First Sylow's Theorem [2]: Let G be a finite group of order $p^n m$, $n \geq 1$ and p does not divide m. Then, (1) G has a subgroup of order p^i for any integer i, $1 \leq i \leq n$; (2) Each p^i-order subgroup H of G is a normal group of order p^{i+1} for $1 \leq i \leq n$.

The existence of a subgroup of order 2^n allows us to form a group of 2^n orthogonal matrices which is capable of generating a transitive group. It is worth mentioning that the product between each element of the group of matrices and a signal vector (feature vector) results in a signal vector (feature vector) also.

Numerical pattern recognition: Each input pattern (numerical character) is an 4-by-8 pixel rectangle. Another way to look at each character pattern is that it consists of eight "2-by-2 primitive patterns". There are in total eight distinct primitive patterns as shown below:

00	00	01	01	11	11	10	10
00	11	01	10	11	00	10	01

Therefore, the procedure for numerical character recognition proposed here consists of primitive pattern recognition. Next, we map the primitive patterns into a binary linear code $Z = (000, 001, 011, 010, 111, 110, 100, 101)$. which, in its turn, is matched to a set of feature vectors, S. The set S represents the eight vertices of a cube, namely $s_1 = (1,1,1)$, $s_2 = (1,1,-1)$, $s_3 = (1,-1,1)$, $s_4 = (1,-1,-1)$, $s_5 = (-1,1,1)$, $s_6 = (-1,1,-1)$ $s_7 = (-1,-1,1)$ and $s_8 = (-1,-1,-1)$.

From those eight signal vectors of S, we can easily find a group of orthogonal matrices B_i which is a transitive group. Signal s_j and orthogonal matrix B_i are related by the transformation $T_{B_i} : s_j \rightarrow B_i s_j$. Note that T_{B_i} transforms a signal vector into another signal vector. Solving the set of transformation T_{B_i} results in eight orthogonal diagonal matrices. The set of matrices $\{B_i\}$ forms a non-cyclic commutative group under the matrix operations. Moreover, these matrices define a transitive group of orthogonal transformations.

It can be shown that there is one-to-one correspondence between elements of sets Z and S. We represent this one-to-one correspondence as $z_i \longleftrightarrow B_i$, which implies the existence of an isomorphism between groups $(Z, \oplus)$ and $(B, .)$, denoted by $\varphi : (Z, \oplus) \rightarrow (B, .)$. It can be easily shown that φ is bijective, and for any z_1, $z_2 \in Z$, $\varphi(z_1 \oplus z_2) = \varphi(z_1).\varphi(z_2)$ Therefore, Z and B are isomorphic.

Noting that the matching between Z and S is a consequence of the isomorphism between Z and B. Define mapping $h : Z \rightarrow S$ as $z_i \rightarrow h(z_i) = T_{B_i(s_j)} = B_i s_j$. The Euclidean distance between any two elements of Z satisfies the relation $d(h(z_i), h(z_j)) = d(h(z_i^{-1} * z_j), h(e))$

It turns out that the primitive pattern classification consists of identifying of feature vectors in the feature space (signal space). Such a geometric property of feature vectors makes the decision procedure simple and straightforward because the decision regions are symmetric.

References

[1] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Trans. Information Theory*, Vol. 37, pp. 1675-1682, 1991.

[2] J.B. Fraleigh, A first Course in Abstract Algebra, Addision Wesley, 1982.

Training Recurrent Networks Using Hessian Information

Pedro Henrique Gouvêa Coelho
DEEE - UFMA - Campus do Bacanga - São Luís - MA - Brazil
e_mail:.coelho@calhau.fapema.br

Abstract - **New training algorithms for fully recurrent neural networks are presented. They are based on Hessian matrices estimates. Simulation results show that the algorithms yields satisfactory results.**

I. Introduction

Recurrent neural networks, having every unit connected to every other unit, are the most general case of neural networks and are highly non-linear dynamical systems exhibiting a rich and dynamical behavior. The architecture is inherently dynamic and usually one-layered, since its complex dynamics confer it powerful representation capabilities. Recurrent networks with the same structure can display different dynamic behavior, as a result of the use of diverse learning algorithms. The network is determined when its structure and learning rule are given, as the network is truly a composition of two dynamical systems: transmission and adjusting systems. The total input-output behavior is therefore a result of the interaction of both. Hence, the importance of learning rules in recurrent neural networks is readily understood . Learning algorithms used for recurrent networks are usually based on the computation of the gradient of a cost function with respect to the weights of the network. There are few learning algorithms applicable to general recurrent neural networks architectures and the most representative is the so called RTRL (Real Time Recurrent Learning) algorithm [1] (Williams and Zipser). This algorithm is truly on line and is a gradient descent type although more computationally expensive than other recurrent neural network algorithms (e.g. the backpropagation through time). However this undesired feature can be compensated by the fact that general fully recurrent architectures usually use far fewer neurons than backpropagation structures.

This paper proposes two new algorithms for fully recurrent neural networks using Hessian information (second derivatives of the cost function with respect to the parameters). The algorithms use estimates to the Hessian matrix with different degrees of computational complexities. Both algorithms use a matrix that is computed recursively on line with elements based on the sensitivity parameter as defined by Williams and Zipser [1]. The second algorithm uses a less computing demand estimate based on a diagonal matrix approximation for the Hessian matrix inspired on the Hessian matrix of the first algorithm. The idea of using a diagonal matrix approximation for the Hessian matrix is not new and was pursued by Becker and Le Cun in a backpropagation feedforward architecture [2]. These methods are known as pseudo-Newton algorithms and have the advantage of faster learning capabilities. They re-scale the learning rate of each weight dynamically to match the curvature of the cost function with respect to that weight.

II. Conclusions

Experiments were done to compare the proposed learning algorithms with existing ones (e.g. RTRL and pseudo-Newton) in the presence of noise. The new algorithms had shorter learning periods, and the first proposed one was the faster, at a cost of a higher computational complexity. The first proposed algorithm can still be an attractive alternative because its high computing demands can be compensated by the use of very small fully connected neural networks. There are some engineering applications that may use as few as two or three fully connected neurons. The proposed algorithm is being used to neural channel equalizers by the author. The availability of prior information could reduce the computing demands of on-line learning methods for recurrent neural networks. Alternatives in this direction are being studied by the author to continue or improve the algorithms.

References

[1] - R.J. Williams and D. Zipper "A learning algorithm for continuosly running fully recurrent neural networks", *Neural Computation*, 1, 1989, pp. 270-280.

[2] - S. Becker and Y. Le Cun, "Improving the convergence of backpropagation learning with second order methods", Proceedings of the 1988 Connectionist Models Summer School, Touretzky, Hinton, and Sejnowski, Eds. , San Matteo, CA: Morgan Kaufmann, pp. 29-37.

An Artificial Neural Net Viterbi Decoder

Xiao-an Wang and Stephen B. Wicker
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

Artificial neural networks (ANN's) have been successfully applied in the fields of signal processing and pattern recognition. In recent years, efforts have been made to design ANN decoders for error control codes. Although the general decoding problem can be viewed as a form of pattern recognition (PR), the information capacity in an error control code is far more extensive than that contained in most PR problems. Because of this, neural net training, a popular design tool for ANN, has not fared well in ANN decoders. So far, trained ANN decoders are limited to very small codes like the (7,4) Hamming code and convolutional codes with no more than 2 memory elements. Meanwhile, algebraic structures of the error control codes are not efficiently used in trained ANN decoders, resulting in inferior performance relative to that of the conventional decoders. For these reasons, the design of ANN decoders has become a process of "neuralizing" the existing digital decoding algorithms which have themselves been derived by fully exploiting the algebraic properties of the codes. The decoding process can be maximally parallelized by neural nets, which greatly increases the decoder throughput. Such ANN decoders have been successfully designed for many important block codes, such as Hamming codes, the (24,12) Golay code and the (32,16) QR code [1].

This paper presents an ANN Viterbi decoder for convolutional codes. In the past, Viterbi decoders have always been implemented using digital circuits. The speed of these digital decoders is directly related to the amount of parallelism in the design. As the constraint length of the code increases, parallelism becomes problematic due to the complexity of the decoder. In this work it is shown that the register exchange type [2] of VA can be completely represented by an ANN structure. However, for large decoding depth Γ, the required dynamic range goes far beyond what an analog neuron can provide. Since the register exchange operation is digital in essence, it is natural to adopt a hybrid design, which is shown in Figure 1 for a standard rate-1/2 code with 2 memory elements.

The analog part of the design implements the input correlation and path selection, as well as a scaling algorithm to keep each neuron holding the partial metric from saturating. The inputs to the decoder are r_0 and r_1 from the binary signalling AWGN channel. All connection gains are +1 unless marked otherwise. The synchronization circuit is not shown in the figure to preserve clarity, The structure in Figure 1 can be easily extended to rate-k/n convolutional codes with M memory elements.

The complexity of a locally connected neural network is characterized by the number of neurons, N. In general N is found to be

$$N = 2^M(2^{k+2} - 2) + 2^n + 1$$

which gives $N = 389$ for a rate-1/2 code with $M = 6$. The neurons can be realized using operational amplifiers (Op-Amps). If each Op-Amp contains 20 transistors, the network will have less than 8,000 transistors. On the other hand, a fully digital implementation for the same code requires 50,000 transistors just for ACS operations [3]. Some other advantages of the ANN Viterbi decoder are:

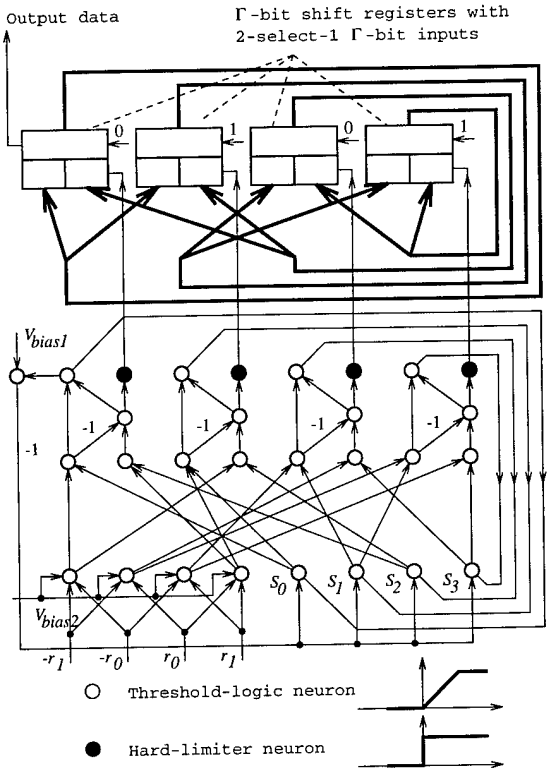


Fig. 1: The ANN Viterbi decoder

- The operations of the ANN decoder are fully parallel.
- All connections have unit gain, which eliminates weight considerations in VLSI implementation.
- The network is only locally connected.
- The characteristics of neurons are very simple to implement.
- The modularity brought by the hybrid design allows further improvements by using more sophisticated memory management techniques.

U.S. and foreign patents are currently pending.

References

[1] S. Ma, H. Zhou, C. S. Chen and J. Yuan, "Decoding complement-closed codes with coordinate-disjoint decomposition," *IEE Proceedings, Part I: Communications, Speech and Vision*, vol. 139, No. 5, pp. 488-494, October 1992.

[2] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, Englewood Cliffs: Prentice Hall, 1995.

[3] J. Sparso, H. N. Jorgensen, E. Paaske, S. Pedersen and T. Rubner-Petersen, "An area-efficient topology for VLSI Viterbi decoders and other shuffle-exchange type structures", *IEEE Journal of Solid-State Circuits*, Vol. 26, No. 2, pp. 90-97, February, 1991.

Combining Neural Network Classification with Fuzzy Vector Quantization and Hidden Markov Models for Robust Isolated Word Speech Recognition

Professor C S Xydeas and Lin Cong

Speech Processing Research Laboratory, Electrical Engineering Division
School of Engineering, University of Manchester Dover Street, Manchester, M13 9PL, UK

Abstract - **This paper proposes a new robust hybrid isolated word speech recognition system which is based on the improved quantization accuracy of FVQ, the strength of HMM in modelling stochastic sequences, and the non-linear classification capability of MLP neural networks. Thus the proposed FVQ/HMM/MLP approach combines effectively the relative contributions of codebook - dependent Fuzzy distortion measures with model - dependent maximum likelihood probability information. Computer simulation results clearly indicate the superiority in recognition accuracy performance of the FVQ/HMM/MLP approach when compared to that obtained from FVQ/HMM or FVQ/MLP schemes.**

I. INTRODUCTION

The system employs N FVQ codebooks and N HMM models. Given an input word, each FVQ codebook produces effectively an associated distortion measure d(O, W_j). In addition, an FVQ output vector is interpreted as a probability mass vector which is accepted by the associated HMM process to yield a maximum likelihood probability P(O/W_j). The above measures can be used directly as inputs to an MLP classifier or can be combined to form a hybrid measure which is then presented to the MLP network. In our noisy speech recognition study the systems under examination are trained on clean speech. Recognition performance is then measured with the input signal corrupted by vehicle or white acoustic noise at different Signal to Noise Ratio (SNR) values.

II. SYSTEM DESCRIPTION

The FVQ/HMM/MLP algorithm employs N VQ codebooks and N HMMs. Each input set of LSP coefficients is then Fuzzy Vector Quantised by C - entries codebooks CB_j j = 1, 2, ..., N. Thus an input word W_j represented by a series $\{x_1, x_2, \ldots, x_{T_j}\}$ of T_j LSP vectors, is vector quantised in "parallel" by N codebooks and a Fuzzy Distortion Measure FD_j[1] is obtained from each VQ process applied to the input word. At the same time, the N parallel codebooks yield N observation sequences which drive N corresponding HMM processes, HMM_j j = 1, 2, ..., N. Thus a maximum likelihood probability P(O/W_j) is obtained from each HMM process in response to an input word. The FD_j and P(O/W_j) measures can be combined to a simple measure [1] and then presented to the MLP network whose output OUT(j) j = 1, 2, ..., N assumes values in the range $0 \le OUT(j) \le 1$. The system selects the unknown input word W_j to be the jth vocabulary word if OUT(j) = max[OUT(j)], j = 1, 2, ..., N. The three layer network used employs P hidden nodes and N input nodes.

Alternatively, the N FD_j and N P(O/W_j) measures can be used as inputs to an MLP classifier having 2N inputs and N outputs. Thus the relative contribution of the above two similarity measures, towards a correct classification, is determined by the neural network. This flexible and powerful method, for "fusing" the output of the FVQ and HMM parts of the system, has been used in the computer simulation experiments discussed in the next section.

III. EXPERIMENTS AND RESULTS

The performance of the proposed FVQ/HMM/MLP scheme has been evaluated, and compared with that obtained from conventional FVQ/MLP [2] and FVQ/HMM systems [1]. Two sets of input words were used in these experiments: set one is based on the ten English digit words, zero to nine, whereas set two employs the 26 English letters. Figure 1 shows the performance of the FVQ/HMM/MLP, FVQ/MLP and FVQ/HMM systems operating on the second set of input words, for different input SNR values, when speech is corrupted by car noise.

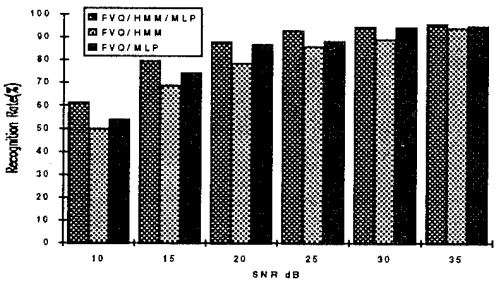


Fig. 1. Recognition performance of the FVQ/HMM, FVQ/MLP and FVQ/HMM/MLP for the car noise (N = 26).

IV. CONCLUSIONS

The proposed speech recognition system provides a superior performance, under noisy input conditions, when compared to conventional schemes [1], [2]. The system achieves a recognition rate of 98.33% and 90% at 30 dB and 20 dB SNR values respectively, when operating on set one of input words.

REFERENCES

[1] L Cong, C S Xydeas and A F Erwood: A Study of Robust Isolated Word Recognition Based on Fuzzy Methods, EUSIPCO - 94, UK

[2] L Cong, C Xydeas and A Erwood: A Fuzzy Vector Quantization and Neural Network Classification for Robust Isolated Word Speech Recognition, ICCS'94, Singapore

An EM-Based Algorithm For Recurrent Neural Networks

Sheng Ma and Chuanyi Ji
Department of Electrical, Computer and System Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180

Abstract — **A stochastic model is established for fully-connected recurrent neural networks with sigmoid units based on Gibbs distributions. EM (Expectation-Maximization) algorithm with a mean field approximation is then applied to train recurrent networks through hidden state estimation. The resulting EM-based algorithm, which reduces training the original recurrent network to training a set of individual feedforward neurons, simplifies the original training process and reduces the training time.**

I. Introduction

The goal of this work is in two-fold. First, we would like to develop a stochastic model to train recurrent networks with sigmoid units. Second, through the model developed, we will derive a novel training algorithm for recurrent networks which drastically simplifies the original training process.

II. A Stochastic Model

Consider a recurrent network with d inputs, L hidden units and one linear output unit. Let $x(n) \in R^d$, $t(n+1) \in R^1$ and $z(n+1) \in R^L$ be an input, a desired output of a recurrent network and a desired output of hidden units (also called desired hidden states) at n-th (and $n+1$-th) epoch, respectively. Let $\{x\}$, $\{z\}$ and $\{t\}$ denote all the inputs, desired hidden states and outputs up to epoch N. Let Θ contains all the parameters of the recurrent network: $w^{(1)}$, $w^{(3)}$ and $w^{(2)}$, the weights between inputs and hidden units, between hidden units, and between hidden and output units, respectively.

A stochastic model can then be established through a conditional probability model based on the Markov property of recurrent networks [1] i.e. $P(\{z\} \mid \{t\},\{x\},z(1),\Theta) = \prod_{n=1}^{N} P(z(n+1) \mid z(n),t(n+1),x(n),\Theta)$ and $P(\{z\},\{t\} \mid \{x\},z(1),\Theta) = \prod_{n=1}^{N} P(z(n+1),t(n+1) \mid z(n),x(n),\Theta)$, where $z(1)$ is the initial desired hidden states. Furthermore, Gibbs distributions can be used which eventually lead to the following probabilities $P(z(n+1) \mid z(n),x(n),\Theta) = A\exp(-\frac{1}{2}(z(n+1)-\hat{z}(n+1))^T\Sigma^{-1}(z(n+1)-\hat{z}(n+1)))$ and $P(z(n+1),t(n+1) \mid z(n),x(n),\Theta) = B\exp(-\lambda_1 E_1(n+1) - \lambda_2 E_2(n+1))$, where $E_1(n+1) = \| z(n+1) - h(n+1) \|^2$ and $E_2(n+1) = (t(n+1) - z(n+1)^T w^{(2)})^2$. λ_1, λ_2, A and B are constants. $\hat{z}(n+1)$ is the expectation of $z(n+1)$ at $n+1$-th epoch when given $z(n)$,$x(n)$ and Θ. $h(n+1)$ is the actual hidden output.

III. EM Algorithm For Recurrent Networks

Once the stochastic model is developed, a new training algorithm is derived through EM algorithm[2]. The essence of EM algorithm is that certain hidden variables (missing data) can be introduced to simplify a maximum likelihood problem. For our case, the hidden variables are desired hidden states $z(n+1)$'s which serve as missing data, whereas the incomplete data consists of pairs of $\{x(n),t(n+1)\}$'s. Using similar derivations as in [3], we can obtain the expected *log* likelihood $Q(\Theta \mid \Theta^p) = \int_{\{z\}} P(\{z\} \mid \{t\},\{x\},\Theta^p) \ln P(\{z\} \mid \{t\},\{x\},\Theta)$, where Θ and Θ^p are the new parameters and the parameters at the previous step, respectively.

Since $Q(\Theta \mid \Theta^p)$ is difficult to evaluate directly, a mean-field approximation[4] is used to estimate $Q(\Theta \mid \Theta^p)$, which eventually leads to an EM-based algorithm for recurrent networks.

E-step: Evaluate the expected desired hidden states recurssively through $\mathbf{E}z_j(n+1) = \hat{h}_j(n+1) + ae(n+1)$, where $\hat{h}_j(n+1) = g(x(n)^T w_j^{(1)} + \mathbf{E}z(n)^T w_j^{(3)})$ for $1 \le j \le L$. $g(u)$ is a sigmoid function. $e(n+1) = t(n+1) - \hat{h}(n+1)^T w^{(2)}$, and a is a constant.

M-step: Using the expected desired hidden states obtained at the E-step as targets for recurrent hidden neurons to find new parameters through two steps.

(a) Find $w_j^{(1)}$'s and $w_j^{(3)}$'s through minimizing the difference between expected desired and "actual" hidden states $\hat{h}(n)$: $\sum_n \| \mathbf{E}z(n) - \hat{h}(n) \|^2$.

(b) Find $w^{(2)}$ through minimizing difference between desired outputs of the network and weighted expected desired hidden targets: $\sum_n (\mathbf{E}z(n+1)^T w^{(2)} - t(n+1))^2$.

The algorithm will iterate between the E- and M-steps until a convergence criterion is achieved.

Notice that (a) and (b) are equivalent to training individual feedforward neurons and can be solved using a fast training algorithm given in [1].

IV. Simulation Results

Learning a teacher recurrent network is chosen as an initial test problem. When RTRL(back-propagation algorithm for recurrent nets), BPTT (back-propagation through time) and our algorithm were required to achieve a similar mean-square-error, our algorithm can be at least 10 times faster.

V. Acknowledgements

The support from National Science Foundation (ECS-9312505) is gratefully acknowledged.

References

[1] Breiman, L.E and Friedman, J.H, "Function Approximation Using Ramps" *Neural Networks for computing, Snowbird, Utah*, 1993.

[2] Dempster, A.P, Laird, N.M and Rubin, D.B "Maximum Likelihood from Incomplete Data via EM Algorithm," *J. of Royal Statistical Society*, B39, 1-33,1977.

[3] Jordan, M. and Jacobs, R.A., "Hierarchical Mixture of Experts," *Neural Computation*, vol. 6, pp 181-214, 1994.

[4] Zhang, J. and J. Modestino "The Mean-field Theory in EM Procedures for Markov Random Fields," *International Symposium of Information Theory*, 1991.

[1] This property comes from the fact that the outputs of a recurrent network and its hidden units at current epoch only depend on the actual outputs of hidden units at previous epoch.

Sufficient Conditions for Norm Convergence of the EM Algorithm[1]

Alfred Hero and Jeffrey Fessler
Dept. of EECS, The University of Michigan, Ann Arbor, Michigan, USA

Abstract — **In this paper we provide sufficient conditions for convergence of a general class of alternating estimation-maximization (EM) type continuous-parameter estimation algorithms with respect to a given norm.**

I. Introduction

Let $\theta = [\theta_1, \ldots, \theta_p]^T$ be a real parameter residing in an open subset Θ of the p-dimensional space $I\!R^p$. Given a general function $Q : \Theta \times \Theta \rightarrow I\!R$ and an initial point $\theta^0 \in \Theta$, consider the following recursive algorithm, called the A-algorithm:

$$\textbf{A-algorithm:} \qquad \theta^{i+1} = \text{argmax}_{\theta \in \Theta} Q(\theta, \theta^i). \tag{1}$$

If there are multiple maxima, then θ^{i+1} can be taken to be any one of them. Let $\theta^* \in \Theta$ be a fixed point of (1).

The A-algorithm contains a large number of popular iterative estimation algorithms such as: ML-EM algorithms (e.g. Dempster, Laird, and Rubin (1977), the penalized EM algorithm (e.g. Hebert and Leahy (1989)) , and EM-type algorithms implemented with E-step or M-step approximations (e.g., Antoniadis and Hero (1994), Green (1990)).

II. Convergence Theorem

A *region of monotone convergence* relative to the vector norm $\|\cdot\|$ of the A-algorithm (1) is defined as any open ball $B(\theta^*, \delta) = \{\theta : \|\theta - \theta^*\| < \delta\}$ centered at $\theta = \theta^*$ with radius $\delta > 0$ such that if the initial point θ^0 is in this region then $\|\theta^i - \theta^*\|$, $i = 1, 2, \ldots$, converges monotonically to zero. Note that as defined, the shape in $I\!R^p$ of the region of monotone convergence depends on the norm used. However in $I\!R^p$ monotone convergence in a given norm implies convergence, however possibly non-monotone, in any other norm.

Define the $p \times p$ matrices obtained by averaging $\nabla^{20} Q(u, \overline{u})$ and $\nabla^{11} Q(u, \overline{u})$ over the line segments $u \in \overrightarrow{\theta\theta^*}$ and $\overline{u} \in \overrightarrow{\overline{\theta}\theta^*}$:

$$A_1(\theta, \overline{\theta}) = -\int_0^1 \nabla^{20} Q(t\theta + (1-t)\theta^*, t\overline{\theta} + (1-t)\theta^*) dt$$

$$A_2(\theta, \overline{\theta}) = \int_0^1 \nabla^{11} Q(t\theta + (1-t)\theta^*, t\overline{\theta} + (1-t)\theta^*) dt.$$

Also, define the following set:

$$\mathcal{S}(\overline{\theta}) = \{\theta \in \Theta \ : \ Q(\theta, \overline{\theta}) \geq Q(\overline{\theta}, \overline{\theta})\}.$$

By construction of the A-algorithm (1), we have $\theta^{i+1} \in \mathcal{S}(\theta^i)$.

Definition 1 *For a given vector norm $\|\cdot\|$ and induced matrix norm $|\!|\!| \cdot |\!|\!|$ define $\mathcal{R}_+ \subset \Theta$ as the largest open ball $B(\theta^*, \delta) = \{\theta : \|\theta - \theta^*\| < \delta\}$ such that for each $\overline{\theta} \in B(\theta^*, \delta)$:*

$$A_1(\theta, \overline{\theta}) > 0, \qquad \textit{for all } \theta \in \mathcal{S}(\overline{\theta}) \tag{2}$$

and for some $0 \leq \alpha < 1$

$$\left|\!\left|\!\left| \left[A_1(\theta, \overline{\theta})\right]^{-1} \cdot A_2(\theta, \overline{\theta}) \right|\!\right|\!\right| \leq \alpha, \qquad \textit{for all } \theta \in \mathcal{S}(\overline{\theta}). \tag{3}$$

[1]This work was partially supported by grants: NSF-BCS-9024370, DOE-FG02-87ER60561, NIH-CA-60711

The following convergence theorem establishes that, if $\mathcal{R}_+$ is not empty, the region in Definition 1 is a region of monotone convergence in the norm $\|\cdot\|$ for an algorithm of the form (1). It can be shown that $\mathcal{R}_+$ is non-empty for sufficiently regular problems (Hero and Fessler (1995)).

Theorem 1 *Let $\theta^* \in \Theta$ be a fixed point of the A algorithm (1), where $\theta^{i+1} = argmax_{\theta \in \Theta} Q(\theta, \theta^i)$, $i = 0, 1, \ldots$. Assume: i) for all $\overline{\theta} \in \Theta$, the maximum $\max_\theta Q(\theta, \overline{\theta})$ is achieved on the interior of the set Θ; ii) $Q(\theta, \overline{\theta})$ is twice continuously differentiable in $\theta \in \Theta$ and $\overline{\theta} \in \Theta$, and iii) the A-algorithm (1) is initialized at a point $\theta^0 \in \mathcal{R}_+$ for a norm $\|\cdot\|$.*

1. *The iterates θ^i, $i = 0, 1, \ldots$ all lie in $\mathcal{R}_+$,*
2. *the successive differences $\Delta\theta^i = \theta^i - \theta^*$ of the A algorithm obey the recursion:*

$$\Delta\theta^{i+1} = [A_1(\theta^{i+1}, \theta^i)]^{-1} A_2(\theta^{i+1}, \theta^i) \cdot \Delta\theta^i, \tag{4}$$

3. *the norm $\|\Delta\theta^i\|$ converges monotonically to zero with at least linear rate, and*
4. *$\Delta\theta^i$ asymptotically converges to zero with root convergence factor*

$$\rho\left(\left[-\nabla^{20} Q(\theta^*, \theta^*)\right]^{-1} \nabla^{11} Q(\theta^*, \theta^*)\right) < 1.$$

III. Tomography Application

In emission computed tomography the objective is to estimate the object intensity vector $\theta = [\theta_1, \ldots, \theta_p]^T$, $\theta_b \geq 0$, from Poisson distributed projection data $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_m]^T$. The Shepp-Vardi implementation of the ML-EM algorithm for estimating the intensity θ has the form:

$$\theta_b^{i+1} = \frac{\theta_b^i}{P_b} \sum_{d=1}^{m} \frac{\mathbf{Y}_d P_{d|b}}{\sum_{j=1}^{p} P_{d|j} \theta_j^i}, \qquad b = 1, \ldots, p, \tag{5}$$

where $P_{d|b}$ is a full rank matrix of transition probabilities from emission locations to projection locations and $P_b = \sum_{d=1}^{m} P_{d|b}$.

Using Theorem 1 we obtain

Theorem 2 *Assume that the unpenalized ECT EM algorithm specified by (5) converges to the strictly positive limit θ^*. Then, for some sufficiently large positive integer M:*

$$\|\ln \theta^{i+1} - \ln \theta^*\| \leq \alpha \|\ln \theta^i - \ln \theta^*\|, \qquad i \geq M,$$

where $\alpha = \rho([\boldsymbol{B} + \boldsymbol{C}]^{-1}\boldsymbol{C})$, $\boldsymbol{B} = \boldsymbol{B}(\theta^*)$, $\boldsymbol{C} = \boldsymbol{C}(\theta^*)$, the norm $\|\bullet\|$ is defined as:

$$\|u\|^2 \stackrel{def}{=} \sum_{b=1}^{p} P_b \theta_b^* \, u_b^2. \tag{6}$$

Lange and Carson (1984) showed that the ECT EM algorithm converges to the maximum likelihood estimate. As long as θ^* is strictly positive, Theorem 2 asserts that in the final iterations of the algorithm the logarithmic differences $\ln \theta^i - \ln \theta^*$ converge monotonically to zero relative to the norm (6).

Deterministic EM Algorithms with Penalties

Joseph A. O'Sullivan and Donald L. Snyder
Department of Electrical Engineering, Campus Box 1127, Washington University, St. Louis, MO 63130
(314) 935-4173, e-mail: *jao@ee.wustl.edu*

I. Introduction

Csiszár [1] presented an axiomatic derivation of the use of the I-divergence as a discrepancy measure between nonnegative vectors. Snyder, Schulz, and O'Sullivan [2] then proposed the use of the I-divergence as an optimality criterion in deblurring problems, and introduced the deterministic version of the EM algorithm. Byrne [3] used a similar scenario to [2], but also looked at reverse entropy measures and included maximum entropy penalties. O'Sullivan [4] introduced roughness penalties for use in stochastic problems where the use of Markov random fields may not arise naturally; these penalties are used here for the deterministic problem. Vardi [5,6] has related papers.

Let $\theta \in \mathbf{R}^p$ be a vector of parameters to be estimated. The available data are $y_m = \sum_{n=1}^{N} H_{mn} x_n$, $1 \le m \le M$, where $\mathbf{y} \in \mathbf{R}_+^M$, $H_{mn} \ge 0$, and $\mathbf{x} \in \mathbf{R}_+^N$ depends on θ. The manner in which $\mathbf{x}$ depends on θ yields slightly different algorithms. The matrix $\mathbf{H}$ is assumed to have at least one positive entry in each column. We show that $\mathbf{x}$ may be considered to be the complete data for θ. The incomplete data I-divergence is shown to equal an averaged complete data I-divergence plus an additional term. This decomposition is a generalization of the decomposition for the stochastic data problem and it simplifies steps used to prove convergence in [2]. The deterministic EM algorithm then consists of minimizing the averaged complete data I-divergence; the averaging step corresponds to the E-step, the minimization is the M-step. Finally, a maximum entropy penalty and a roughness penalty are incorporated into the problem.

II. Deterministic EM Algorithm Derivation

The problem is to find the θ that minimizes $I(\mathbf{y}|\mathbf{H}\mathbf{x}(\theta))$, where $I(\mathbf{y}|\eta) = \sum_{m=1}^{M} [y_m \log \frac{y_m}{\eta_m} - y_m + \eta_m]$, and log means natural log. Let $\tilde{\mathbf{x}} \in \mathbf{R}_+^N$ and define a function of $\tilde{\mathbf{x}}$ by $g_{mn}(\tilde{\mathbf{x}}) = \frac{Y \tilde{x}_n H_{mn}}{\sum_{n'} H_{mn'} \tilde{x}_{n'}}$, where $Y = \sum_{m=1}^{M} y_m$. Also, denote by $\mathbf{h} \cdot \mathbf{x}$ the $N \times 1$ vector whose nth entry is $x_n \sum_{m=1}^{M} H_{mn}$.

Theorem 1:

$$I(\mathbf{y}|\mathbf{H}\mathbf{x}) = \sum_{m=1}^{M} \frac{y_m}{Y} [I_n(g_{mn}(\tilde{\mathbf{x}})|\mathbf{h} \cdot \mathbf{x}) - I_n(g_{mn}(\tilde{\mathbf{x}})|g_{mn}(\mathbf{x}))] + F(\mathbf{y}, \tilde{\mathbf{x}}),$$

where F does not depend on $\mathbf{x}$, and $\tilde{\mathbf{x}}$ is arbitrary.

The notation $I_n(\cdot|\cdot)$ indicates that the I-divergence is computed over the n index only. The vector $\mathbf{y}$ may be referred to as the incomplete data and $I(\mathbf{y}|\mathbf{H}\mathbf{x})$ is the incomplete data I-divergence. The theorem states the the incomplete data I-divergence equals the sum of three terms. The first is an averaged I-divergence involving the vector $\mathbf{h} \cdot \mathbf{x}$ and is called the complete data I-divergence; $\mathbf{x}$ is the complete data vector. The second term is an I-divergence that is used to guarantee monotonicity of the sequence of likelihood values. The third term is an extra term that does not depend on the complete data.

The deterministic EM algorithm then has the following steps at iteration $k+1$ given an estimate $\hat{\theta}^k$ and the corresponding $\mathbf{x}^k = \mathbf{x}(\theta^k)$ from iteration k:

E-Step: Compute $Q(\mathbf{x}|\mathbf{x}^k) = \sum_{m=1}^{M} \frac{y_m}{Y} I_n(g_{mn}(\mathbf{x}^k)|\mathbf{h} \cdot \mathbf{x})$

M-Step: Find $\theta^{k+1} = \text{argmin } Q(\mathbf{x}(\theta)|\mathbf{x}^k)$.

The objective function is nonincreasing since (using $\tilde{\mathbf{x}} = \mathbf{x}^k$)

$$I(\mathbf{y}|\mathbf{H}\mathbf{x}^{k+1}) - I(\mathbf{y}|\mathbf{H}\mathbf{x}^k) = \sum_{m=1}^{M} \frac{y_m}{Y} [I_n(g_{mn}(\mathbf{x}^k)|\mathbf{h} \cdot \mathbf{x}^{k+1}) - I_n(g_{mn}(\mathbf{x}^k)|\mathbf{h} \cdot \mathbf{x}^k) - I_n(g_{mn}(\mathbf{x}^k)|g_{mn}(\mathbf{x}^{k+1}))].$$

The sum of the first two terms in the bracket is nonpositive by the M-step, and the last term is nonpositive because the I-divergence is nonnegative. For discussions of convergence see [2,3,7]. If $\theta = \mathbf{x}$, the algorithm derived in [2] results,

$$\hat{x}_n^{k+1} = \frac{\hat{x}_n^k}{\sum_{m=1}^{M} H_{mn}} \sum_{m=1}^{M} \frac{y_m H_{mn}}{\sum_{n'=1}^{N} H_{mn'} \hat{x}_{n'}^k}.$$

Byrne [3] introduced maximum entropy penalties ($I(\mathbf{x}|\xi)$ or $I(\xi|\mathbf{x})$) into the deterministic problem; prior information is assembled into a prior guess ξ. O'Sullivan [4] introduced roughness penalties that penalize discrepancies with neighbors. Let $\{\mathbf{S}_i, 1 \le i \le I\}$ be a set of $N \times N$ permutation matrices. Then the roughness penalties are of the form $\sum_{i=1}^{I} I(\mathbf{x}|\mathbf{S}_i\mathbf{x})$. Furthermore, a generalized EM algorithm was introduced in [4] based on the resulting neighborhood structures. The minimum penalized I-divergence problem is to compute the vector $\mathbf{x}$ that minimizes $I(\mathbf{y}|\mathbf{H}\mathbf{x}) + \alpha I(\xi|\mathbf{x}) + \beta \sum_{i=1}^{I} I(\mathbf{x}|\mathbf{S}_i\mathbf{x})$. The GEM algorithm from [4] may be used, replacing the complete and incomplete data spaces for that stochastic problem by the corresponding spaces from this deterministic problem, to obtain a sequence of iterates that converges to the optimum.

References

1. I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *Annals Stat.*, vol. 14, no. 4, pp. 2032-2066, 1991.
2. D. L. Snyder, T. J. Schulz, and J. A. O'Sullivan, "Deblurring subject to nonnegativity constraints," *IEEE Trans. Signal Proc.*, vol. 40, no. 5, pp. 1143-1150, May 1992.
3. C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Proc.*, vol. 2, pp. 96-103, Jan. 1993.
4. J. A. O'Sullivan, "Roughness penalties on finite domains," *IEEE Trans. Image Proc.*, to appear, 1995.
5. Y. Vardi and D. Lee, "From image deblurring to optimal investments: maximum likelihood solution for positive linear inverse problems," (with discussion) *J. Royal Stat. Soc.*, Series B, vol. 4, no. 3, pp. 569-612, 1993.
6. Y. Vardi, "Applications of the EM algorithm to linear inverse problems with positivity constraints," to appear in a book on image analysis and speech recognition, L. A. Shepp and S. E. Levinson, Eds.
7. I. Csiszar and G. Tusnady, "Information geometry and alternating minimization procedures," *Statistics and decisions,* Supp. No. 1, pp. 205-237, 1984.

Hidden Markov Models Estimation via the Most Informative Stopping Times for Viterbi Algorithm

Joseph A. Kogan

Courant Institute of Mathematical Sciences, NYU, New York, NY 10012, USA, email: koganj@acf4.nuy.edu

Abstract — **We propose a sequential approach for studying the Viterbi algorithm via a renewal sequence of the most informative stopping times which allows us in particular to obtain new asymptotic "single-letter" decoding conditions of equivalency between the Baum-Welch, segmental K-means and vector quantization algorithms of the hidden Markov models parameters estimation.**

I. Introduction

Let $\{h_t\}$ be a finite hidden Markov chain (HMC) indirectly observed through a process $\{z_t\}$, $z_t \in R^D, t = 0,\dots,N$. Given a sequence of observations z_0^N and a set $\lambda = \{\pi_{h_0}, a_{h_{t-1}h_t}, b(z_t/h_t)\}$ of prior, transition and observation probabilities, respectively, the Viterbi algorithm (VA) allows us to find the most likely state-sequence (MLSS) $\hat{h}_0^N$ of the HMC via maximizing the next additive criterion $d_N(\hat{h}_0^N) = \max_{h_0^N} \ln P\{h_0^N, z_0^N\}$ by a dynamic programming (DP) method. Then the MLSS $\hat{h}_0^{N-1}$ or the optimal segmentation of the observations z_0^{N-1} can be obtained by the backtracking $t = N-1,\dots,0$: $\hat{h}_t \doteq k_{t+1}(h_{t+1})$, where $\hat{h}_N = \arg\max_{h_N} d_N(\hat{h}_0^{N-1}(h_N))$, (see, [1]-[3]).

The direct implementation of VA requires to store the values of $\hat{h}_t$ what fills up a table $\mathbf{K}(m \times N)$ with columns of back pointers $k_t : \hat{H}_t \longrightarrow \hat{H}_{t-1}, t = N, N-1,\dots$ with $\hat{H}_N = H = \{0,1,\dots,m-1\}$ but if for a some moment $s, \exists j \in H$: $k_{s+1}(\hat{H}_{s+1}) \equiv j$ for all $\hat{h}_t \in \hat{H}_t = H, t > s$, then $\hat{h}_s \equiv j$ is called **a special column (SC)** in the table $\mathbf{K}$ of optimal VA decisions [2] and the moments of the *SCs* appearing are **the most informative stopping times (MISTs)** for the Viterbi recognition of HMS [3], [4] because after their appearing further observations do not change the previous decisions of the VA.

II. Results

We establish the renewal properties of the MIST sequence and the duality between the Wald's sequential analysis and the VA which allow us to develop a sequential version of the segmental K-means algorithm for reducing the biases of estimates if the set of parameters λ is unknown.

Then we consider the asymptotic conditions of equivalency between the Baum-Welch (BW), segmental K-means (SKM) algorithms and vector quantization (VQ) approach which has important applications in speech recognition (see, [5], [6]). When the set of parameters $\{\lambda\}$ is unknown, it can be estimated, for instance, by the BW: $\lambda^* \doteq \arg\max_\lambda P_\lambda(Z)$, or by the SKM: $\hat{\lambda} \doteq \arg\max_\lambda \max_h P_\lambda(Z,h)$ algorithms what can be achieved by the following iterative maximizations respectively

$$\text{BW: } \lambda_{i+1} \doteq \arg\max_\lambda \textstyle\sum_h P_{\lambda_i}(h/Z) \ln P_\lambda(Z,h),$$
$$\text{SKM: } \lambda_{i+1} \doteq \arg\max_\lambda \textstyle\sum_h \delta(h - h(\lambda_i)) \ln P_\lambda(Z,h),$$

where $h = h_0^N$, $Z = z_0^N$ and $\delta(\cdot)$ is the Kronecker δ-function. Thus, if $P_{\lambda_i}(h/Z) \to \delta(h - h(\lambda_i)), \forall i$, then $h(\lambda_i)$ is a dominant MLSS for $\forall i$ and $BWA \sim SKMA$. A sufficient condition of the existence of such a dominant MLSS $h^*_{0,T} = h^*_0,\dots,h^*_T$, where $h^*_t = \arg\min_h D^{-1} \ln b(z_t/\theta_h)$, has been given in [5]:

$$-\lim_{D\to\infty} D^{-1} \ln b(z_t/\theta_{h_t}) = H_{h_t}, \qquad (1)$$

where H_{h_t} is a constant entropy depending on h_t and θ_{h_t} is a set of parameters. Furthermore, the probability of deviation from the MLSS given Z, decays uniformly exponentially and does not depend on the length N of the sequence h. Then from [3] one can have

Theorem 1 *Given (1), the dominant MLSS is asymptotically single-letter decoding, as $D \to \infty$.*

Thus, in the limit, the MISTs will appear at each step what coincides with the case of a generalized single-letter decoding [2]. For the Gaussian HMC with the autoregression covariance matrix associated with state h_t we can, by using the renewal properties of the MIST sequence, further strengthen the result of [5], [6] that as $D \to \infty$, the SKMA becomes equivalent to the VQ approach which in this case minimizes the Itakura-Saito distortion measure.

Theorem 2 *If $p_{ij} \geq \delta > 0$, for $\forall SCi$ (as $N \to \infty$), then*

$$\lim_{D\to\infty} \frac{1}{D} \sum_{n\in\{\tau_0,\tau_1\}} \ln P_\lambda(z_n, h_n) = -\sum_{n\in\{\tau_0,\tau_1\}} [H(z_n) + \frac{d_{IS}(z_n, \lambda_{h_n})}{2}],$$

where $H(z_n)$ is the empirical entropy of z_n, d_{IS} is the corresponding discrete Itakura-Saito distortion measure, τ_k is the kth moment of the SC appearing, and

$$\lim_{D\to\infty} \lim_{N\to\infty} \frac{\ln \max_h P_{\hat{\lambda}}(Z,h)}{DN} = -E_i[H(z_n) + \frac{d_{IS}(z_n, \hat{\lambda}_{h_n})}{2}].$$

References

[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition, "*Proceedings of the IEEE*, vol. 77, pp. 257-286, 1989.

[2] J. A. Kogan, "Optimal segmentation of structural experimental curves by the dynamic programming method," *Automation and Remote Control*, No. 7, pt. 2, pp. 934–942, 1988.

[3] J. A. Kogan, "Exact Viterbi recognition of hidden Markovian sequences via the most informative stopping times," submitted to *IMA Proceedings on "Image models (and thier speech models cousins)"*.

[4] J. A. Kogan, "The most informative stopping times for Viterbi algorithm: sequential properties," in *Proceedings 1994 IEEE-IMS Workshop on Information Theory and Statistics*, Alexandria, Virginia, Oct. 1994.

[5] N. Merhav, Y. Ephraim, "Hidden Markov modeling using a dominant state sequence with application to speech recognition," *Computer Speech and Language*, pp. 327–339, 1991.

[6] N. Merhav, Y. Ephraim, "Maximum likelihood hidden Markov modeling using a dominant sequence of states," *IEEE Trans., on ASSP*, vol. 39, No. 9, 1991.

Model Parameter Estimation for 2D Noncausal Gauss-Markov Random Fields

R. Cusani, E. Baccarelli, S. Galli

INFOCOM Dpt., University of Rome 'La Sapienza', Via Eudossiana 18, 00184 Roma, Italy

Abstract - **An original procedure for estimating the model parameters of a noncausal Gauss-Markov Random Field (GMRF) from noisy observations is proposed. Starting from a suitable 'local' representation of the field and taking into account the symmetry property of the so-called 'potential fields' [3] describing the GMRF, a linear equation system relating the model parameters to the (generally, non-stationary) 2D autocorrelation function (acf) of the observed field is derived. Its solution for a known (or estimated) acf directly gives the parameter estimates of the GMRF. The unkown variance of the eventually present observation noise can be also estimated jointly with the model parameters.**

SUMMARY

A discrete-index 2D zero-mean Gaussian random process $\{x(\underline{s}) \in R^1, \underline{s} \in I\}$ defined over a rectangular lattice I and constituting a noncausal d^{th}-order homogeneous GMRF with respect to (wrt) an assigned 'support region' (or 'neighbourhood system' [3,4]) $\eta(d)$ admits the 'innovations representation' [1,2]

$$x(\underline{s}) = \sum_{\underline{r} \in \eta(d)} \phi(\underline{r})\, x(\underline{s}+\underline{r}) + u(\underline{s}), \quad \underline{s} \in I^*(\eta(d)). \tag{1}$$

In (1) the set $\eta(d)$ is assumed symmetric and constituted by an even number of points 2L(d) [4]; $I^*(\eta(d))$ is the set of 'internal points' of I wrt $\eta(d)$; $\{\phi(\underline{r}) \in R^1, \underline{r} \in \eta(d)\}$ are the so-called 'field potentials', related as reported in [2] to the acf $\{R_u(\underline{r})\}$ of the 2D stationary zero-mean Gaussian 'innovations process' $\{u(\underline{s}) \in R^1, \underline{s} \in I^*(\eta(d))\}$, with variance K_u.

From the obvious symmetry property $R_u(\underline{r}) = R_u(-\underline{r})$ we have: $\phi(\underline{r}) = \phi(-\underline{r})$, $\underline{r} \in \eta(d)$. This allows to partition of the support region $\eta(d)$ in the sub-sets $\eta_+(d)$, $\eta_-(d) \subset \eta(d)$, each constituted by L(d) sites and such that if $\underline{r} \in \eta_+(d)$ then $-\underline{r} \in \eta_-(d)$ for every $\underline{r} \in \eta(d)$. In this way (1) can be rewritten as

$$x(\underline{s}) = \sum_{\underline{r} \in \eta_-(d)} \phi(\underline{r}) \left[x(\underline{s}+\underline{r}) + x(\underline{s}-\underline{r}) \right] + u(\underline{s}), \quad \underline{s} \in I^*(\eta(d)). \tag{2}$$

The representation of the GMRF in (2) is then completed by specifying the associated boundary conditions (b.c.), i.e. the statistics of the random vector constituted by the r.v.s extracted from the random field $\{x(\underline{s})\}$ at the boundary points of the lattice I.

It is also assumed that the GMRF is corrupted by a 2D stationary zero-mean additive white noise process $\{w(\underline{s}) \in R^1, \underline{s} \in I\}$ independent from $\{x(\underline{s})\}$ and with (unknown) variance σ_w^2, so that the resulting observation process $\{y(\underline{s}) \in R^1, \underline{s} \in I\}$ is defined as: $y(\underline{s}) = x(\underline{s}) + w(\underline{s})$.

An original procedure for estimating the model parameters of a GMRF of an arbitrary order can be obtained from the 'local' representation in (2). In fact, from the model in (2) the following set of linear algebraic equations can be built up:

$$R_Y(\underline{s}; \underline{s}+\underline{m}) = \sum_{\underline{r} \in \eta_-(d)} \phi(\underline{r}) \left[R_Y(\underline{s}+\underline{r}; \underline{s}+\underline{m}) + R_Y(\underline{s}-\underline{r}; \underline{s}+\underline{m}) \right] + \left[\sigma_w^2 + K_u \right] \delta(\underline{m}),$$
$$\underline{s} \in I^*(\eta(d)), \quad \underline{s}+\underline{m} \in I, \tag{3}$$

thus relating the unknown model parameters to the 2D acf $\{R_Y(\underline{s};\underline{t})\}$ of the process $\{Y(\underline{s})\}$; the acf is assumed 'a priori' known or estimated from the available observations ($\delta(\underline{m})$ in (3) is the Kronecker delta).

Writing (3) for a set of L(d) sites $\underline{s} \in I^*(\eta(d))$ suitably chosen and for $\underline{m} \notin \{\eta_-(d) \cup \{\underline{0}\}\}$ (so that $\delta(\underline{m})=0$), a matrix linear algebraic equation system is directly derived, and from its solution the field potentials $\{\phi(\underline{r})\}$ are obtained. Such a system can be considered as the extension to the case of 2D noncausal GMRFs of the so-called 'high-order' Yule-Walker equations for the parameter identification of 1D causal AR processes. From the field potentials, writing (3) for $\underline{s} \in I^*(\eta(d))$ and for any $\underline{m} \in \eta_-(d)$ such that $\phi(\underline{m}) \neq 0$, the noise variance σ_w^2 is then calculated; finally, the parameter K_u is computed from (3) written for $\underline{m} = \underline{0}$.

The illustrated parameter estimation procedure is *fully general*: in fact it is valid for GMRFs defined on both *finite or infinite lattices* and for *any kind* of assumed boundary conditions, periodic or non-periodic, their influence being embedded in the acf of the field itself. Moreover, it can be easily particularized to the case when the noise variance σ_w^2 is known, or when the observation noise is absent.

Comparing the proposed solution to alternative methods available in the literature, some improvements can be outlined. More in detail, having exploited the symmetry $\phi(\underline{r}) = \phi(-\underline{r})$ gives an algebraic system with *half size* with respect to the system in [1]. On the other hand, the procedure in [4] is based on an iterative search algorithm, thus giving a computational complexity proportional to the size of the field, while the proposed solution is based on a 'local' description of the GMRF so that it does *not* involve time-consuming iterative search algorithms and its complexity is *independent* from the the size of the field. Finally, in the proposed approach the variance of the observation noise is estimated *together* with the model parameters while the procedures in [1] and [4] requires that it is known (or separately estimated). The results of some computer simulations of the above procedure are reported in Tab.I and in [7].

REFERENCES

[1] A.K. Jain, "Advances in Mathematical Models for Image Processing", *Proc. IEEE*, vol.69, no.5, pp.502-528, May 1981.

[2] J.W. Woods, "Two-dimensional discrete Markovian fields", *IEEE Trans. IT*, vol.18, pp.232-240, March 1972.

[3] J.M.F. Moura, N. Balram, "Recursive structure of non-causal Gauss Markov fields", *IEEE Trans. IT*, vol.28, no.2, March 1992.

[4] N. Balram, J.M.F. Moura, "Noncausal Gauss Markov random fields: Parameter structure and estimation", *IEEE Trans. IT*, vol.39, no.4, pp.1333-1355, July 1993.

[5] R. Cusani, E. Baccarelli, G. Di Blasio, "Model Parameter Estimation for Reciprocal Gauss-Markov Random Processes", *IEEE Trans. on SP*, vol.43, no.3, pp.792-795, March 1995.

[6] E. Baccarelli, 'Bilateral Markovian Processes', Ph.D. dissertation, University of Rome, Italy, October 1992 (in italian).

[7] R. Cusani, E. Baccarelli, "Identification of 2D Noncausal Gauss-Markov Random Fields", submitted to *IEEE Trans. on SP*.

	True	Estimated	True	Estimated	True	Estimated
ϕ (-1, -1)	5.0 -2	4.27 -2	1.2 -1	1.10 -1	5.0 -2	5.75 -2
ϕ (-1, 0)	5.0 -2	4.89 -2	1.2 -1	1.17 -1	9.0 -2	8.80 -2
ϕ (-1, +1)	5.0 -2	6.92 -2	1.2 -1	1.21 -1	5.0 -2	5.93 -2
ϕ (0, +1)	5.0 -2	4.80 -2	1.2 -1	1.38 -1	9.0 -2	9.80 -2
Ku	10.0	10.01	10.0	10.23	10.0	9.815

Tab.I - True and estimated parameter-values for three cases of noisy-free second-order (i.e., $\sigma_w^2=0$, d=2) 2D GMRFs with pinned-to-zero boundary conditions. The field potentials $\{\phi(\underline{r})\}$ and K_u are calculated as in (3) by estimating the acfs $\{R_X(\underline{s};\underline{t})\}$ from 10^4 independent realizations.

Achievable Regions in the Bias-Variance Plane for Parametric Estimation Problems [1]

Alfred Hero* and Mohamed Usman**
*Dept. of EECS, The University of Michigan, Ann Arbor, Michigan, USA
**Punjab Institute of Computer Science, Lahore, Pakistan

Abstract — **In this paper we use the uniform Cramer-Rao (CR) lower bound [1] to generate bias-variance tradeoff curves which separate achievable from unachievable regions in the estimator bias variance plane.**

I. Introduction

Let $\underline{\theta} = [\theta_1, ..., \theta_n]^T \in \Theta$ be a vector of unknown parameters which parameterize the distribution of an observed random variable $\mathbf{Y}$. Let $\hat{\theta}_1 = \hat{\theta}_1(\mathbf{Y})$ be an estimator of the scalar θ_1 and define the estimator bias function $b_1 = b_1(\underline{\theta}) = E_{\underline{\theta}}[\hat{\theta}_1] - \theta_1$ and the variance function $\sigma^2 = \sigma^2(\underline{\theta}) = E_{\underline{\theta}}[(\hat{\theta}_1 - \theta_1)^2]$. The goal of this work is to quantify fundamental tradeoffs between the bias and variance functions for any parametric estimation problem. When considered as surfaces over the parameter space Θ, the bias and variance provide a very informative description of estimator performance, for example they jointly specify the MSE.However, since comparison of performance surfaces over a large set Θ is usually impractical, the bias and variance in a small neighborhood is of greater interest. In this case, the bias gradient $\nabla_{\underline{\theta}} b_1$ is more useful since it is insensitive to constant and hence removable biases. It can be shown that $\nabla_{\underline{\theta}} b_1$ is directly related to the width of the point spread function for penalized maximum likelihood deconvolution problems [2]. The weighted norm of the bias gradient is indirectly related to the variation of the bias function over Θ by: $|\Delta b_1(\underline{\theta})| \leq \|\nabla_{\underline{\theta}} b_1\|_D + o(det|D|)$, where $\|\underline{u}\|_D^2 = \underline{u}^T D^T D \underline{u}$ and D is an invertible matrix whose determinant is proportional to the volume of the region.

II. The Bias-Variance Tradeoff Curve

The tradeoff curve is derived from a generalization of the bound on estimator variance presented in [1]. Unlike the bound of [1], this bound applies to the case of singular Fisher information matrices (FIM), an important case arising in deconvolution problems, and permits use of any weighted l_2 norm of the bias gradient.

Theorem 1 *For a fixed scalar $\delta \in [0, 1]$ let $\hat{\theta}_1$ be an estimator whose bias gradient satisfies the norm constraint $\|\nabla_{\underline{\theta}} b_1\|_D^2 = \underline{u}^T D^T D \underline{u} \leq \delta^2$, where D is an arbitrary non-singular matrix. Define the oblique projection operator ($n \times n$ matrix) $\mathcal{P}_{F_Y} = F_Y[F_Y D^T D F_Y]^+ F_Y D^T D$ which maps n-dimensional space onto the column space of the FIM F_Y, and define the n-element unit vector $\underline{e}_1 = [1, 0, ..., 0]^T$. Then the variance of $\hat{\theta}_1$ satisfies:*

$$var_{\underline{\theta}}(\hat{\theta}_1) \geq B(\underline{\theta}, \delta), \tag{1}$$

where if $\|\mathcal{P}_{F_Y} \underline{e}_1\|_D \leq \delta$ then $B(\underline{\theta}, \delta) = 0$, while if $\|\mathcal{P}_{F_Y} \underline{e}_1\|_D > \delta$ then:

$$B(\underline{\theta}, \delta) = \underline{e}_1^T F_Y^+ \underline{e}_1 - \underline{e}_1^T \left[F_Y^+ (\lambda \cdot D^T D + F_Y^+)^{-1} F_Y^+\right] \underline{e}_1 - \lambda \delta^2 \tag{2}$$

In (2) $\lambda > 0$ is determined by the unique non-negative solution of the following equation:

$$g(\lambda) = \underline{e}_1^T \left[F_Y^+ \left(\lambda \cdot D^T D + F_Y\right)^{-2} F_Y^+\right] \underline{e}_1 = \delta^2. \tag{3}$$

By calculating the family of points $\{(B(\underline{\theta}, \delta), \delta) : \delta \in [0, 1]\}$ we sweep out a curve in the *bias-variance plane* which lower bounds any estimator plotted in the plane. Figure 1 illustrates this curve for a simple one dimensional Gaussian deconvolution problem and the unweighted l_2 norm (D=identity) [2].

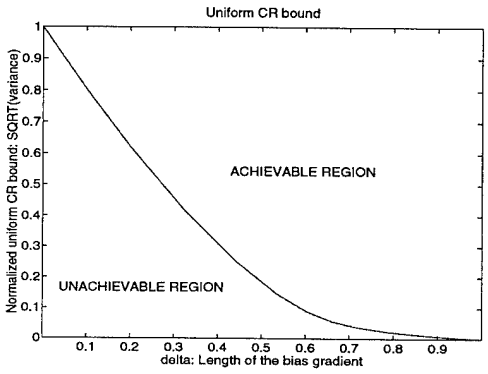


Figure 1: Bias-Variance Plane and Lower Bound.

The region above and including the curve is the so called 'achievable' region where all the realizable pairs of estimator variance and bias-gradients exist. Note that if an estimator lies on the curve then lower variance can only be bought at the price of increased bias and vice versa. For this example the regularized least squares estimator attains optimal bias-variance tradeoff, i.e. it hits the lower bound for all values of δ [2]. In this case the bias gradient norm δ was swept out by varying the smoothing (regularization) parameter of the estimator.

In general to place an estimator somewhere within the achievable region of Figure 1 requires the variance and length of the estimator bias gradient. In most cases the variance and the bias-gradient length are analytically intractable and must be empirically estimated. Since the sample mean estimate of the bias gradient norm has severe positive bias some form of bias correction is necessary. We have developed a bootstrap estimator and a $(1-\alpha)$% lower confidence bound for this purpose.

[1] This work was partially supported by NSF Grant BCS-9024370

References

[1] A.O. Hero, "A Cramer-Rao type lower bound for essentially unbiased parameter estimation," Technical Report TR-890, MIT Lincoln Laboratory, 1992.

[2] M. Usman and A.O. Hero, "Bias-variance tradeoffs for parametric estimation problems using the uniform CR bound," in revision for publication in *IEEE Trans. on Signal Processing/*.

New Spherical Designs in Three and Four Dimensions

R. H. Hardin and N. J. A. Sloane
Mathematical Sciences Research Center
AT&T Bell Laboratories
Murray Hill, NJ 07974 USA

Abstract — **A number of new spherical t-designs in three and four dimensions are described. Evidence is presented to suggest that in three dimensions the resulting catalog gives a complete list of all designs of strength $t \leq 9$.**

I. Introduction

A set of N points $\wp = \{P_1, \ldots, P_N\}$ on the unit sphere $\Omega_d = S^{d-1} = \{x = (x_1, \ldots, x_d) \in \mathbb{R}^d : x \cdot x = 1\}$ forms a *spherical t-design* if the identity

$$\int_{\Omega_d} f(x) d\mu(x) = \frac{1}{N} \sum_{i=1}^{N} f(P_i) \qquad (1)$$

(where μ is uniform measure on Ω_d normalized to have total measure 1) holds for all polynomials f of degree $\leq t$ ([3]; [4]; [2, §3.2]). In the present paper we are concerned only with the cases $d = 3$ and 4.

II. Spherical t-Designs in Three Dimensions

In three dimensions it is trivial that 1-designs exist if and only if $N \geq 2$, and Mimura [7] showed that 2-designs exist if and only if $N = 4, \geq 6$. Bajnok [1] found 3-designs for $N = 6, 8, \geq 10$ and conjectured that they do not exist for $N = 7$ and 9. In [5] we showed that 4-designs exist for $N = 12, 14, \geq 16$, and conjectured that no others exist. Reznick [8] showed that 5-designs exist for $N = 12, 16, 18, 20, 22, 24, \geq 26$. We have found 5-designs with $N = 23$ and 25, and, our search having repeatedly failed in the remaining cases, conjecture that 5-designs do not exist for $N = 13$–15, 17, 19 and 21.

Let $\tau(N)$ denote the largest value of t for which an N-point 3-dimensional spherical t-design exists. Since a t-design is also a t'-design for all $t' \leq t$, an N-point spherical t-design exists if and only if $\tau(N) \geq t$. Our results lead us to believe that the following are the values of $\tau(1), \ldots, \tau(50)$:

$$\begin{array}{l} 0,1,1,2,1,3,2,3,2,3, \\ 3,5,3,4,3,5,4,5,4,5, \\ 4,5,5,7,5,6,5,6,6,7, \\ 6,7,6,7,6,8,7,7,7,8, \\ 7,8,7,8,8,8,8,9,8,9 \end{array}$$

This is part of a much larger table that will appear in [6].

The results of this table then suggest that, in three dimensions, spherical 6-designs with N points exist for $N = 24, 26, \geq 28$; 7-designs for $N = 24, 30, 32, 34, \geq 36$; 8-designs for $N = 36, 40, 42, \geq 44$; 9-designs for $N = 48, 50, 52, \geq 54$; 10-designs for $N = 60, 62, \geq 64$; 11-designs for $N = 70, 72, \geq 74$; and 12-designs for $N = 84, \geq 86$. The existence of some of these designs is established analytically, while others are given by very accurate numerical coordinates.

The 24-point 7-design was first found by McLaren in 1963, and — although not identified as such by McLaren — consists of the vertices of an "improved" snub cube, obtained from Archimedes' regular snub cube (which is only a 3-design) by slightly shrinking each square face and expanding each triangular face.

One of our constructions gives a sequence of putative spherical t-designs in three dimensions with $N = 12m$ points ($m \geq 2$) where $N = \frac{1}{2}t^2(1 + o(1))$ as $t \to \infty$.

III. Spherical t-Designs in Four Dimensions

Analogous results have been obtained in four dimensions and will be described if time permits.

References

[1] B. Bajnok, paper presented at Amer. Math. Soc. meeting, College Station, Texas, Oct. 22, 1993.

[2] J. H. Conway and N. J. A. Sloane, *Sphere Packing, Lattices and Groups*, Springer-Verlag, NY, 2nd ed., 1993.

[3] P. Delsarte, J.-M. Goethals and J. J. Seidel, Spherical codes and designs, *Geom. Dedicata*, **6** (1977), 363–388.

[4] J.-M. Goethals and J. J. Seidel, Cubature formulae, polytopes and spherical designs, in C. Davis et al., eds., *The Geometric Vein: The Coxeter Festschrift*, Springer-Verlag, NY, 1981, pp. 203–218.

[5] R. H. Hardin and N. J. A. Sloane, New spherical 4-designs, *Discrete Math.*, **106/107** (1992), 255–264.

[6] R. H. Hardin and N. J. A. Sloane, An improved snub cube and other new spherical designs in three dimensions, *Discrete and Computational Geometry*, submitted.

[7] Y. Mimura, A construction of spherical 2-designs, *Graphs and Combinatorics*, **6** (1990), 369–372.

[8] B. Reznick, Some constructions of spherical 5-designs, *Linear Algebra and Its Applications*, 1995, to appear.

Computing the Voronoi cell of a lattice: The diamond-cutting algorithm

Emanuele Viterbo and Ezio Biglieri[1]
Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy)
fax: +39 11 5644099 • e-mail: <name>@polito.it

Abstract — **A computational algorithm is described for the numerical evaluation of some lattice parameters such as density, thickness, dimensionless second moment (or quantizing constant), etc. By using this algorithm, previously unknown quantizing constants of some interesting lattices can be obtained.**

I. Introduction

The complete geometric structure of a lattice can be found from the description of its Voronoi cell. The knowledge of the Voronoi cell solves at once the problem of the computation of relevant lattice parameters such as packing radius, covering radius, kissing number, center density, thickness, normalized second moment (or quantizing constant).

The Voronoi cell of certain highly symmetric lattices can be determined analytically but no such result is available for an arbitrary lattice. In this paper we propose an algorithm which exactly computes the Voronoi cell of a full-rank arbitrary lattice. The exact knowledge of the Voronoi cell (i.e., knowledge of the coordinates of its vertices, edges, etc.) enables one to compute all the lattice parameters within any degree of accuracy.

The Voronoi cell of lattice is an $\mathbf{0}$-symmetric convex polytope, i.e., a bounded region delimited by a finite number of hyperplanes symmetric about the origin. The basic elements of a polytope $\mathcal{P}$ are its *k-faces*, where k is the dimension. The 0-faces are called *vertices of* $\mathcal{P}$, the 1-faces, *edges of* $\mathcal{P}$ and the $(d-1)$-faces, *facets of* $\mathcal{P}$. For convenience we identify $\mathcal{P}$ with the d-face and the empty set with the (-1)-face. To give a complete description of a polytope we must know all the relations among its faces. For $-1 \leq k \leq d-1$ a k-face f and a $(k+1)$-face g are *incident upon* each other if f belongs to the boundary of g; in this case, f is called a *subface of* g and g a *superface of* f. The d-face represents the whole polytope and is the only superface of all the facets. The (-1)-face has no subfaces and is the only subface of all the vertices. The *incidence graph* $I(\mathcal{P})$ of $\mathcal{P}$ is an undirected graph defined as follows: for each k-face ($k = -1, 0, 1, \ldots d$) of $\mathcal{P}$, $I(\mathcal{P})$ has a node $\nu(f)$; if f and g are incident upon each other then $\nu(f)$ and $\nu(g)$ are connected by an arc.

II. The diamond-cutting algorithm

This algorithm computes the incidence graph of the Voronoi region $\mathcal{V}$ of a lattice. Its name was chosen due to its resemblance to the procedure for cutting a raw diamond into a brilliant. Let us consider a lattice Λ defined by an arbitrary basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_d\}$. Given a point $\mathbf{p}$ we will denote with $h(\mathbf{p})$ the hyperplane passing through the point $\mathbf{p}$ and normal to the vector $\mathbf{p}$. The distance of $h(\mathbf{p})$ from the origin is equal to $\|\mathbf{p}\|$.

Preparation Given the lattice basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_d\}$ construct the parallelotope $\mathcal{Q}$ defined by the hyperplanes $h(\pm\frac{1}{2}\mathbf{v}_i)$ for $i = 1, \ldots, d$. $\mathcal{Q}$ contains the Voronoi cell. The corresponding incidence graph $I(\mathcal{Q})$ has 3^d nodes. Finally, set $\mathcal{V} := \mathcal{Q}$.

Cutting Consider all hyperplanes $h(\frac{\lambda_1}{2}\mathbf{v}_1 + \frac{\lambda_2}{2}\mathbf{v}_2 + \cdots \frac{\lambda_d}{2}\mathbf{v}_d)$, with λ_i integers, which cut $\mathcal{V}$ and update $I(\mathcal{V})$, by introducing the nodes corresponding to the new faces and erasing those corresponding to the faces which are left out. For this operation we have adapted Edelsbrunner's algorithm for the incrementation of arrangements [2].

Finish Compute $\mathrm{vol}(\mathcal{V})$, the volume of $\mathcal{V}$. If $\mathrm{vol}(\mathcal{V}) > \det(\Lambda)^{1/2}$ keep on cutting, else end the algorithm and output the incidence graph $\mathcal{I}(\mathcal{V})$.

III. Results

By introducing some additional information into the nodes of the incidence graph, it is possible to compute all the lattice parameters once the Voronoi cell is found. In particular we easily find the *packing radius*, the *kissing number*, the *covering radius* and the related parameters. Finding the *quantization constant* requires a slightly more complex procedure which recursively computes the volume and second order moment of $\mathcal{V}$ about $\mathbf{0}$ in terms of the volume and of the second-order moment of the subfaces.

Using the diamond-cutting algorithm we have computed some previously unknown values of the quantizing constants for some particularly interesting lattices. Of special interest are the previously unknown quantizing constants for the two locally optimal lattice coverings in $\mathbf{R}^4$ found by Dickson (Di_{4a} : 0.076993; Di_{4b} : 0.077465) and for a 5-dimensional extreme lattice covering, which belongs to the class introduced by Barnes and Trenerry (0.076278). As these lattices do not improve upon the best known lattice quantizers, the conjecture about the optimal lattice quantizers being the duals of the densest lattices still holds.

Most of the computational problems related to lattices are either known or conjectured to be NP-hard [1, p. 40]. The principal limitation in the application of the diamond-cutting algorithm is the exponentially increasing memory requirement. The possibility of reducing the memory requirements appears remote especially if we want to preserve the generality of the algorithm.

References

[1] J. H. Conway, N. J. A. Sloane, *Sphere packings, lattices and groups*. 2nd ed. Berlin: Springer-Verlag, 1992.

[2] H. Edelsbrunner, *Algorithms in combinatorial geometry*. Berlin: Springer-Verlag, 1987.

[1] This research was sponsored by the Italian National Research Council (CNR) under "Progetto Finalizzato Trasporti."

A New Sphere Packing in 20-Dimensional Euclidean Space

Alexander Vardy

Coordinated Science Laboratory, University of Illinois, 1308 W. Main Street, Urbana, IL 61801, USA

vardy@golay.csl.uiuc.edu

Abstract — **We describe a new nonlattice sphere packing $\mathcal{J}_{20} \subset \mathbb{R}^{20}$ which is denser than any previously known sphere packing in $\mathbb{R}^{20}$. Properties of $\mathcal{J}_{20}$ are investigated, ans several alternative representations of the new packing are presented. One of these was recently recognized by Conway and Sloane as the first example of the so-called antipode packings, leading them to the discovery of new densest-known sphere packings also in dimensions 22 and 44–47.**

I. Introduction

It is well-known since the celebrated work of Shannon [4] that the design of efficient transmission codes for band-limited channels with additive white Gaussian noise is equivalent to the problem of constructing dense arrangements of nonoverlapping spheres in $\mathbb{R}^n$. In the study of dense sphere packings in $\mathbb{R}^n$, a particular effort has been devoted to dimensions $n \leq 24$. For $n \leq 24$ major progress was achieved by John Leech with the construction of his famous Leech lattice Λ_{24}, and the sequence of laminated lattices $\Lambda_0, \Lambda_1, \ldots \Lambda_{24}$, which may be obtained as cross-sections of Λ_{24}. Presently, the laminated lattices are the densest packings known in dimensions $n \leq 29$, except for $n = 10, 11, 12, 13$. For $n = 12$ the Coxeter-Todd lattice K_{12} is the densest known packing. For $n = 10, 11, 13$ nonlattice packings denser than the laminated sequence were found by Leech and Sloane [3] in 1970. Notwithstanding the vast body of research devoted to constructions of dense sphere packings in recent years — see [1] and references therein — no further progress for $n \leq 24$ has been reported in the intervening two and a half decades.

II. The Construction

Given a sequence of binary codes $C_0, C_1, \ldots, C_m$, consider a packing Λ consisting of all the points $x \in \mathbb{Z}^n$ with the following property: the 2^i-s row in the coordinate array of x is a codeword of C_i for $i = 0, 1, \ldots, m$. We use the notation $\Lambda = C_0 + 2C_1 + \cdots + 2^m C_m + 2^{m+1}\mathbb{Z}^n$, to describe such a packing. Now, let C and C^* be two orthogonal (n, M_1, d_1), respectively (n, M_2, d_2), binary codes with $d_1, d_2 \geq n/4 + 2$. We shall use $\mathbf{0}, \mathbf{1}$ to denote (codes consisting of) the all-zero and the all-one n-tuples, respectively. The $(n, 2^{n-1}, 2)$ binary code consisting of all the vectors in $\mathbb{F}_2^n$ of even weight, respectively odd weight, is denoted $\mathcal{E}_n$, respectively $\mathcal{O}_n$. Consider two sphere packings $\mathcal{J}_e, \mathcal{J}_o \subset \mathbb{R}^n$, defined as follows:

$$\begin{aligned} \mathcal{J}_e &= \mathbf{0} + 2C' + 4\mathcal{E}_n + 8\mathbb{Z}^n \\ \mathcal{J}_o &= \mathbf{1} + 2C^* + 4\mathcal{O}_n + 8\mathbb{Z}^n \end{aligned} \tag{1}$$

where $C' = \mathbf{1} + C$. Let $\mathcal{J} = \mathcal{J}_e \cup \mathcal{J}_o$. We show that for $n \leq 24$, the center density of $\mathcal{J}$ is given by

$$\delta(\mathcal{J}) = \frac{(n+8)^{n/2}(M_1 + M_2)}{2^{3n+1}} \tag{2}$$

Although (2) holds for all $n \leq 24$, it is clear from the condition $d_1, d_2 \geq n/4+2$ that the construction of (1) would be most successful for n divisible by 4. For $n = 8, 24$, we take the $(8, 2^4, 4)$ Hamming code and the $(24, 2^{12}, 8)$ Golay code and, by virtue of the fact that these codes are self-orthogonal, reproduce the lattice packings E_8 and Λ_{24}, respectively. For $n = 20$ our construction calls for two orthogonal $(20, 512, 7)$ codes. A $(20, 2^9, 7)$ linear code C is known (cf. [1, p.248]), and the question is whether its dual contains another $(20, 512, 7)$ subcode. This question is settled in the affirmative using a quaternary representation for C and $C^\perp$, similar to the constructions of the Golay code from the $(6, 4^3, 4)$ hexacode, and of the Nordstrom-Robinson code from the $(4, 4^2, 3)$ quadracode over $\mathbb{F}_4$. Thus, the codes C and C^* may be identified with certain binary images of two different $(5, 4^2, 4)$ subcodes of the $(5, 4^3, 3)$ perfect Hamming code over $\mathbb{F}_4$. The resulting nonlattice packing $\mathcal{J}_{20}$ has center density $7^{10} \cdot 2^{-31} = 0.1315\ldots$ This is denser than the best previously known packing Λ_{20} whose center density is $1/8$.

III. Properties of $\mathcal{J}_{20}$

We provide several alternative representations of $\mathcal{J}_{20}$ and investigate some of its properties. In particular, we show that $\mathcal{J}_{20}$ may be constructed as an $\mathcal{H}$-packing, where $\mathcal{H}$ is the ring of Hurwitz quaternions. Furthermore, we prove that although $\mathcal{J}_{20}$ is not a lattice it is distance invariant. This allows us to express the theta series of $\mathcal{J}_{20}$ in terms of the theta functions $\theta_2(z)$,$\theta_3(z)$,$\theta_4(z)$ and the weight distribution of the $(20, 2^9, 7)$ binary code C. Precise enumeration of the first six shells of the new packing is presented. In particular, the kissing number of $\mathcal{J}_{20}$ is shown to be 15360, which is slightly less then the kissing number of Λ_{20} given by 17400. This demonstrates once again that in general the answers to the packing problem and the kissing number problem may differ (cf. [1, p.23]).

Although we establish the distance invariance of $\mathcal{J}_{20}$, we were unable to determine whether $\mathcal{J}_{20}$ has the stronger property of geometrical uniformity. This was recently settled by Conway and Sloane [2], who showed that the affine automorphism group of $\mathcal{J}_{20}$ is not only transitive on the spheres, but also doubly-transitive on adjacent spheres. In fact, Conway and Sloane [2] provide a complete characterization of $Aut(\mathcal{J}_{20})$ in terms of the automorphism group of the Leech lattice.

Finally, we provide yet another representation of $\mathcal{J}_{20}$ as the union of four cosets of $2\Lambda_{20}$. Conway and Sloane [2] show that this representation of $\mathcal{J}_{20}$ is a special case of their new antipode construction of sphere packings. The antipode construction of [2] is remarkable in that it readily establishes the existence of sphere packings in dimensions 22,44,45,46,47 that are denser than previously known. All these packings were discovered in [2]. We note here that in most cases (including $\mathcal{J}_{16}$ and $\mathcal{J}_{20}$), the antipode set is a simplex.

*This work was supported by the NSF Grant NCR-9409688

References

[1] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd Ed., Springer-Verlag, New York, 1993.

[2] J.H. Conway and N.J.A. Sloane, "The antipode construction for sphere packings," preprint.

[3] J. Leech and N.J.A. Sloane, "Sphere packings and error-correcting codes," *Canad. J. Math.*, vol. 23, pp. 718–745, 1971.

[4] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and pp. 623–656, 1948.

Asymptotically Optimal Spherical Codes[1]

Jon Hamkins[2] and Kenneth Zeger[2]

[2] Coordinated Science Lab., Dept. of Elect. and Comp. Engineering, University of Illinois, Urbana-Champaign, IL 61801
email: hamkins@uiuc.edu, zeger@uiuc.edu

Abstract — **A new class of spherical codes is presented which are designed analogously to laminated lattice construction. For many minimum angular separations, these "laminated spherical codes" outperform the best known spherical codes. In fact, for fixed dimension $k \leq 49$, the density of the laminated spherical code approaches the density of the $(k-1)$-dimensional laminated lattice Λ_{k-1}, as the minimum angular separation $\theta \to 0$. In particular, the three-dimensional laminated spherical code is asymptotically optimal, in the sense that its density approaches the Fejes Tóth upper bound as $\theta \to 0$. The laminated spherical codes are also structured, which simplifies decoding.**

A *spherical code* $\mathcal{C}(k,\theta)$ is a set of points on the surface of a k-dimensional unit radius sphere S_k having minimum angular separation θ. The *density* of $\mathcal{C}(k,\theta)$, denoted $\Delta_{\mathcal{C}(k,\theta)}$, is the ratio of the surface area of $|\mathcal{C}(k,\theta)|$ disjoint spherical caps centered at the codepoints and with angular radius $\theta/2$, to the surface area of S_k. Let $\Delta(k,\theta) = \max_{\mathcal{C}(k,\theta)} \Delta_{\mathcal{C}(k,\theta)}$. Note that the maximum number of codepoints in any k-dimensional spherical code with minimum angular separation θ can be determined directly from $\Delta(k,\theta)$. We refer to a family of codes $\mathcal{C}(k,\theta)$ as *asymptotically optimal* if $\Delta_{\mathcal{C}(k,\theta)}/\Delta(k,\theta) \to 1$ as $\theta \to 0$.

For fixed dimension k and small minimum angular separation θ, [Fej59] ($k=3$) and [Cox68] ($k \geq 4$) provide the tightest upper bound and [GHSW87] provides the tightest known lower bound on $\Delta(k,\theta)$. However, there is a gap between these bounds as $\theta \to 0$. In this paper we introduce a new spherical code construction analogous to laminated lattice construction. We call these codes *laminated spherical codes*. These new codes have larger asymptotic (for small θ) densities than any previously known spherical codes.

The laminated spherical codes are obtained by placing codepoints on concentric $(k-1)$-dimensional spheres and projecting each codepoint onto S_k by adding a kth coordinate to form a vector of unit norm. The set of points on each $(k-1)$-dimensional sphere is either a $(k-1)$-dimensional laminated spherical code, or another code formed from its deep holes. By nesting the concentric spheres closely, and placing codepoints of one sphere at the radial extension of the deep holes of codepoints of the previous sphere, a method similar to constructing laminated lattices (e.g., [CS93]) is used to construct our spherical codes, which we denote by $\mathcal{C}^\Lambda$. As more of these concentric spheres are stacked up, codepoints start spreading out, and the density lessens. To counteract this, a buffer zone is placed between concentric spheres, and a new, tighter packed $(k-1)$-dimensional spherical code is placed in the next sphere. A recursion describes the sequence of radii necessary to insure that both the desired minimum angular separation is maintained and the desired density is obtained.

Our construction has similarities to those of [Yag58] and [GHSW87] in that a projection from $k-1$ dimensions to k dimensions is used; the difference lies in the placement of points prior to the projection. Our technique is practical for creating codes of any size and thus provides a lower bound on achievable minimum distance as a function of code size.

Let $\Delta_{\mathcal{C}^\Lambda}(k) = \limsup_{\theta \to 0} \Delta_{\mathcal{C}^\Lambda(k,\theta)}$, and let Δ_{Λ_k} be the density of the sphere packing constructed from the laminated lattice Λ_k. In the laminated spherical code, layers ($(k-1)$-dimensional spheres) are stacked similarly to layers of lattices in a laminated lattice, and as a result, $\Delta_{\mathcal{C}^\Lambda}(k)$ is equal to the density of the sphere packing generated by Λ_{k-1}.

Theorem 1 $\Delta_{\mathcal{C}^\Lambda}(k,d) = \Delta_{\Lambda_{k-1}} - O(d^{1/k})$.

Corollary 1 *$\mathcal{C}^\Lambda(3,d)$ is asymptotically optimal and the Fejes Tóth upper bound is asymptotically tight.*

Corollary 2 *If there exists a family of spherical codes $\mathcal{C}(k,d)$ whose asymptotic density is higher than $\Delta_{\mathcal{C}^\Lambda}(k,d)$, then there exists a $(k-1)$-dimensional sphere packing denser than that generated by Λ_{k-1}.*

Theorem 2 *There is an optimal decoder for $\mathcal{C}^\Lambda(k,\theta)$ using $O(\sqrt{|\mathcal{C}^\Lambda(k,\theta)|})$ space and $O(\log|\mathcal{C}^\Lambda(k,\theta)|)$ time, or an optimal decoder using $O(1)$ space and $O(\sqrt{|\mathcal{C}^\Lambda(k,\theta)|})$ time.*

References

[Cox68] H. S. M. Coxeter. *Twelve Geometric Essays.* Southern Illinois University Press, 1968.

[CS93] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices, and Groups.* Springer-Verlag, 1993.

[GHSW87] A. A. El Gamal, L. A. Hemachandra, I. Shperling, and V. K. Wei. Using simulated annealing to design good codes. *IEEE Trans. Inf. Thy.*, IT-33(1), January 1987.

[Fej59] L. Fejes Tóth. Kugelunterdeckungen und Kugelüberdeckungen in Räumen konstanter Krümmung. *Archiv Math.*, 10:307–313, 1959.

[Wyn65] A.D. Wyner. Capabilities of bounded discrepancy decoding. *Bell Sys. Tech. J.*, pages 1061–1122, July-Aug. 1965.

[Yag58] I. M. Yaglom. Some results concerning distributions in n-dimensional space. Appendix to Russian edition of Fejes Toth's Lagerungen in der Ebene, auf der Kugel und in Raum, 1958.

[1]The research was supported in part by the National Science Foundation, the Joint Services Electronics Program, and Engineering Research Associates Co.

Applications of TCM with σ-Tree Constellations over the AWGN Channel

Mahdi Y. Zaidan[1], Christopher F. Barnes[2], and Stephen B. Wicker[1]
School of Elec. & Comp. Engineering
Georgia Institute of Technology, Atlanta, Georgia, USA

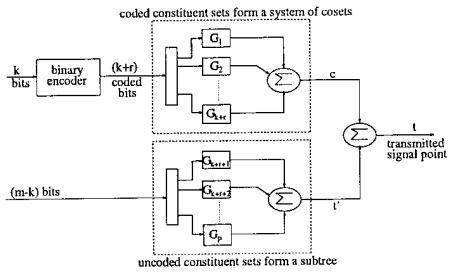
Abstract — **σ-trees are a class of geometric structures that include lattices as a constrained special case. These structures allow for signal sets that are more spherical in shape than lattices in spaces of arbitrary dimensionality. In this paper, the use of σ-trees in the construction of *non-lattice* TCM codes is established and investigated.**

I. Introduction

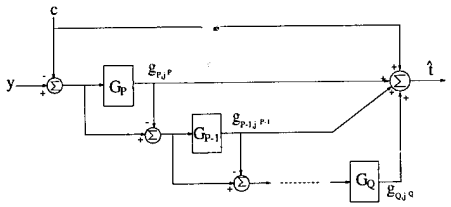
A P-level σ-tree signal constellation $T_{(1:P)}$ consists of a set of 2^P points in N-space that is formed from the direct sum of an ordered P-member collection binary constituent sets $(G_1, G_2, \ldots, G_P)$. One vector (called a generator), $g_{p,j} \in G_p$, $j \in \{0,1\}$, is selected from each constituent set and all selected vectors are summed to form a point, $\mathbf{t}$, in the signal constellation. A $(P-Q+1)$-level subtree $T_{(Q:P)}$ of $T_{(1:P)}$ is a σ-tree that is formed from the last $(P-Q+1)$ constituent sets; i.e. $T_{(Q:P)} = G_Q + G_{Q+1} + \cdots + G_P$. The design of these constellations is based on an iterative algorithm that uses training data drawn from multidimensional probability distribution functions [1, 2]. More interestingly, a σ-tree signal constellation T has a sequence of subtrees T' that induces a partition of T into partition chains with expanded intra-subtree minimum distances.

II. σ-Tree TCM Codes

A σ-tree coset code $\mathbf{C}(T/T'; C)$ is based on a σ-tree T, a σ-subtree T', and a binary encoder C. Figure (1a) illustrates the general encoder structure. The order of the constituent sets $G_1, G_2, \ldots G_{m+r}$ plays an essential role in determining a useful partition of T. For the one-dimensional σ-tree T, the constituent set with the lowest energy has to be in the first level of the tree, then continuing in ascending order until we have the constituent set with the largest energy in the last level [3]. To transmit m bits per N dimensions, the signal constellation must be based on an $(m+r)$-level σ-tree T, partitioned into 2^{k+r} subsets, each consisting of 2^{m-k} points from a different coset of the $(m-k)$-level σ-subtree T'. Constituent sets are divided into *coded* and *uncoded* constituent sets, based on the data bits to address them. The direct sum of the uncoded constituent sets form the σ-subtree T', while the direct sum of the coded constituent sets form a system of cosets. Of the incoming m data bits, k bits are applied to a binary encoder to get a (k+r)-coded bits with which to select a subset of the σ-tree TCM code. This is performed with each of the (k+r)-coded bits selecting one generator from the binary coded constituent sets. The direct sum of the generators form a coset representative, $\mathbf{c}$. The remaining (m-k) uncoded bits selects a point t' from the σ-subtree $T_{(k+r+1:m+r)}$, added to the coset representative to form the transmitted signal point



(a) Encoder.



(b) Decoder.

Figure 1: σ-tree TCM encoder and decoder

$\mathbf{t} = \mathbf{c} + \mathbf{t}'$; i.e. $\mathbf{t}(\mathbf{j}) = \sum_{p=1}^{m+r} \mathbf{g}_p(j^p)$ where j^p is the pth element of the (m+r)-tuples binary label $\mathbf{j}$.

An optimum *subset decoder* shown in Fig. (1b), developed for the one-dimensional case, works as follows. For a $(P-Q+1)$-level subtree T_Q of a P-level binary σ-tree T_1, there are 2^{P-Q+1} parallel transitions between each pair of states. To choose one of these parallel transitions, $(P-Q+1)$ decisions are needed. First, the received channel output is translated by a coset representative $c(j^{Q-1}j^{Q-2}\cdots j^1)$ of the signal subset $S(\mathbf{j})$ assigned to the parallel transitions. Then the translated channel output is applied sequentially to the $(P-Q+1)$ constituent sets of T_Q starting with G_P, the constituent set of largest energy. At each stage, a generator is determined, then is subtracted from the current stage's input and the result is passed to the next stage (constituent set) till we end with the constituent set G_Q. The direct sum of the decoded generators $g_{P,j^P}, g_{P-1,j^{P-1}}, \cdots g_{Q,j^Q}$ and the coset $c(\mathbf{j})$ forms the decoded signal $\hat{t}$.

References

[1] C. F. Barnes and R. L. Frost, "Residual vector quantizers with jointly optimized code books," *IEEE Trans. Inform. Theory*, vol. IT-39, no. 2, pp. 565–580, Mar. 1993.

[2] C. F. Barnes, "Tree structured signal space codes," *DIMACS Series in Discrete Math. and Theory Comp. Science*, vol. 14, pp. 33–53, Jan 1993.

[3] M. Y. Zaidan, C. F. Barnes, and S. B. Wicker, "Trellis coded modulation with σ-tree codes," *submitted for publication IEEE Trans Inform. Theory*, Sept. 1994.

[1]This work was supported by Grant NCR-9216686
[2]This author is with Georgia Tech. Research Institute

Generalized Minimum Distance Decoding of Reed-Muller Codes and Barnes-Wall Lattices

Chun Wang, Bazhong Shen and Kenneth K. Tzeng[1]
Dept. Elect. Eng. & Comp. Sci., Lehigh University, Bethlehem, PA 18015-3084

Abstract — **Low complexity soft decision decoding algorithms for Reed–Muller codes and Barnes–Wall lattices are presented. These algorithms are constructed based on the usage of generalized minimum distance (GMD) decoding recursively. Evaluation of the algorithms on AWGN channel through computer simulation indicates a slight degradation in performance, compared to maximum likelihood decoding, but with considerable reduction in complexity.**

I. Introduction

Decoding Reed–Muller codes and Barnes–Wall lattices become very important because of extensive studies of various codes and lattices in coded modulation application in recent years. Most previous decoding algorithms relied on trellis decoding. However, trellis can become very complicated for codes of longer length or lattices of higher dimension. Therefore, following the approach suggested by Forney [3], we apply hard decision decoding algorithm via GMD decoding [1] to realize low complexity soft decision decoding of Reed–Muller codes and Barnes–Wall lattices. In [4] Taipale and Pursley proposed an improvement to Forney's GMD decoding algorithm. However, it still may fail to find an acceptable codeword. In this paper, we provide a measure of compensation and present low complexity soft decision decoding algorithms for Reed–Muller codes and Barnes–Wall lattices by recursively using GMD decoding. Evaluation of the algorithms on AWGN channel through computer simulation indicates a slight degradation in performance, compared to maximum likelihood decoding, but with considerable reduction in complexity.

II. GMD Decoding of Reed–Muller Codes and Barnes–Wall Lattices

We first show that the original majority logic decoding algorithm for Reed–Muller codes [2] can be easily extended to an error–and–erasure decoding procedure. Then we can incorporate the criterion in [4] to derive an improved GMD decoding procedure. Our soft decision decoding algorithm is then realized, based upon the $(u|u+v)$ construction of Reed–Muller codes, by recursively applying the improved GMD decoding procedure. Namely, if a received vector can not be decoded to a codeword in $RM(r,m)$, then decode it to codewords in $RM(r-1,m-1)$ and $RM(r,m-1)$ respectively. Finally, an acceptable codeword in $RM(r,m)$ can be obtained. The complexity of this algorithm for decoding Reed–Muller codes in the worst case is $n=2^m$ or $2^{2(m-r)} < n^2$, while in the average case, it will be much lower.

It is known that the connection between Barnes–Wall lattices and Reed–Muller codes can be described by various code formulas. Therefore it is obvious that the decoding of Barnes–Wall lattices can be directly derived from the decoding of Reed–Muller codes.

[1]This work was supported by the NSF Grant NCR-9406043.

III. Simulation Results

The error performance of the proposed algorithm for decoding $RM(1,3)$ and $RM(2,4)$ in AWGN channel is shown in Fig. 1, and further, performance for decoding $BW_8(E_8)$ and $BW_{16}(\Lambda_{16})$ assuming 16QAM signaling in AWGN channel is shown in Fig. 2.

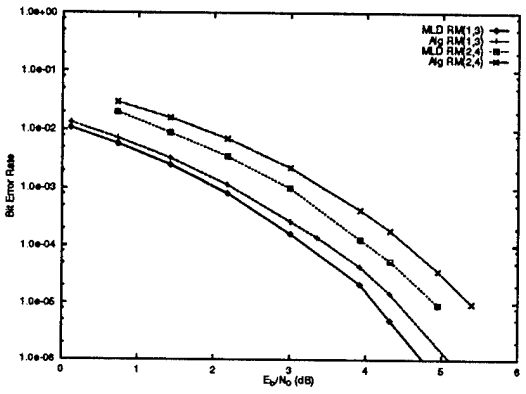


Figure 1: Proposed algorithm vs MLD algorithm for decoding RM(1,3) and RM(2,4)

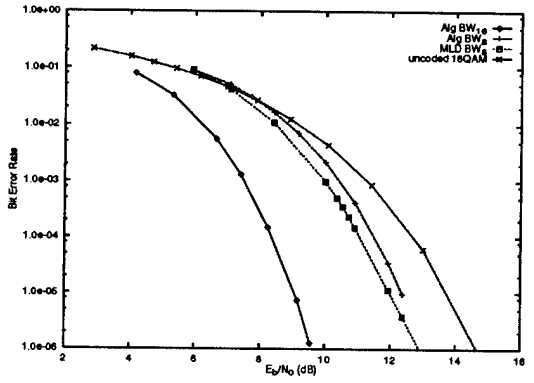


Figure 2: Bit error rate of coded 16QAM using BW_8 and BW_{16} lattices

References

[1] G. D. Forney, Jr., *Concatenated Codes* , Cambridge, Mass.: M.I.T. Press, 1966.

[2] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, New York: North-Holland, 1977.

[3] G. D. Forney, Jr., "A Bounded-Distance Decoding Algorithm for the Leech Lattice, with Generalizations", *IEEE Trans. Inform. Theory*, pp. 906-909, vol. 35, July 1989.

[4] D. J. Taipale and M. B. Pursley, "An Improvement to Generalized-Minimum-Distance Decoding", *IEEE Trans. Inform. Theory*, pp. 167-172, vol. 37, Jan. 1991.

Constellation Shaping for the Gaussian Channel[1]

Chris Heegard[2]

School of Electrical Engineering, Cornell University, Ithaca, New York, USA

Abstract — **In this paper we present a new view to the problem of constellation shaping. Both a new procedure and information theoretic analysis are discussed. The talk presents an approach to understanding constellation shaping that avoids the "continuous approximation" analysis of performance. A unique "type-mapping" approach to shaping is derived and related to monomial orderings on a ring of polynomials.**

I. Introduction

Constellation shaping is method of improving the effectiveness of digital communications over noisy, bandlimited channels. This topic has drawn considerable interest in recent days [1, 2, 3]. The roots of the topic go back to the paper on Lattice Codes and Cosets by Conway and Sloane[4], while the current framework for discussing the topic was outlined by Forney and Wei[5]. Three basic approaches to the problem were given by Lang and Longstaff [6], based on "shell mapping", Calderbank and Ozarow [7], based on "nonequiprobable signaling", and Forney[8] based on "trellis shaping". The recent high-speed telephone modem standard, v.34 ("v.fast") incorporates a version of shell mapping in the standard.

II. Discrete Analysis

The paper presents an approach to constellation shaping that avoids the "continuous approximation" (CA) analysis of performance. The crux of the CA method is related to the asymptotic shaping gain that can be derived from the entropy power of the uniform distribution. If X is a uniform random variable on the interval $[-A,\ A]$, then it has a "power" $P(X) = E(X)^3 = A^2/3$ and a differential entropy $h(X) = \frac{1}{2}\log(4A^2)$; a Gaussian random variable Y, with zero mean and variance σ^2, has "power" $P(Y) = \sigma^2$ and differential entropy $h(Y) = \frac{1}{2}\log(2\pi e\sigma^2)$. For equivalent entropy, the Gaussian power $P(Y) = \frac{6}{\pi e}P(X)$ which is -1.53 dB less then the Uniform power. This means, for transmission over power constrained channels, Gaussian distributed signaling has a 1.53 dB advantage over uniformly distributed signals.

In practice, however, discrete signal sets are always used. For example, consider how 2 bits might be transmitted over a $\mathcal{R}$ valued channel. The basic approach is to use the 4-PAM signal set $\{-3, -1, +1, +3\}$ with a uniform distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Then the entropy (rate) is 2 bits and the average power $\frac{1}{2}1 + \frac{1}{2}9 = 5$. To achieve a shaping gain, the signal set is increased and a code is used to induce a non-uniform distribution. For example, if the 6-PAM signal set, $\{-5, -3, -1, +1, +3, +5\}$ is used a gain can be achieved by selecting a blocklength $n = 4$ and signaling with the $2^{nR} = 2^{4\cdot 2} = 256$ least power signals. These signals induce a non-uniform marginal distribution of $(1, 28, 35, 35, 28, 1)/128$ resulting in a power of 4.875, a .11 dB improvement over 4-PAM . The optimum distribution for rate 2, 6-PAM is iid with marginal distribution $(0.0155, 0.1258, 0.3587, 0.3587, 0.1258, 0.0155)$; this has entropy of 2 bits and power 3.7569, a 1.2414 dB improvement! By going to 8-PAM , a 1.2525 dB is feasible with the maximum gain tops out at 1.2526 dB. Thus for a rate of 2 bits, the 1.53 dB gain is never obtainable (*i.e.*, for the 1.53 dB gain both n and R must grow to infinity).

The basic methods of studying tradeoffs are developed and explicit formulas are derived. The relationship to the capacity of the additive white Gaussian channel are discuss where it is shown that shaping techniques bridge the "uniform distribution" gap.

III. Type-Mapping

The basic methods of constellation shaping can be roughly characterized as forms of *coset coding* (*i.e.*, codes for which messages are associated with cosets of a subgroup such as linear subspace) and *enumerative coding* (*i.e.*, codes for which messages are enumerations of vectors). A unique "type-mapping" approach, an enumerative technique, is derived and related to monomial orderings on a ring of polynomials. It is shown how this approach provides a rate flexible and optimal tradeoff between peak and average power

References

[1] A. K. Khandani and P. Kabal, "Shaping multidimensional signal spaces—part i: Optimum shaping, shell mapping," *IEEE Transactions on Information Theory*, vol. IT-39, pp. 1799–1808, November 1993.

[2] A. K. Khandani and P. Kabal, "Shaping multidimensional signal spaces—part ii: Shell-addressed constellations," *IEEE Transactions on Information Theory*, vol. IT-39, pp. 1809–1819, November 1993.

[3] R. Laroia, N. Farvardin, and S. A. Tretter, "On optimal shaping of multidimensional constellations," *IEEE Transactions on Information Theory*, vol. IT-40, pp. 1044–1056, July 1994.

[4] J. H. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and cosets," *IEEE Transactions on Information Theory*, vol. IT-29, pp. 820–824, November 1983.

[5] J. G. David Forney and L. Wei, "Multidimensional constellations - part i: Introduction, figures of merit, and generalized cross constellations," *IEEE Journal on Selected Areas of Communications*, vol. SAC-7, pp. 877–892, August 1989.

[6] G. R. Lang and F. M. Longstaff, "A leech lattice modem," *IEEE Journal on Selected Areas of Communications*, vol. SAC-7, pp. 968–973, August 1989.

[7] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the gaussian channel," *IEEE Transactions on Information Theory*, vol. IT-36, pp. 726–740, July 1990.

[8] J. G. David Forney, "Trellis shaping," *IEEE Transactions on Information Theory*, vol. IT-38, pp. 281–300, March 1992.

[1]This work was supported in part by NSF grant NCR-9207331 and by the United States Army Research Office through the Army Center of Excellence for Symbolic Methods in Algorithmic Mathematics (ACSyAM), Mathematical Sciences Institute of Cornell University, Contract DAAL03-91-C-0027.

[2]heegard@ee.cornell.edu

Random Exploration of the Three Regular Polytopes

Nelson M. Blachman
GTE Government Systems Corporation
Mountain View, California 94039-7188

Abstract — **There are just three regular polytopes in Euclidean $(n > 4)$-space. Their dimensions are determined, including the distance from the centroid to the periphery in a random direction—that of a white-Gaussian-noise vector. As $n \to \infty$, this distance becomes very predictable. It differs from the distance near which almost all of the volume and surface of the polytope lie.**

There are infinitely many regular polygons, and there are 5 or 9 regular 3-dimensional solids. In Euclidean 4-space there are 6 or 16 regular polytopes. The second number—9 or 16—includes the possibility that faces will intersect one another internally and may also intersect themselves, as in the case of the regular pentagram (five-pointed star). But in spaces of $n \geq 5$ dimensions there are only 3 regular polytopes: the hypercube, which, for $n = 2$, 3, and 4, is a square, cube, and tesseract, respectively; the cross polytope, which is the dual of the hypercube and, for $n = 2$, 3, and 4, is a square, octahedron, and 16-hedroid; and the simplex, which is self-dual and, for $n = 2$, 3, and 4, is a triangle, tetrahedron, and pentahedroid.

The discrete set of different signals that might be transmitted during any signaling interval may be represented by a set of points in such a space. To each of these signal points belongs a Voronoi-polytope decision region. White Gaussian noise in the transmission channel will add random contributions to the coordinates of the transmitted-signal point, moving it a somewhat random distance in a uniformly distributed random direction and causing a reception error if it moves the signal point outside its Voronoi polytope. It is therefore of interest to understand how far such a polytope extends in a random direction and to compare that distance with the rms distance to a random point distributed uniformly over the volume of the polytope. Such questions are most easily answered for the simplest polytopes, i.e., the regular polytopes, and some of the phenomena exhibited by the regular polytopes will also occur in the others.

In each case we suppose that the regular polytopes have edges of unit length. Table I lists the height H_n, the distance I_{nk} from the center to the k-dimensional faces, the volune V_n, the radius $I_n = I_{n,n-1}$ of the inscribed sphere, the length L_n of a ray in a random direction from the center to the periphery, the radius S_n of a sphere having the same volume, the rms distance ρ_n to interior points, and the radius $C_n = I_{n0}$ of the circumscribed sphere for the 3 unit-edge regular polytopes.

The last five dimensions appear in the order of increasing size when $n \geq 15$. When $n < 15$, $\rho_n^{\text{cross}} < S_n^{\text{cross}}$; when $n < 10$, $\rho_n^{\text{cube}} < S_n^{\text{cube}}$; and when $n < 5$, $\rho_n^{\text{simp}} < S_n^{\text{simp}}$. Moreover, $\rho_n < I_n$ for the cross polytope if $n < 4$, for the cube if $n < 3$, and for the simplex if $n = 1$. For $n \gg 1$ the radius of the sphere having the same area as any of the three polytopes is asymptotically equal to its S_n.

Comparison of the fourth moment of the distance from the center of each regular polytope with the square of the second moment, ρ_n^4, shows that, for $n \gg 1$, nearly all of the volume lies within a thin spherical shell of radius ρ_n. The fact that $L_n < \rho_n$ for large n indicates that nearly all of the volume of these polytopes lies within a very small hypersolid angle about the center when $n \gg 1$. Setting ρ_n equal to I_{nk}, we find the largest k such that the boundary faces of dimensionality less than k lie wholly outside the spherical shell containing nearly all of the volume of the polytope, viz., $k = n/2$ for the simplex, $k = 2n/3$ for the hypercube, and $k = \frac{1}{2}n + 1$ for the cross polytope. Full details should appear next year in the *IEEE Transactions on Information Theory.*

Table I

Dimensions of the Regular Polytopes	Simplex	Hypercube	Cross Polytope
Edge	1	1	1
Height, H_n	$\sqrt{\frac{n+1}{2n}}$	1	$\sqrt{\frac{2}{n}}$
From center to k-dimensional face, I_{nk}	$\sqrt{\frac{n-k}{2(k+1)(n+1)}}$	$\frac{\sqrt{n-k}}{2}$	$\frac{1}{\sqrt{2(k+1)}}$
Content, V_n	$\frac{\sqrt{n+1}}{2^{n/2}n!}$	1	$\frac{2^{n/2}}{n!}$
Inradius, $I_n = I_{n,n-1}$	$\frac{1}{\sqrt{2n(n+1)}}$	$\frac{1}{2}$	$\frac{1}{\sqrt{2n}}$
Length of random ray from center, L_n	$\sim \frac{1}{2\sqrt{n \log n}}$	$\sim \sqrt{\frac{n}{8 \log n}}$	$\sim \sqrt{\frac{\pi}{4n}}$
Radius of equal sphere, S_n	$\sim \sqrt{\frac{e}{4\pi n}}$	$\sim \sqrt{\frac{n}{2\pi e}}$	$\sim \sqrt{\frac{e}{\pi n}}$
RMS radius, ρ_n	$\sqrt{\frac{n}{2(n+1)(n+2)}}$	$\sqrt{\frac{n}{12}}$	$\sqrt{\frac{n}{(n+1)(n+2)}}$
Circumradius, $C_n = I_{n0}$	$\sqrt{\frac{n}{2(n+1)}}$	$\frac{\sqrt{n}}{2}$	$\frac{1}{\sqrt{2}}$

Codes for the Lee Metric and Lattices for the l_1-Distance

Mohamed Siala and Ghassan Kawas Kaleh
Ecole Nationale Supérieure des Télécommunications,
46, rue Barrault, 75634 Paris 13, France

Forney has proposed an iterated construction called the squaring construction for simplified derivation and representation of the Barnes-Wall lattices. He used as a starting partition chain the two-dimensional infinite two-way partition ... $Z^2/\ Z^2/\ \mathcal{R}Z^2/\ 2Z^2/\ 2\mathcal{R}Z^2/\ \ldots$ with minimum Euclidean distances $\ldots 1/1/2/4/8/\ldots$, where $\mathcal{R}$ is a two-dimensional rotation operator . We apply this construction to the one-dimensional infinite two-way partition $\ldots Z/Z/2Z/4Z/8Z/\ldots$ with minimum l_1-distance $\ldots 1/1/2/4/8/\ldots$ which has clearly the same properties as the previous partition.The resulting lattices of dimension $N = 2^n$ for the l_1-distance can therefore be regarded as the duals of the Barnes-Wall lattices of dimension $2N$ for the Euclidean distance. Since the 2-depth of each of these lattices is equal to n they necessarily contain the $2^n Z^N$ lattice. The coset representatives of these lattices in $\nu 2^n Z^N$, where ν is an arbitrary nonnegative integer, are good codes for the Lee distance since they outperform the negacyclic codes in low dimensions. Maximum Likelihood (ML) soft detection can be performed easily on these lattices and codes since they have a simple trellis structure. Furthermore low complexity detection algorithms such as multistage decoding can be used without noticeable performance degradation. This is not the case for negacyclic codes where only algebraic hard decoding is performed easily using for example Euclid's algorithm.

The explicit expression of the lattices obtained by the Squaring Construction motivates us to consider a more general construction based on multilevel coding first proposed by Imai and Hirakawa. We consider jointly a μ-level code $C = [C_0, C_1, \ldots, C_{\mu-1}]$ and a finite partition chain $Z/qZ/q^2Z/\ldots/q^\mu Z$, where each code C_i is an (N, K_i, d_i) block code over the Galois Field GF($q = p^m$) with Hamming distance d_i, and m is an arbitrary nonnegative integer. An N-dimensional code Λ can be defined as the set of integer N-tuples λ that are congruent to $q^{\mu-1}c_{\mu-1} + \cdots + c_1 q + c_0$ modulo q^μ, where c_i is a codeword in the code C_i, i.e., the coefficients of q^i in the q-ary representation of λ are codewords in $C_i, 0 \leq i \leq \mu - 1$. The resulting code Λ is generally nonlinear and a necessary condition for it to be a lattice is that the component codes C_i satisfy the condition $C_i \subseteq C_{i+1}$. Also for a good design of Λ the Hamming distances of the component codes $C_{\mu-1}, \ldots, C_1, C_0$ should be chosen in the form $q, \ldots, q^{\mu-1}, q^\mu$. For the lattices obtained by the Squaring Construction $q = 2$ and the component codes are Reed-Muller codes that satisfy the two previous conditions on the component codes C_i.

In Section I we apply, as we have mentioned before, the Squaring Construction to the one-dimensional infinite two-way partition $\ldots Z/Z/2Z/4Z/8Z/\ldots$. We generalize the notion of the 2-depth of a binary lattice introduced by Forney to the case of nonbinary lattices and nonlinear codes and use this notion as a measure of the implementation complexity of the corresponding lattice. Furthermore, we derive the two-dimensional density of each lattice obtained by this construction and determine its behavior when the lattice dimension goes to infinity. We give also an explicit expression of the asymptotic value of this density as a function of the lattice 2-depth, which we assume fixed, when the lattice dimension goes to infinity.

For comparison reasons, we present in Sections II and III the negacyclic and shortened BCH codes designed for the Lee-metric, introduced respectively by Berlekamp and Roth and Siegel . We apply Construction A to these codes and derive dense lattices for the l_1-distance. Moreover, we give an explicit expression of the behavior of the two-dimensional density of these lattices when their dimension goes to infinity and show that it is the same in the two cases. We show also that the expression of the two-dimensional asymptotic density for fixed lattice 2-depth is identical in both cases to that found for the lattices obtained by the Squaring Construction.

Multilevel coding is considered in Section IV. As we have said before this construction provides a class of lattices and nonlinear codes which includes the lattices obtained by the Squaring Construction. We show that when considering the one-dimensional two-way partition $Z/2Z/2^2Z/\cdots/2^{\mu-1}Z$ and using binary BCH codes as component codes we obtain an approximate lattice density which is one quarter that obtained in the case of negacyclic and shortened BCH codes. However, for fixed 2-depth, the asymptotic two-dimensional density is found to be equal to that obtained for lattices based on negacyclic and shortened BCH codes.

In Section V we consider two applications of Lee-metric codes and l_1-distance lattices. The first one is concerned with shift, insertion and deletion error correction in peak-detection magnetic recording channels. The second one deals with error correction when transmitting data through the Rician channel. We have considered two four-dimensional constellations, with (2 bit/s)/Hz as spectral efficiency, based on the Schlfli lattice D_4, which is dense for the Euclidean distance, and the lattice E_4^{L}, which is dense for the l_1-distance and obtained by the Squaring Construction. The simulation results show that a coding gain of the order of 2 dB can be achieved by the constellation based on E_4^{L} over that based on D_4 for symbol error rates of the order of 10^{-4} when considering a Rician channel with specific characteristics to be detailed later. We show also that even if the lattice Λ_8^{L} obtained by the Squaring Construction, which is dense for the l_1-distance, can achieve a large asymptotic coding gain over the Gosset lattice E_8, which is dense for the Euclidean distance, it cannot provide positive coding gains for moderate signal-to-noise ratios because its kissing number is too large compared to that of E_8.

On the Redundancy of Lossy Source Coding*

Zhen Zhang[1], En-hui Yang[2], and Victor K. Wei[3]

I. Introduction

The redundancy of a source code is the difference between its expected performance and the optimum performance theoretically attainable(OPTA). The redundancy problem of source coding is to investigate the trade-off between the minimum redundancy over a class of codes having a common parameter(such as block length) and the common parameter. The significance of the redundancy problem is obvious when one takes into account the following facts: first, as compared with OPTA, the minimum redundancy gives the second-order theoretical performance and, therefore, is one of the basic problems in source coding theory; second, the redundancy problem provides a basis for comparison of different source coding algorithms; and finally, the minimum redundancy can tell algorithm-designers how much room they do have to improve the performances of their algorithms.

In this paper, we shall assume that the common parameter associated with the codes considered is block length. We shall refer the minimum redundancy over the class of all codes having block length n and some specified type as the nth-order redundancy. (In what follows, different names will be given for different types of codes.) In lossless source coding, the OPTA is the Shannon entropy and there exists extensive literature studying the nth-order redundancy. Typical results are: (1) when source statistics is known, the nth-order redundancy is $O(n^{-1})$; (2) when the statistics of a source is unknown, the nth-order redundancy grows as $O(\ln n/n)$.

In lossy source coding, the OPTA of a memoryless source p is its rate distortion function $R(p,d)$ when the memoryless source p is encoded by block codes at fixed distortion level d, i.e., d-semifaithful codes, and is its distortion rate function $d(p,R)$ when p is encoded by block codes at fixed rate level R. If $R(p,d)(d(p,R)$, resp.) is the OPTA, then the corresponding nth-order redundancy shall be refered to as the nth-order rate(distortion, resp.) redundancy. Unlike the case of lossless coding, in lossy source coding only a few research works on redundancy have been done. Specifically, Pilc[1] considered for the first time the problem of nth-order distortion redundancy. For a memoryless source p with finite source and reproduction alphabets, he proved that the n-th order distortion redundancy of p is upper bounded by $(-(1+\epsilon)\frac{\partial}{\partial R}d(p,R)\ln n/2n)(1+o(1))$ and argued that the n-th order distortion redundancy is lower bounded by $(-\frac{\partial}{\partial R}d(p,R)\ln n/2n)(1+o(1))$, where $\frac{\partial}{\partial R}d(p,R)$ is the derivative of $d(p,R)$ with respect to R. Recently, Yu and Speed[2] proved that for memoryless sources with finite source and reproduction alphabets, the n-th order universal rate redundancy is upper bounded by $(KJ+J+4)\ln n/n+o(n^{-1})$ and conjectured that $O(\ln n/n)$ is the optimal rate at which the n-th order rate redundancy converges to 0 as $n\to\infty$, where J and K are the sizes of the source alphabet and the reproduction alphabet, respectively. Linder, Lugosi and Zeger recently considered the case of real alphabet and proved that for memoryless sources, the n-th order universal distortion redundancy is upper bounded by $O(\sqrt{\ln n/n})$. Unfortunately, Pilc's argument to his lower bound is heavily based upon the unjustified assumption that the output of any block code can be approximated by an independent and identically distributed random vector. Whether or not the Pilc's lower bound is true is a question left open for more 25 years. Before our work, therefore, nontrivial lower bounds are still unknown to either the n-th order rate redundancy or the n-th order distortion redundancy.

The aim of this paper is to answer the above open questions. We derive a closed formula for the nth-order distortion redundancy and prove that the nth-order rate redundancy is upper bounded by $(\ln n)/n+o((\ln n)/n)$ and lower bounded by $(\ln n)/2n+o((\ln n)/n)$. As by-products, these results give positive answers to both the Pilc's open problem and the recent Yu-Speed's conjecture.

II. Statement of Main Results

Let $\{X_i\}_1^\infty$ be an I.I.D source taking values in a source alphabet $\mathbf{A}$ and having a generic distribution p. Let $\mathbf{B}$ be our reproduction alphabet. Denote by J and K the cardinalities of $\mathbf{A}$ and $\mathbf{B}$, resp. . Let $\rho:\mathbf{A}\times\mathbf{B}\to[0,\infty)$ be a single letter distortion measure. Denote by $R(p,d)(d(p,R)$, resp.) the rate distortion(distortion rate, resp.) function of p with respect to the fidelity criterion generated by ρ. If $\mathbf{C}_n\subset\mathbf{B}^n$ is a block code of order n with $|\mathbf{C}_n|\le e^{nR}$(in this paper, coding rates are measured in nats), the distortion redundancy $D_n(\mathbf{C}_n)$ of $\mathbf{C}_n$ is defined as $\rho_n(\mathbf{C}_n)-d(p,R)$, where $\rho_n(\mathbf{C}_n)$ is the average distortion resulting from the encoding of $\{X_i\}$ by $\mathbf{C}_n$. The nth-order redundancy $\mathcal{D}_n(R)$ is the minimum number of $D_n(\mathbf{C}_n)$ over all block codes $\mathbf{C}_n$ of order n with $|\mathbf{C}_n|\le e^{nR}$. For d-semifaithful codes of order n, we can similarly define the nth-order rate redundancy $\mathcal{R}_n(d)$. The following two theorems give the asymptotics of $\mathcal{D}_n(R)$ and $\mathcal{R}_n(d)$.

Theorem 1 *Let $R>0$. For sufficiently large n, we have*

$$\mathcal{D}_n(R)=-\frac{\partial}{\partial R}d(p,R)\frac{\ln n}{2n}+o\left(\frac{\ln n}{n}\right).$$

Theorem 2 *Assume $R(p,d)>0$. Then for sufficiently large n,*

$$\frac{\ln n}{2n}+o\left(\frac{\ln n}{n}\right)\le\mathcal{R}_n(d)\le\frac{\ln n}{n}+o\left(\frac{\ln n}{n}\right).$$

During the process of proving Theorems 1 and 2, we develop a deep theory on types and d-ball covering, which is also very interesting on its own.

*This work was supported in part by National Sciences Foundation under grant NCR 9205265.

[1]Commun. Science Institute, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565.

[2]Dept. of Math., Nankai University, Tianjin 300071, P.R. China.

[3]Dept. Infor. Eng., Chinese University of Hong Kong, Hong Kong.

References

[1] R. J. Pilc, "The transmission distortion of a source as a function of the encoding block length," *Bell Syst. Tech. J.*, Vol. 47, pp. 827–885, 1968.

[2] B. Yu and T. P. Speed, "A rate of convergence result for a universal D-semifaithful code," *IEEE Trans. Inform. Theory*, Vol. IT-39, pp. 813–820, 1993.

Mutual Information and Mean Square Error

Shunsuke Ihara
School of Informatics and Sciences, Nagoya University
Nagoya, 464-01 Japan

Abstract — **We derive some information theoretic inequalities to evaluate channel capacity and mean square error. We prove an inequality for the capacity of an additive noise channel with feedback. We also prove an inequality for mutual information and mean square error. The inequality is applied to bound minimum mean square transmission errors.**

Summary

We first study the capacity of an additive noise channel with feedback. In general, especially in non-Gaussian cases, it is a hard task to calculate the capacity exactly. So it is important to give effective lower or upper bounds on the capacity. Let ξ be a stochastic process representing an additive noise. We employ the notation ξ^* to denote a Gaussian process with the same mean and covariance functions as the process ξ. Corresponding to the channel with additive noise ξ, we consider a Gaussian channel with additive noise ξ^*.

Theorem 1 Assume that the channels are with feedback. Then, under an average power constraint, the capacity C of the channel with additive noise ξ is bounded by

$$C^* \le C \le C^* + H(\xi\|\xi^*), \tag{1}$$

where C^* is the capacity of the corresponding Gaussian channel and $H(\xi\|\xi^*)$ is the relative entropy (or information divergence) of ξ with respect to ξ^*.

In the case where the channels are without feedback, (1) has been obtained in [2].

It is interesting to recall the duality between the result (1) on the channel capacity and a result due to Binia et al. [1] on the rate distortion function. Denote by $R(D;\xi)$ the rate distortion function of a stochastic process ξ with mean square distortion. Then it is known that

$$R(D;\xi^*) - H(\xi\|\xi^*) \le R(D;\xi) \le R(D;\xi^*), \quad D > 0,$$

or equivalently

$$D[R + H(\xi\|\xi^*);\, \xi^*] \le D(R;\xi) \le D(R;\xi^*), \quad R > 0,$$

where $D(R;\xi)$ is the distortion rate function of ξ.

The second result relates the mean square error to the mutual information. We denote by $d(\xi,\eta)^2$ the mean square error between stochastic processes (or random variables) ξ and η.

Theorem 2 The mean square error is lower bounded by

$$d(\xi,\eta) \ge D[I(\xi,\eta) + H(\xi\|\xi^*);\, \xi^*], \tag{2}$$

where $I(\xi,\eta)$ is the mutual information between ξ and η.

The results (1) on the channel capacity C and (2) on the mean square error $d(\xi,\eta)^2$ are expressed in terms of the capacity C^* of the related Gaussian channel, the distortion rate function $D(\cdot;\xi^*)$ of the related Gaussian process, the mutual information, and the relative entropy. Results on the capacity of Gaussian channels are available in the literature [3] (and references therein). The rate distortion function $R(D;\xi^*)$ of the Gaussian process ξ^* is given in a closed form of D, and $D(R;\xi^*)$ is the inverse function of $R(D;\xi^*)$. The relative entropy $H(\xi\|\xi^*)$ may be regarded as the non-Gaussianness of ξ.

The inequality (2) is useful to evaluate the reproduction error in information transmission over a channel.

Theorem 3 Let a stochastic process ξ be a message to be transmitted over a channel of capacity C. Then the minimum mean square transmission error $\Delta(\xi)^2$ over the channel is bounded by

$$\Delta(\xi) \ge D[C + H(\xi\|\xi^*);\, \xi^*].$$

If a message ξ is a random variable with variance σ^2, then

$$\Delta(\xi) \ge \sigma \exp[-C + H(\xi\|\xi^*);\, \xi^*].$$

References

[1] J. Binia, M. Zakai and J. Ziv, "On the ε-entropy and the rate-distortion function of certain non-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 517–524, July 1974.

[2] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Inform. Control*, vol. 37, pp. 34–39, 1978.

[3] S. Ihara, *Information Theory for Continuous Systems*. Singapore: World Scientific, 1993.

Critical Distortion of Potts Model

Zhongxing Ye[1]
Dept. of Applied Math.
Jiao Tong University
Shanghai 200030
P.R.China

Toby Berger[2]
School of Elec. Eng.
Cornel University
Ithaea,NY 14850
U.S.A.

Abstract — It is shown by developing the previous technique that the critical distortion d_c of the q-ary Potts models on a number of lattices is related to the radius of convergence R of the Mayer's series by $d_c=(q-1)R/(1+R)$. A recursive approach is applied to estimate R as well as d_c by using the matrix representation of Mayer's series. For those Potts models of which the Mayer's series are not available, we derive an unified form of lower bound for d_c.

SUMMARY

A q-ary Potts model on Z^k is a random field $X=\{X_i, i\in Z^k\}$ with the following Gibbs distribution

$$\pi_x=\frac{1}{Z}\exp\{\Sigma_{<i,j>,\|i-j\|=1}J(1-\delta(x_i,x_j)\} \quad (1)$$

where

$$\delta(x_i,x_j)=\begin{cases}0 & \text{if } x_i=x_j\\ 1 & \text{if } x_i\neq x_j\end{cases}$$

$x_i\in Q=\{0,1,2,\cdots,q-1\}$

and the summation is taking over all the nearest neighboring pairs of sites on lattice Z^k. The Ising model is recovered when q=2.

The per-site ε -entropy for $X^{V^{(n)}}=\{X_i,i\in V^{(n)}\}$ on an finite subset $V^{(n)}=\{i=(i_1,\cdots,i_k),|i_j|\leqslant n\}\subset Z^k$, is defined by

$$R_X{}^{V^{(n)}}(d)=\inf\frac{1}{|V|}I(X^{V^{(n)}},Y^{V^{(n)}})$$

where the inf is over all random fields $Y^{V^{(n)}}$ such that

$$\frac{1}{|V|}\Sigma_i\rho(X_i,Y_i)\leqslant d$$

where $\rho(.,.)$ is Hamming distance on A*A.Then the per-site ε -entropy for X is defined by

$$R_x(d)=\lim_{n\to\infty}R_X{}^V(d)$$

if the limit exists.

Bassalygo and Dobrushin[1] proved for a wide class of q-ary random fields on Z^k that for sufficient small d:

$$R_x(d)=H_\infty(X)-\varphi(d) \quad (2)$$

where $H_\infty(X)$ is the entropy rate of the random field X, and $\varphi(d)=-d\log d-(1-d)\log(1-d)+d\log(q-1)$.

The critical distortion d_c is defined by

$$d_c=\sup\{d:R_x(d)=H_\infty(X)-\varphi(d)\} \quad (3)$$

They proved the existence of positive d_c using cluster expansion method, but didn't provide any estimation or bounds of d_c.

In this work we show that for random field in (1), d_c is related to the radius of convergence R of the Mayer's series associated with the Potts model by:

$$d_c=(q-1)\frac{R}{1+R} \quad (4)$$

We have provided a recursive approach[2] to compute R as well as d_c by the matrix rerepresentation of Mayer's series. In particular, we have applied this method to calculate the d_c for Ising models defined on several 2 or 3-dimensional even lattices.Let N denotes the number of the nearest neighboring sites of each site on the lattice.We found that d_c decreases as N or the dimension of lattices increase.

In the case that the series expansions are not available we can bound d_c using Ruelle's Theorem[3] from statistical mechanics. We derive the following lower bound for d_c:

$$d_c\geqslant\begin{cases}\frac{(q-1)b_1^{2k}}{1+b_1^{2k}} & \text{if } J<0\\ \frac{(q-1)b_2^{2k}}{1+b_2^{2k}} & \text{if } J>0\end{cases}$$

where

$$b_1=\frac{q-1-[(q-1)(q-1+e^J)(1-e^J)]^{1/2}}{(q-1)(q-2+e^J)}$$

$$b_2=\frac{(q-2+e^J)-[(q-1+e^J)(e^J-1)]^{1/2}}{(q-2)(q-1+e^J)+1}$$

When q=2,k=1,this bound coincides with the exact value of Gray's critical distortion.

Reference

[1] L.A.Bassalygo and R.L.Dobrushin,"ε -entropy of the random field",Prob. Peredach. Inform. vol.23,no.1,pp3-15,1987

[2] Zhongxing Ye and Toby Berger, "A new method to estimate the critical distortion of random fields", IEEE Trans. on Information Theory,vol. 38,no.1,pp152-157,1992

[3] D.Ruelle,"Some remarks on the location of zeroes of the partition function for lattice systems", Commun. Math. Phys.,vol.31,pp265-277,1973

1.Supported in part by China NNSF and US NSF Grants IRI-90005849

2.Supported in part by US NSF Grant IRI-90005849

On the Role of Mismatch in Rate Distortion Theory

Amos Lapidoth[1]

Information Systems Laboratory, Stanford Univerisity, Stanford, CA 94305-4055

Abstract — **Using a codebook $\mathcal{C}$, a source sequence is described by the codeword that is closest to it according to the distortion measure $d_0(x,\hat{x}_0)$. Based on this description, the source sequence is reconstructed to minimize the distortion measured by $d_1(x,\hat{x}_1)$, where in general $d_1(x,\hat{x}_1) \neq d_0(x,\hat{x}_0)$. We study the minimum resulting $d_1(x,\hat{x}_1)$-distortion between the reconstructed sequence and the source sequence as we optimize over the codebook subject to a rate constraint. Using a random coding argument we derive an upper bound on the resulting distortion. Applying this bound to blocks of source symbols we construct a sequence of bounds which are shown to converge to the least distortion achievable in this setup. This solves the rate distortion dual of an open problem related to the capacity of channels with a given decoding rule—the mismatch capacity.**

Addressing a different kind of mismatch, we also study the mean squared error description of non-Gaussian sources with Gaussian codebooks. It is shown that the use of a Gaussian codebook to compress any ergodic source results in an average distortion which depends on the source via its second moment only. The source with a given second moment that is most difficult to describe is the memoryless zero-mean Gaussian source, and it is best described using a Gaussian codebook. Once a Gaussian codebook is used, we show that all sources of a given second moment become equally hard to describe.

I. Mismatched Description

The design and implementation of lossy block source compression is usually done in three steps. The first step is to find a single-letter distortion measure that best describes the needs and sensitivities of the end-user (reconstructor). Based on this distortion measure and on the probability law that governs the source behavior, a codebook is designed to minimize the average distortion subject to some rate and complexity constraints. Finally the source output sequence is described by the index of the closest codeword to the source sequence according to the distortion measure. The end-user then reconstructs the source sequence based on the index, the codebook and the distortion measure.

Our interest is in a situation where the distortion measure $d_1(x,\hat{x}_1)$ that best describes the sensitivities of the end-user is different from $d_0(x,\hat{x}_0)$ according to which the source is encoded. Such a situation can arise if encoding according to $d_0(x,\hat{x}_0)$ is easier to implement than encoding to minimize $d_1(x,\hat{x}_1)$, or when one attempts to reconstruct a source that was compressed using a standard lossy compression algorithm over which one has no control. The codebook and the two distortion measures are known to the end-user. Only the codeword nearest to the source sequence, not the source sequence itself, is available to him, and he needs to reconstruct the source sequence to minimize the $d_1(x,\hat{x}_1)$-distortion. A formal statement of the problem follows.

A blocklength n code $\mathcal{C}$ of size 2^{nR} over a finite alphabet $\hat{\mathcal{X}}_0$ is used to encode a memoryless source of law $p(x)$ that takes value in a finite alphabet $\mathcal{X}$. A source sequence $\mathbf{x}$ is described by the codeword $\hat{\mathbf{x}}_0(i)$ that is nearest to $\mathbf{x}$ according to the single-letter bounded distortion function $d_0(x,\hat{x}_0)$. Based on the description $\hat{\mathbf{x}}_0(i)$ and the knowledge of the codebook $\mathcal{C}$, we wish to reconstruct the source sequence to minimize the average distortion defined by the bounded distortion function $d_1(x,\hat{x}_1)$, where in general $d_1(x,\hat{x}_1) \neq d_0(x,\hat{x}_0)$. In fact, the reconstruction alphabets $\hat{\mathcal{X}}_0$ and $\hat{\mathcal{X}}_1$ may well be different. We study the minimum, over all codebooks $\mathcal{C}$ of rate R, of the average distortion between the reconstructed sequence $\hat{\mathbf{x}}_1(i)$ and the source sequence $\mathbf{x}$. This quantity is denoted by $D_1(R)$.

Using a random coding argument, an upper bound on $D_1(R)$ is derived. We show that this bound is in general not tight, and derive a monotonic sequence of upper bounds which converges to $D_1(R)$. This solves the rate distortion dual of an open problem related to the capacity of channels with a given decoding metric [1].

II. A Rate Distortion Saddlepoint

Here we focus on a different kind of mismatch—one where the source distribution is not the one for which the codebook was optimized. We consider real-valued ergodic sources and the mean squared error distortion measure. We study that performance that one can expect when one describes such a source using a "Gaussian codebook", where a Gaussian codebook is a random codebook whose codewords are drawn independently and uniformly over an n-dimensional Euclidean sphere. Using a result due to Wyner [2] we show the following.

Theorem 1 *Consider the ensemble of codebooks generated by drawing 2^{nR} codewords independently and uniformly over the n-dimensional sphere of radius r_n centered about the origin. Let $\mathbf{x}$ be an n-length source sequence generated by an ergodic source of second moment σ^2, and let $0 < D < \sigma^2$.*

(a) *If $R < \frac{1}{2}\log(\sigma^2/D)$ then irrespective of the radii*

$$Pr\left(\exists \hat{\mathbf{x}} \in \mathcal{C} \text{ s.t. } \|\mathbf{x}-\hat{\mathbf{x}}\|^2 \leq nD\right) \stackrel{n\to\infty}{\longrightarrow} 0.$$

(b) *If $R > \frac{1}{2}\log(\sigma^2/D)$ and $r_n = \sqrt{n(\sigma^2 - D)}$ then*

$$Pr\left(\exists \hat{\mathbf{x}} \in \mathcal{C} \text{ s.t. } \|\mathbf{x}-\hat{\mathbf{x}}\|^2 \leq nD\right) \stackrel{n\to\infty}{\longrightarrow} 1.$$

Acknowledgements

Stimulating discussions with Tom Cover, Aaron Wyner, and Emre Telatar are gratefully acknowledged.

References

[1] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. on Inform. Theory*, vol. 41, pp. 35–43, Jan. 1995.

[2] A.D. Wyner, "Random packing and coverings of the unit n-sphere," *Bell Syst. Tech. J.*, vol. 46, pp. 2111–2118, Nov. 1967.

[1]This research was carried out in part while the author was with AT&T Bell Laboratories, Murray Hill, NJ.

On Rate–Distortion Bounds of Sub–Gaussian Random Vectors

Frank Müller[1]

Institut für Elektrische Nachrichtentechnik, Aachen University of Technology (RWTH), 52056 Aachen, Germany

Abstract — **The Shannon lower bound on the rate distortion function of sub–Gaussian vectors is considered. It can be shown that the Shannon lower bound can be decomposed into a sum of the rate distortion function of a corresponding Gaussian vector and a correction term which accounts for the differing distribution shape. This correction term is numerically evaluated.**

I. Introduction

Sub-Gaussian random vectors can serve as source models for speech samples, coefficients in image transform or subband coding or as model for displaced frame differences as they occur in hybrid video coding. This is due to mainly two aspects. First, sub-Gaussian vectors show elliptically shaped contours of equal distribution, thus belonging to the class of *spherically invariant random vectors*; second, the univariate distribution shape is peaky and "thick-tailed" compared to the Gaussian distribution. Both, spherical invariance and peaky distribution fit well to the actual statistics of a wide variety of sources.

Leung and Cambanis [1] gave the Shannon lower bounds of spherically invariant random processes and vectors. Except for using the squared error distortion criterion their work was very general. In this contribution the Shannon lower bounds of sub-Gaussian random vectors will be evaluated. In order to keep the average distortion finite, the absolute error criterion is employed using results from [2].

II. Sub-Gaussian random vectors

Let $\mathbf{X}$ be a random vector with pdf $f(\mathbf{x})$. The (multivariate) characteristic function (cf) of $\mathbf{X}$ is then defined by $\Phi(\mathbf{t}) = \mathsf{E}e^{j\mathbf{t}^T\mathrm{X}}$, where $\mathbf{t}$ denotes a vector of same dimension as $\mathbf{X}$ and E denotes expectation.

A spherically invariant random vector (SIRV) is a random vector defined by the property that its characteristic function (cf) can always be written as

$$\Phi(\mathbf{t}) = h(u) \quad \text{with} \quad u = \mathbf{t}^T C\mathbf{t}, \tag{1}$$

where C is a positive definite matrix. Note, that Gaussian random vectors are included here as a special case with $h(u) = \exp(-u/2)$ and C being the covariance matrix of the vector.

Spherically invariant vectors are completely specified by the univariate marginal density function and the linear statistical dependencies (expressed in terms of C) between the components.

A random vector $\mathbf{X}$ is called sub-Gaussian, if and only if its characteristic function is given by

$$\Phi_{\mathrm{X}}(\mathbf{t}) = \exp\left[-\left(\mathbf{t}^T C\mathbf{t}\right)^{\alpha/2}\right], \tag{2}$$

where C is a positive definite matrix and $1 < \alpha \leq 2$ [3].

Note, that sub-Gaussian random vectors are SIRVs [4] and that zero mean Gaussian random vectors are included in the case $\alpha = 2$. Compared to Gaussian random vectors of same dimension, sub-Gaussian vectors are parameterized with only one additional parameter α which accounts for different distribution shapes.

III. Evaluation of Shannon lower bounds

Following [1][2], the Shannon lower bound $R_{SL}^{(n)}(D)$ of a sub-Gaussian random vector with cf (2), decomposes into a sum of the Shannon lower bound $R_{SL,G}^{(n)}(D)$ of a Gaussian vector with the same matrix C and a correction term $K(n,\alpha)$ depending only on the vector dimension n and the parameter α. Interestingly, the correction term does neither depend on the matrix C nor on the distortion D.

Because sub–Gaussian random vectors are SIRVs, the bounds can be determined via Hankel transform of the function $\exp(-|t|^\alpha)$ (see e.g. [5]). Values for the correction term in case of $\alpha = 1$ are given in the table. In this case the sub-Gaussian distribution is spherically invariant with Cauchy marginals. For α between 1 and 2 the correction term falls into the range between zero and the corresponding value in the table and can be determined (at least in principle) numerically.

n	1	2	4	8	16	32	∞
$K(n,1)$	1.10	0.94	0.79	0.66	0.57	0.51	0.39

Tab. 1: Correction term (in bit/sample) for sub-Gaussian random vector of dimension n with $\alpha = 1$.

Addition of the Shannon lower bound of a Gaussian vector (which is well known) leads then to the Shannon lower bound of a sub-Gaussian vector for any parameters C, α and n.

Acknowledgements

Thanks are due to Dipl.–Ing. Ole Harmjanz for pointing out the relevance of [2] for the evaluation of the Shannon lower bound and for the analytical and numerical calculations leading to the table.

References

[1] H. M. Leung and S. Cambanis, "On the rate distortion functions of spherically invariant vectors and sequences," *IEEE Trans. Inform. Theory*, vol. 24, pp. 367–373, 1978.

[2] Y. Yamada, S. Tazaki and R.M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory*, vol. 26, pp. 6–14, 1980.

[3] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, pp. 986-1010, 1993.

[4] F. Müller and B. Hürtgen, "A new spherically invariant joint distribution model for image signals," Proc. VIIth European Signal Processing Conference (EUSIPCO 94), vol. 2, pp. 1082–1085, Edinburgh, Scotland, Sept. 1994.

[5] I.F. Blake and J.B. Thomas, "On a class of processes arising in linear estimation theory," *IEEE Trans. Inform. Theory* , vol. 14, pp. 12–16, 1968.

[1] E-mail: mueller@ient.rwth-aachen.de

Rate Distortion Efficiency of Subband Coding with Crossband Prediction

Ping Wah Wong

Hewlett Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304

Abstract — **Traditional subband coding, where each subband is encoded independently, has been shown by Fischer to be suboptimal in the rate-distortion sense. We show that if we use prediction across subbands, the resulting coder is asymptotically rate-distortion optimal at high rate.**

I. Introduction

In a typical subband coding system, the signal is first decomposed into subsignals, then a bit allocation algorithm is used to determine the rate to encode each subsignal, and finally each subsignal is encoded independent of the others. It is shown [1, 2] that for Gaussian sources and for ideal (brick-wall) subband filters, subband coding can achieve at high rate a coding gain over PCM. Recently, Fischer [3] showed that subband coding for Gaussian sources with QMF filters is generally suboptimal in the rate distortion sense.

Theorem: (Fischer 1992) Consider a Gaussian process x_n with spectral density $S_x(f)$, which is decomposed by a QMF system into the subsignals s_n and d_n. If we encode s_n and d_n independently of each other, the optimal performance satisfies

$$\sigma_x^2\gamma_x^2 \le 2\sqrt{\sigma_s^2\gamma_s^2\sigma_d^2\gamma_d^2} = \exp\left\{\int_{-0.25}^{0.25} \log_e\left[\Delta(f) + S_x(f)S_x(f+0.5)\right] df\right\};$$

where

$$\Delta(f) = |H(f)|^2|H(f+0.5)|^2\left[S_x(f) - S_x(f+0.5)\right]^2.$$

The inequality is strict if $\Delta(f) > 0$ on a subset of $[-0.25, 0.25]$ of positive measure. ∎

The implication of the inequality is that the performance of subband coding at high rate is strictly lower bounded by the rate-distortion function of the source except for several special cases where $\Delta(f) = 0$, e.g., when the filter $H(f)$ is ideal (hence also the complimentary filter $G(f)$), or when $S_x(f)$ is symmetric about $f = 1/4$.

II. Subband Coding with Crossband Prediction

Consider the subband coder with crossband prediction as shown in Fig. 1. We first encode s_n to get $\widehat{s}_n$. We then use a linear predictor to generate $\widehat{d}_n$, a predicted version of d_n, from $\widehat{s}_n$, and then encode the prediction error $e_n = d_n - \widehat{d}_n$. To calculate the energy of the prediction error, we assume that s_n is available at the predictor input. This obviously cannot be the case at the decoder, and hence the results in this summary is only asymptotically exact at high rate. The mean square prediction error achieved using an optimum linear predictor is (see, for example, pages 432-435 of [4])

$$E[e_n^2] = E[(d_n - \widehat{d}_n)^2] = \int_{-0.5}^{0.5} \left[S_d(f) - |S_{ds}(f)|^2 S_s^{\oplus}(-f)\right] df,$$

where

$$\alpha^{\oplus}(f) = \begin{cases} 1/\alpha(f) & \text{if } \alpha(f) > 0 \\ 0 & \text{if } \alpha(f) = 0. \end{cases}$$

It is clear that the predictor error is also Gaussian. If we encode the components s_n and e_n using the optimum bit allocation, the resulting distortion is

$$D(R) = 2\sqrt{\sigma_s^2\gamma_s^2\sigma_e^2\gamma_e^2}\, 2^{-2R}.$$

We can then prove the following theorem:

Theorem: Let x_n be a Gaussian process with spectral density $S_x(f)$. It is decomposed using a two band QMF system into s_n and d_n, and then optimally encoded using the crossband predictive coder. The equality

$$2\sqrt{\sigma_s^2\gamma_s^2\sigma_e^2\gamma_e^2} = \sigma_x^2\gamma_x^2 \tag{1}$$

holds, which implies that the subband coding system with cross band prediction is asymptotically optimal in the rate distortion sense at high rates.

Proof: Proceed with

$$2\sqrt{\sigma_s^2\gamma_s^2\sigma_e^2\gamma_e^2} = \exp\left\{\int_{-0.25}^{0.25} \log_e\left(4S_s(2f)S_d(2f) - 4|S_{ds}(2f)|^2\right) df\right\} \tag{2}$$

As shown in [3],

$$4S_s(2f)S_d(2f) = S_x(f)S_x(f+0.5) + \Delta(f). \tag{3}$$

Using the equality $G(f) = e^{-j2\pi f}H(-f-0.5)$, we have

$$\begin{aligned} &4|S_{ds}(2f)|^2 \\ &= |e^{-j2\pi f}H(-f-0.5)H(-f)S_x(f) \\ &\qquad + e^{-j2\pi(f+0.5)}H(-f)H(-f-0.5)S_x(f+0.5)|^2 \\ &= \Delta(f). \end{aligned} \tag{4}$$

Substituting (3) and (4) into (2), we get the desired result. ∎

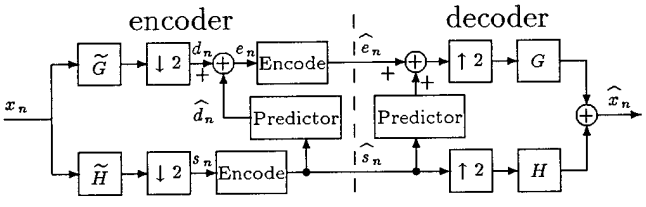


Figure 1: A subband coder with crossband prediction.

References

[1] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1278–1288, Oct. 1986.

[2] W. A. Pearlman, "Performance bounds for subband coding," in *Subband Image Coding* (J. W. Woods, ed.), ch. 1, pp. 1–41, Norwell MA: Kluwer Academic Publishers, 1991.

[3] T. R. Fischer, "On the rate-distortion efficiency of subband coding," *IEEE Trans. Info. Theory*, pp. 426–428, Mar. 1992.

[4] A. N. Shiryayev, *Probability*. New York: Springer-Verlag, 1984.

OPERATIONAL RATE DISTORTION THEORY

Ilan Sadeh
CS. Dept. Ben Gurion University, Beer Sheva, Israel, sade@ivory.bgu.ac.il

The paper treats data compression from the viewpoint of information theory where a certain error probability is tolerable. We obtain bounds for the minimal rate given an error probability for blockcoding of general stationary ergodic sources. An application of the theory of large deviations provides numerical methods to compute for memoryless sources, the minimal compression rate given a tolerable error probability. Interesting connections between Cramer's functions and Shannon's theory for lossy coding are found.

1. The deterministic partition approach

Given $u \in U$ and $v \in V$ a distortion-measure is any real positive function $d : [U \times V] \rightarrow \mathcal{R}^+$. Let $\rho_l(\bar{u};\bar{v})$ - denote the distortion for a block- the average of the per letter distortions for the letters that comprise the block. $\rho_l(\bar{u};\bar{v}) = \frac{1}{l}\sum_{i=1}^{l} d(\bar{u}_i;\bar{v}_i)$. Let D be a given tolerable level of distortion relative to the memoryless distortion measure $d(u,v)$.

The set of all possible codewords is partitioned into two disjoint subsets: Codebook and its complement set. The Codebook contains all the codewords in the code. Each sourceword $\bar{u}$ of length l is mapped onto exactly one of the codewords in the Codebook provided the distortion of the block is not larger than lD. Otherwise, the sourceword is included in the Error set and a coding failure is said to have occurred.

Definition 1: A $D-Ball$ covering of a codeword $\bar{v}$, denoted $\Upsilon(\bar{v})$, is a set of all sourcewords such that $\Upsilon(\bar{v}) = \left\{\bar{u}\middle|\rho_l(\bar{u},\bar{v}) \leq D\right\}$.
That is, we define spheres around all the possible codewords $\bar{v}$. But these spheres do not define probabilities on the codewords. Each sourceword should be mapped to exactly one codeword. Thus, we denote the set of the sourcewords that map to the codeword $\bar{v}$ after a partition, as $\mathbf{A}(\bar{v})$. Consequently the induced l-order entropy is, $H_v(l) = -\frac{1}{l}E\log\Pr(\bar{v})$.

Definition 2: An *acceptable partition* of blocklength l is a partition on the space of l length sourcewords such that for all $\bar{v}$, the associated subset $\mathbf{A}(\bar{v})$ satisfies $\mathbf{A}(\bar{v}) \subseteq \Upsilon(\bar{v})$ and that $\lim_{l\rightarrow\infty} H_v(l)$ exists.

Definition 3: The set $D-Ball(\bar{u})$ is defined as, $D-Ball(\bar{u}) = \left\{\bar{v}\middle|\rho_l(\bar{u},\bar{v}) \leq D\right\}$.

Lossy AEP Theorem:
For any acceptable partition of blocklength l and given any $\delta > 0$, the set of all possible sourcewords of blocklength l produced by the source can be partitioned into two sets, $Error$ and $Error^c$, for which the following statements hold:
1. Assuming a stationary and ergodic output process, the probability of a sourceword belonging to $Error$, vanishes as l tends to infinity.
2. If a sourceword $\bar{u}$ is in $Error^c$ then its associated codeword $\bar{v}$ is in the Codebook and its probability of occurrence is more than $e^{-l(H_v(l)+\delta)}$.
3. The number of codewords in the Codebook is at most $e^{l(H_v(l)+\delta)}$.
Given is a stationary ergodic source u with known probabilities for all blocklengths l, an acceptable average distortion D and a tolerable error probability P_e. Assuming the l order entropy induced by the chosen acceptable partition is $H_v(l)$, then the optimal code set is, $\Gamma_l(D,\delta) = \left\{\bar{v}\middle|\Pr(\bar{v}) \geq e^{-l(H_v(l)+\delta)}\right\}$ where a value δ is determined by the error probability. The error set is defined by, $Error_l(\delta,D) = \left\{\bar{u}\middle|\min_{\bar{v}:\rho(\bar{u},\bar{v})\leq D} -\frac{1}{l}\log\Pr(\bar{v}) - H_v(l) > \delta\right\}$.

2. Bounds on Memoryless Sources.

For a given P_e, a bound on the average distortion level D and a blocklength l, we find the best compression rate. The results, developed for memoryless source might be generalized for classes of sources for which there is a well-developed body of large deviations results for the source output process.
Our approach to the problem is based on transformation of the deterministic problem to a stochastic one and calculation of the error probability and the rate, using large deviations theory. Optimizing over all possible transitions matrices for a given error probability provides the solution. The loss of $\Psi_Q(\delta)$ amount of information in the transmission results in the compression by gaining $\psi_Q(\delta)$ nats. The term δ is determined by the tolerable error probability. We obtain a "conservation law", where the amount of the lost information is equal to the gain in the compression, only in the lossless case. It is an interesting interpretation for the two Cramer's functions in context of lossy data compression.

Distortion Measures for Variable Rate Coding

Stan McClellan[†] and Jerry D. Gibson[‡]

[†] Dept. of Electrical & Computer Engineering, U. Alabama at Birmingham
[‡] Dept. of Electrical Engineering, Texas A&M University

Abstract — **We apply the relative entropy functional to sets of Line-Spectrum Pairs (LSPs) and transform-based generalized spectral pmfs of [1] and present experimental results for sequence segmentation and vector quantization which show that the relative entropy of these quantities is a useful indicator for variable-rate speech coding.**

I. Activity Measures

Numerous methods for evaluating spectral differences (or distortion) have been explored in the literature. Many of the popular techniques pertain to optimal one-step-ahead linear prediction, or LPC models in speech processing systems. These approaches have been used to minimize distortion in vector quantizer (VQ) design for fixed-rate coding and for performance evaluation of different coding systems.

Recently, *spectral entropy* has been proposed as a different indicator of spectral information content and coefficient rate [1]. Here, we combine previous results which use subband spectral flatness measures for time-domain speech segmentation [2, 3] with a different application of the concept of spectral distance. This approach produces encoding cues which allow for the efficient allocation of rate in both the time and frequency domains.

The information-theoretic functional *relative entropy* is a convenient indicator of distance, since it produces a measure of the difference between a target distribution and a source distribution. The usual *entropy* functional is a special case of relative entropy where the source distribution is assumed to be uniform, and this case is of particular interest in waveform segmentation [1–3]. Thus, the use of relative entropy on appropriately normalized spectral data can be helpful in describing the flatness of the spectrum with respect to an average energy level, or in determining the evolution of nonstationary spectral representations.

For example, by dividing the normalized spectrum into upper and lower halfbands and applying the entropy functional to each halfband, we can derive an instantaneous indicator of useful bandwidth. This technique can be applied recursively to further refine the estimate. These halfband indications can be used to reduce encoding rate in the context of a scalar coder [3] by dynamically changing the sampling rate of the signal and in the context of a vector coder [4] by changing the allocation of rate for spectral VQ.

II. Line Spectral Entropy

The Line Spectral Frequencies (LSF) or Line Spectrum Pairs (LSP) introduced by Itakura are an alternative LPC spectral representation with several convenient properties (ordering/interlacing, independence, dynamic range) which have been examined closely in the context of LPC quantization. As a result of these properties, an LSP vector can be interpreted as a generalized pmf of vocal tract resonances, and so application of the *entropy functional* produces intuitive results. High values of the "line-spectral entropy" indicate a flat spectrum, and low values indicate a textured spectrum.

The relative entropy between two pmfs is defined by

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}. \quad (1)$$

Considering the pmfs in (1) to be derived from LSPs (as generalized pmfs), the relative entropy can provide some indication of the similarity between two spectral envelopes. This leads to some interesting interpretations for the selection of optimal paths to minimize distortion and detection of change-points in speech waveforms. In this case, the relative entropy provides a measure of *stationarity* for the AR process estimates which have been derived from local segments of speech data. Since $D(p||q)$ is minimized by $q \approx p$, a small value of $D(p_i||p_{i-1})$ (where the subscript indicates a frame time) indicates a slowly varying spectrum whereas large values indicate a rapidly changing spectral envelope. This measure can be applied to any subset of elements of the LSPs to determine the rate of evolution of that group of resonances. Also, if we assume for each i that the spectrum of the current (i^{th}) frame has evolved in one frame time from complete whiteness,

$$D(p_i||p_{i-1}) = \log m - H(p_i) \quad (2)$$

since p_{i-1} is the uniform distribution of m resonances. So, the line-spectral entropy can be seen as a particular interpretation of relative entropy which measures the spectral evolution with respect to whiteness at each frame time.

III. Acknowledgement

This work was supported in part by NSF Grant NCR-9303805.

References

[1] J. D. Gibson, S. P. Stanners, and S. A. McClellan, "Spectral entropy and coefficient rate for speech coding," in *Conf. rec. 27th Annual Asilomar Conf.*, (Pacific Grove), pp. 925–929, November 1993.

[2] S. A. McClellan and J. D. Gibson, "Spectral entropy: An alternative indicator for rate allocation?," in *Proc. IEEE ICASSP-94*, (Adelaide), pp. I.201–I.204, April 1994.

[3] S. A. McClellan and J. D. Gibson, "Variable rate tree coding of speech," in *Proc. IEEE Wichita Conf. on Comm., Netw., and Sig. Proc.*, (Wichita), pp. 134–139, April 1994.

[4] S. A. McClellan and J. D. Gibson, "Variable rate CELP based on subband flatness," in *Proc. IEEE Int'l Conf. on Communications*, (Seattle), June 1995.

Multiple-Access Channels with Correlated Sources — Coding Subject to a Fidelity Criterion

Masoud Salehi[1]
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115.

Abstract — **Characterization of the achievable distortion region, when correlated information sources are transmitted via a multiple-access channel, is studied. An inner bound for the set of achievable distortions is obtained and it is shown that certain known results in multi-terminal source and channel coding can be considered as special cases of this result.**

I. Introduction

Transmission of arbitrarily correlated information sources over a multiple-access channel was first addressed in [1], where sufficient conditions for reliable transmission were derived. However, determining the necessary conditions for this communication model still remains an open problem. In particular it has been shown that the conditions derived in [1] are not in general necessary conditions. Nevertheless, so far no other conditions that are more general than those in [1] are known.

Coding of correlated information sources subject to a fidelity criterion has been considered in a number of works including [2], [3], and [4]. This problem, in general, also remains an open problem and characterization of the rate-distortion region for this case is not yet known except for some special cases.

In this work we consider a communication model in which two correlated information sources are to be transmitted via a multiple-access channel and to be reproduced at the receiver subject to two distortion measures. We derive a set of achievable distortions for this source-channel configuration and show that many of the previously known results on transmission of correlated information sources via a multiple-access channel and rate-distortion region for correlated information sources can be considered as special cases of this result.

II. The Communication Model

Two discrete memoryless correlated information sources $\{(S_k, T_k)\}$ are modeled by independent drawings of two random variables S and T which are distributed according to $p^*(s,t)$. The corresponding alphabets are denoted by $\mathcal{S}$ and $\mathcal{T}$. A discrete memoryless multiple-access channel with two transmitters and one receiver is described in terms of its input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, the output alphabet $\mathcal{Y}$, and the conditional probability mass function $p^*(y|x_1,x_2)$.

Sources S and T are connected to the first and the second transmitters respectively. It is assumed that for each (S,T) pair generated by the sources, one (X_1,X_2) pair can be transmitted over the channel. At the receiver the decoder estimates $\{(\hat{S}_k,\hat{T}_k)\}$ as the source outputs, where $(\hat{S}_k,\hat{T}_k) \in \hat{\mathcal{S}}\times\hat{\mathcal{T}}$ and $\hat{\mathcal{S}}$ and $\hat{\mathcal{T}}$ denote the reproduction alphabets for the two sources. Two distortion functions $d_1 : \mathcal{S}\times\hat{\mathcal{S}} \to R^+$ and $d_2 : \mathcal{T}\times\hat{\mathcal{T}} \to R^+$ represent the corresponding fidelity criteria.

A distortion pair (D_1,D_2) is achievable if for any $\delta_1 > 0$ and $\delta_2 > 0$ there exist an integer n, two encoding functions $f_1 : \mathcal{S}^n \to \mathcal{X}_1^n$ and $f_2 : \mathcal{T}^n \to \mathcal{X}_2^n$, and one decoding function $g : \mathcal{Y}^n \to \hat{\mathcal{S}}^n \times \hat{\mathcal{T}}^n$ such that

$$\frac{1}{n}\sum_{k=1}^{n} E[d_1(S_k,\hat{S}_k)] \leq D_1+\delta_1$$

$$\frac{1}{n}\sum_{k=1}^{n} E[d_2(T_k,\hat{T}_k)] \leq D_2+\delta_2$$

Let $\mathcal{D} \subset R^{+2}$ denote the set of all achievable distortion pairs (D_1,D_2).

III. Main Result

Our main result is the derivation of an inner bound for the set $\mathcal{D}$ as stated in the following theorem.

Theorem: If there exist

1. Auxiliary random variables U and V taking values in finite sets $\mathcal{U}$ and $\mathcal{V}$ such that $U \to S \to T \to V$ make a Markov chain.
2. Functions $h_1 : \mathcal{U}\times\mathcal{V} \to \hat{\mathcal{S}}$ and $h_2 : \mathcal{U}\times\mathcal{V} \to \hat{\mathcal{T}}$.

such that

$$\begin{aligned} I(S;U|V) &< I(X_1;Y|X_2,V) \\ I(T;V|U) &< I(X_2;Y|X_1,U) \\ I(S,T;U,V) &< I(X_1,X_2;Y) \end{aligned}$$

for some

$$\begin{aligned} p(s,t,u,v,x_1,x_2,y) &= p^*(s,t)p(u|s)p(v|t) \\ &\quad \times p(x_1|u,s)p(x_2|v,t)p^*(y|x_1,x_2) \end{aligned}$$

and D_1 and D_2 are given by

$$\begin{aligned} D_1 &= E[d_1(S,h_1(U,V))] \\ D_2 &= E[d_2(T,h_2(U,V))] \end{aligned}$$

then $(D_1,D_2) \in \mathcal{D}$.

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. Berger, T. Housewright, J. Omura, S. Tung, and J. Wolfowitz, "An upper bound for the rate distortion function for source coding with partial side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-25, pp. 664–666, November 1979.

[3] A. H. Kaspi and T. Berger, "Rate distortion for correlated sources with partially separated encoders," *IEEE Transactions on Information Theory*, vol. IT-28, pp. 828–840, November 1982.

[4] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Transactions on Information Theory*, vol. IT-35, pp. 228–236, March 1989.

[1] This work was partially supported by the NSF Grant NCR-9101560.

Coding For Channels With Cost Constraints

Ali S. Khayrallah
EE Department
University of Delaware
Newark, DE 19716, USA

David L. Neuhoff
EECS Department
University of Michigan
Ann Arbor, MI 48109, USA

Abstract — **We address the problem of finite-state code construction for the costly channel. Adler et al. developed the powerful state-splitting algorithm for use in the construction of finite-state codes for hard-constrained channels. We extend the state-splitting algorithm to the costly channel. We present several examples of costly channels related to magnetic recording, the telegraph channel, and shaping gain in modulation. We design a number of synchronous and asynchronous codes, some of which come very close to achieving capacity.**

I. Introduction

In a costly channel, sequences of symbols are assigned costs (possibly infinite). A constraint in the form of an average cost is imposed on the sequences. A costly channel is a natural generalization of a hard-constrained channel (or subshift), where sequences are assigned either cost zero or infinity. Many hard-constrained channels of interest have a finite-state structure, and can be represented by finite directed graphs. Similarly, finite-state costly channels can be represented by finite directed graphs with an additional cost labeling.

We present a method for constructing finite-state codes for the costly channel. Our finite-state codes come in two varieties. The first is a synchronous (fixed-length to fixed-length) code. The second is an asynchronous (variable-length to fixed-length) code. The latter has a higher rate, but it has the drawbacks common to all asynchronous schemes, in particular the potential for error propagation. At the heart of our method is a modified version of the state splitting algorithm of Adler, Coppersmith, and Hassner. The capacity-cost function $C(\rho)$ is the maximum code rate for a given target cost ρ. Our asynchronous codes come very close to achieving $C(\rho)$, while the synchronous codes achieve a lower rate, but still come pretty close to $C(\rho)$. Given a graph G representing the costly channel, $C(\rho)$ is achieved by a Markov chain defined on the edges of G. We associate with G a modified adjacency matrix $\mathbf{B}$ that reflects the target cost ρ. Then

$$C(\rho) = \log \lambda + \mu \rho \log e$$

where λ is the largest eigenvalue of $\mathbf{B}$, and $\mu = dC/d\rho$.

II. Code Construction

Our construction is summarized as follows. For a given ρ, we choose $n \geq 1$ and $m \geq 1$ such that m/n does not exceed a function related to $C(\rho)$. Then we construct an asynchronous encoder graph with power n and with smallest state outdegree equal to 2^m (and every outdegree a power of 2), and consequently, whose rate exceeds m/n. We also construct a synchronous encoder with every outdegree equal to 2^m, and rate m/n.

We assume that the information source is binary IID with a uniform distribution. The source induces a stationary Markov chain on the encoder graph, where the edges leaving a state have a uniform conditional probability. It also yields a coding rate, and a coding cost.

The idea of the code construction is to obtain an encoder graph such that the source-induced Markov chain coincides with the optimal Markov chain that achieves capacity. Then the code will actually achieve capacity. It turns out that in most cases, we can only approximate the optimal Markov chain, but the resulting codes are still very good.

Our construction consists of three stages. It uses state splitting and edge pruning. First, we use state splitting to obtain a uniform cost graph, that is one where all edges leaving a state have the same cost. Secondly, we use state splitting in a way similar to Adler et al. Let $\mathbf{v}$ denote the eigenvector corresponding to λ. We perform a sequence of state splittings to obtain a graph whose $\mathbf{v}$ (or a related approximate eigenvector $\mathbf{x}$) is equal to the all ones vector. Thirdly, we use edge pruning to obtain a graph with the appropriate state outdegrees. Edge pruning must be done carefully, since it affects both the coding rate and the coding cost.

III. Examples

We introduce costly channel models for magnetic recording, namely variations on the (1,3) and (2,7) hard-constrained channels. We also consider Shannon's telegraph model as a costly channel, and relate his definition of capacity to our capacity-cost function. Finally we show an application of our technique to the problem of shaping in amplitude modulation. Our codes are consistently good, and several almost achieve capacity. Their complexity is low, judging by the number of encoder states.

On the Capacity of M-ary Run-Length-Limited Codes[1]

Steven W. McLaughlin[2], Jian Luo[2] and Qun Xie[3]

Abstract — **We present two results on the Shannon capacity of M-ary (d,k) codes. First we show that 100-percent efficient fixed-rate codes are impossible for all values of (M,d,k), $0 \le d < k < \infty$, $M < \infty$, thereby extending a result of Ashley and Siegel to M-ary channels. Second, we show that (unlike the binary case) for $k = \infty$, there exist an infinite number of 100-percent efficient M-ary (d,k) codes and we construct one such code.**

I. Summary

Traditional magnetic and optical recording employ saturation recording, where the channel input is constrained to be a binary sequence satisfying run-length limiting (RLL) or (d,k) constraints. A binary (d,k) sequence is one where the number of zeroes between consecutive ones is at least d and at most k.

The recording media in [1] supports unsaturated, M-ary ($M \ge 3$) signaling while requiring that run-length-limiting constraints be satisfied. Assuming an M-ary symbol alphabet, $A = \{0, 1, \ldots, M-1\}$, $M < \infty$, an M-ary run-length-limited or (M,d,k) sequence [2] is one where at least d and at most k zeroes occur between nonzero symbols. Binary (d,k) codes are M-ary (d,k) codes with $M = 2$.

In [3] (and the applicable corrections in [4]) it was shown that for binary (d,k) codes, there exist no 100-percent efficient codes. Specifically, the Shannon capacity of all binary RLL (d,k) constraints is irrational for all values of (d,k), $0 \le d < k \le \infty$, and hence, there exist no fixed rate codes that achieve capacity. In this paper we present two propositions on (M,d,k) codes. First, for any integer M, 100-percent efficient fixed-rate codes are impossible for all values of (d,k), $0 \le d < k < \infty$, thereby extending [3] to the M-ary channel. Secondly, unlike [3], for $k = \infty$ there do exist (an ∞ number of) 100-percent efficient codes, and we construct one such code using the state splitting algorithm [5].

The RLL (M,d,k) constraint is often represented by a finite state transition diagram (FSTD) G. Associated to the FSTD with $k+1$ vertices is a state-transition matrix T, a $(k+1) \times (k+1)$ matrix defined by $T = [t_{ij}]$ where t_{ij} is the number of edges in G from state i to state j.

Shannon showed that if the FSTD G has distinct labels on the outgoing edges of each state, the capacity of a system constrained by sequences from G is $C = \log_2 \lambda$ (bits per symbol) where λ is the largest real eigenvalue of the adjacency matrix T associated with G. We have assumed base-2 logarithms because it is the most common case, but all results are extendable to non-base-2 logarithms.

We consider fixed rate $r = m/n$ encoders that map m user bits to n channel symbols satisfying the (M,d,k) constraints. The capacity C is therefore an upper bound to all achievable rates r, and the code efficiency $E = r/C$ is the ratio of the actual coding rate to the largest rate achievable. We give the following two propositions and one new capacity achieving code. Denote the binary capacity of the (M,d,k) constraint as $C(M,d,k)$.

Proposition 1: $C(M,d,k)$ is irrational for all (M,d,k), $0 \le d < k < \infty, M < \infty$.

Since this capacity is irrational, there exist no 100 percent efficient fixed rate $r = m/n$ codes, namely $r < C(M,d,k)$.

Proposition 2: For any $0 \le d < \infty$, and $k = \infty$ the set of M's for which $C(M,d,\infty)$ is rational is $\{M : M = 2^{dm}(2^m - 1) + 1, \text{integer } m \ge 1\}$.

Since this capacity is rational for some M, there exist 100 percent fixed rate $r = m/n = C$ codes that achieve capacity $C(M,d,k)$. What follows is a construction of one such code satisfying $(5,2,\infty)$ constraints using the state splitting algorithm. For details on the state splitting algorithm see [5].

$(5,2,\infty)$ *code*: The adjacency matrix for the $(5,2,\infty)$ is

$$T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & 0 & 1 \end{bmatrix} \tag{1}$$

with capacity $C = 1$. Choosing $m = n = 1$ one can design a rate $r = m/n = 1 = C$ code. An approximate eigenvector $\mathbf{v}$ satisfying $T\mathbf{v} \ge 2\mathbf{v}$ is $\mathbf{v} = (1,2,4)^T$. After two rounds of splitting and some simple merging a five-state encoder results [6]. A sliding block decoder with a sliding window six symbols wide (corresponding to memory $m = 3$ and anticipation $a = 2$) is sufficient.

II. Acknowledgment

The authors gratefully acknowledge Optex Communications Corporation whose support initiated this work.

References

[1] A. Earman "Optical data storage with electron trapping materials using M-ary data channel coding," *Proceedings of the SPIE 1663, Optical Data Storage*, pp. 92-103, San Jose, CA, 1992.

[2] D.T. Tang and L.R. Bahl, "Block Codes for a Class of Constrained Noiseless Channels," *Information and Control*, vol. 17, 1970, pp. 436-461.

[3] J. Ashley and P. Siegel, "A Note on the Shannon Capacity of Run-length Limited Codes," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 4, pp. 601-605, July 1987.

[4] J. Ashley, M. Hilden, P. Perry and P. Siegel, "Correction to "A Note on the Shannon Capacity of Runlength-Limited Codes"," *IEEE Trans. Inform. Theory*, vol. IT-39, no. 3, pp. 1110-1112, May 1993.

[5] R. Adler, D. Coppersmith and M. Hassner, "Algorithms for sliding block codes," *IEEE Trans. Info. Thy* vol. 29, no.1, pp. 5-22, Jan. 1983.

[6] S. McLaughlin, J. Luo and Q. Xie "On the capacity of M-ary run-length-limited codes," to appear *IEEE Trans. on Information Theory.*

[1]This work was sponsored by the National Science Foundation under award no. NCR-9309008.

[2]Electrical Engineering Dept., Rochester Institute of Technology, Rochester, NY 14623.

[3]Mathematics Dept., University of Rochester, Rochester, NY 14627.

Joint Multilevel RLL and Error Correction Coding

Mohamed Siala and Ghassan Kawas Kaleh
Ecole Nationale Supérieure des Télécommunications,
46, rue Barrault, 75634 Paris 13, France

Runlength-limited (RLL) codes, also known as (d,k) RLL codes, are used in digital magnetic recording and have potential use in Soliton optical communication. Let 0^m denotes a sequence of m successive zeros. We define the alphabet:

$$\mathcal{H}_{dk} = \left\{\text{“}0^d1\text{”}, \text{“}0^{d+1}1\text{”}, \ldots, \text{“}0^k1\text{”}\right\}$$

A (d,k) phrase is an element in $\mathcal{H}_{dk}$. A (d,k) runlength-limited sequence is defined as the concatenation of such phrases.

We have presented in ISIT'94 a new approach for constructing simple and efficient variable-length (d,k) RLL codes which can be decoded with no memory and no anticipation. An n-dimensional RLL code $\mathcal{C}_{dk}^{nL}$ is defined as

$$\begin{aligned}\mathcal{C}_{dk}^{nL} &= \{p=(p_1,p_2,\ldots,p_n)\in(\mathcal{H}_{dk})^n : \\ &\quad l(p_1)+l(p_2)+\cdots+l(p_n)\le nL\}\end{aligned}$$

where $l(p)$ is the length of an element p in $\mathcal{H}_{dk}$ and L is a normalized threshold.

We show here that if the size of $\mathcal{H}_{dk}$, $k-d+1$, is equal to b^m for some arbitrary integers b and m all greater than or equal to 2, the encoding/decoding algorithms can be greatly simplified, using m parallel simple codes, called component codes, with the same properties as the original code $\mathcal{C}_{dk}^{nL}$.

The block diagram of the proposed coding system is depicted in the Figure. A binary information sequence I is demultiplexed into m binary subsequences $I_0, I_1, \ldots, I_{m-1}$. Each subsequence is encoded by an independent b-ary shaping set, denoted by $\mathcal{S}_i, i=0,1,\ldots,m-1$. The rate of the code $\mathcal{S}_i$ is $R_i = k_i/n_i$. Its output b-ary n_i-tuples $x_j^i, j=\ldots,-1,0,1,\ldots$, are concatenated into a b-ary infinite sequence x^i. Let Z_b denote the b-ary alphabet $\{0,1,\ldots,b-1\}$, and x_t^i the output (in Z_b) of the encoder for $\mathcal{S}_i$ at time t, where t is an integer. The b-ary symbols $x_t^0, x_t^1, \ldots, x_t^{m-1} \in Z_b$, are mapped synchronously onto a phrase length

$$l_t \triangleq d+1+\sum_{i=0}^{m-1} x_t^i b^i, \qquad (1)$$

in

$$\mathcal{J}_{dk} \triangleq \{d+1, d+2, \ldots, k+1\}.$$

The phrase length l_t is then mapped into the phrase p_t, in $\mathcal{H}_{dk}$, with length equal to l_t. The resulting phrase p_t is subsequently recorded on the magnetic media or transmitted using soliton pulses.

To simplify the notations, we assume that the readback signal is available after a null time delay. The phrase at time t in the readback (for recording systems) or received (for fiber optic transmission using solitons) signal, denoted by $\hat{p}_t$, is assumed to be in $\mathcal{H}_{dk}$. Denote by $\hat{l}_t$ the length of $\hat{p}_t$. The inverse mapper outputs the estimates $\hat{x}_t^i \in Z_b, i=0,1,\ldots,m-1$, verifying

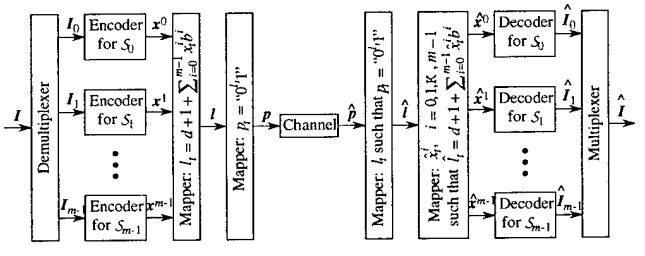
$$\sum_{i=0}^{m-1} \hat{x}_t^i b^i = \hat{l}_t - (d+1),$$

to the decoding circuit. The estimates $\hat{x}_i$ of the codewords x^i are obtained from the final estimated sequences $\{\ldots, \hat{x}_{t-1}^i, \hat{x}_t^i, \hat{x}_{t+1}^i, \ldots\}$, $i=0,1,\ldots,m-1$. These sequences are split into the n_i-tuples $\hat{x}_j^i, j=\ldots,-1,0,1,\ldots$. These n_i-tuples, which are estimates of the n_i-tuples x_j^i at the encoder side, are concatenated into the estimate $\hat{x}^i$ of the sequence x^i at the output of the shaping sets $\mathcal{S}_i$.

The decoders for $\mathcal{S}_i$ use these sequence estimates and deliver the estimates $\hat{I}_i$, $i=0,1,\ldots,m-1$, of the component information sequences I_i. The estimate $\hat{I}$ of the desired final information sequence I is constructed by combining the digits of the sequences $\hat{I}_i$, $i=0,1,\ldots,m-1$.

Moreover, we show that this approach to RLL coding can naturally incorporate multilevel error correction coding. Recall that the most common type of errors in digital magnetic recording are: Shift errors, Drop-in errors, Drop-out errors, Insertion errors and Deletion errors. Although deletions and insertions are not as common as the other types of errors, they can involve a catastrophic propagation of errors when using a conventional rate p/q finite state encoder , since they can change q, the length of the (d,k) constrained sequences generated by this code. Deletions and insertions of zeros ("0") as well as shift errors do not change the number of phrases in the (d,k) RLL codewords. From this, we conclude that our approach to RLL coding allows us not only to correct shift errors but also insertions and deletions of zeros. The main idea for achieving this is to use Lee-metric codes. The component codes of the multilevel block code are chosen such that this code has a large minimum Lee distance and such that the (d,k) RLL code combined with the error correction multilevel code provides the highest possible rate. We emphasize that multilevel (d,k) RLL coding combined with multilevel block coding for error correction is well suited to multistage decoding .

We also show that for an appropriate choice of the normalized thresholds of these parallel codes, the rate of the corresponding (d,k) RLL code converges to the capacity of the (d,k) constraint, as the dimension of each of the parallel codes goes to infinity.



(4, 20) Runlength Limited Modulation Code for High Density Storage System

Min-Goo Kim[1] and Jae Hong Lee

Department of Electronics Engineering Seoul National University
San 56-1, Shinlim-dong, Kwanak-ku, Seoul 151-742, Korea

Abstract - **In magnetic and optical storage systems, RLL (run length limited) modulation code is widely used. An RLL code having a limited amount of error propagation and high density ratio is attractive for high density storage system. This paper presents a new fixed-length RLL modulation code with *d=4* and *k=20*, and coderate *m/n=4/11*. A code design is based on the *Look-Ahead* method. This code has finite error propagation limited to at most two codewords.**

I. INTRODUCTION

To increase storage-capacity, runlength limited modulation code has been widely used for data storage systems such as magnetic disk/tape and optical disk, because RLL code has density ratio (DR) larger than 1. Most RLL codes for practical storage systems have low DR. EFM (Eight to Fourteen Modulation) code and (1, 7), (2, 7) RLL codes which are widely used for contemporary storage systems have DR less than 1.5. Thus these codes are less attractive for high density storage system. As storage system evolves to high density storage system, it is necessary to design an RLL code with high DR and low hardware complexity and encoding/decoding delay[1].

(d, k) RLL code is a binary sequence, where the number of "zeros" between consecutive "ones" must be at least d and at most k. Density ratio is determined by d and coderate $R=m/n$ where m is information length and n is codelength. DR increases as d increases. However DR is limited by C which is the capacity of the *(d, k)* constrained noiseless channel[1], [2]. Many class of design techniques have been proposed. Alder, Coppersmith, and Hassner showed that Sliding-block code algorithm can produce an *m/n (d, k)* RLL code if $R(=m/n) < C$ for positive integers m and n[3]. Jacoby and et al. developed and have evolved Look-Ahead coding method (LA) which is attractive for practical consideration, if m and n are small[4]. RLL code has error propagation and it is not always simple to achieve the minimum amount of error propagation. Blaum showed that the error propagation is a major issue in storage system adopting error correcting codes[5].

In this paper, we suggest a new fixed-codelength (4, 20) RLL code with coderate *R=4/11* which is based on Look-Ahead coding method with full-codelength look-ahead. We will show that (4, 20) RLL code has error propagation limited to at most two codewords and DR of *20/11*.

II. (4, 20) RLL MODULATION CODE

For high density storage system, it is necessary for RLL code to have high DR, small error propagation, and low *(k+1)/(d+1)*[1]. In NRZI recording scheme, DR is *(m/n)(d+1)*. It was shown that for positive integers m and n, a *(d, k)* RLL code with $R=m/n$ exists if $R<C$, where C is the capacity of *(d, k)* constrained noiseless channel. Channel capacity of (4, ∞) RLL code is 0.4056. Therefore an RLL code with *d=4* and *R=4/11* is feasible[1], [2].

(4, 20) RLL code is a fixed codelength RLL code with R=4/11 and it translates information block of 4 bits into a codeword block of 11 bits. (4, 20) RLL code is composed of 16 codewords. All codewords satisfy *d=4* constraint. There is always the possibility that consecutive codewords violate *d=4* constraint. When *d=4* constraint is violated, we require substitutions in order to eliminate successive "ones". There are three cases of violation. Let a *precursive codeword* $P=(p_1, p_2, \ldots, p_{11})$ be the first codeword and a *successive codeword* $S=(s_1, s_2, \cdots, s_{11})$ be the second codeword among consecutively two codewords. In TABLE I, three violation cases are given where *'x'* denotes don't care bit. Precursive and successive codewords are substituted by Rule I, II, III when violation cases occur.

TABLE I. CASES OF VIOLATION AND SUBSTITUTION RULES

Case	Check-Bits for Violation	Rule
	$p_7\ p_8\ p_9\ p_{10}\ p_{11}\ s_1\ s_2\ s_3$	
(1)	x x 1 0 0 : 0 1 x	Rule I
(2)	x x 1 0 0 : 1 x x	Rule II
(3)	0 1 0 0 0 : 1 x x	Rule III-1 (P_3='0')
		Rule III-2 (P_3='1')

Density ratio of (4, 20) RLL code is 20/11 which is 38% greater than (1, 7) RLL code and 29% greater than EFM code. Also *(k+1)/(d+1)* is 21/5 which is similar to that of (1, 7) RLL code. (4, 20) RLL code has finite error propagation limited to at most two codewords because all substituted codewords have always "five zeros" at the position of 7, 8, 9, 10, 11*'th* bits. Thus encoding is completed by considering only two consecutive codewords and it is impossible for errors on a codeword to propagate into codewords more than two.

III. CONCLUSIONS

We designed a fixed-codelength (4, 20) RLL modulation code for high density storage systems. Density ratio of (4, 20) RLL code is greater than (1, 7) RLL code and EFM code. Also it has finite error propagation limited to at most two codewords. (4, 20) RLL code has low complexity of hardware and it is feasible to be implemented by look-up table of small size.

[1]This work was supported by a grant from the SAIT (Samsung Advanced Institute of Technology), Korea.

REFERENCES

[1] P. H. Siegel, "Recording codes for digital magnetic storage," *IEEE Trans. Magnet.*, vol. MAG-21, no. 5, pp. 1344-1349, Sept. 1985.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell syst. Tech J.*, vol. 27, pp. 379-656, 1948.

[3] R. L. Alder, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes. An application of symbolic dynamic to information theory," *IEEE Trans. Inform. Theory.*, vol. IT-29, pp. 5-22, Jan. 1983.

[4] G. V. Jacoby and R. Kost, "Binary two-third rate code with full word look-ahead," *IEEE Trans. Magnet.*, vol. MAG-20, no. 5, pp. 709-714, Sept. 1984.

[5] M. Blaum, "Combining ECC with modulation: Performance comparisons", *IEEE Trans. Inform. Theory.*, vol. IT-37, no. 3, pp.945-949, May 1991.

On Properties of Binary Maxentropic DC-free Runlength-Limited Sequences

Volker Braun

Institute for Experimental Mathematics, Ellernstr. 29, 45326 Essen, Germany

Abstract — **We present the most remarkable results obtained from a numerical analysis of relevant statistical properties of binary maxentropic DC-free runlength-limited (DCRLL) sequences. In particular, we consider the sum variance and its relation to the low-frequency characteristic or the redundancy. Further, we present an approximation of the runlength distribution of binary maxentropic pure charge constrained sequences.**

I. Introduction

Binary DC-free runlength-limited (DCRLL) sequences are widely applied in digital storage systems, for example in the CD player [2]. The fact that there is still no profound knowledge of the relevant statistical properties of these sequences motivated us to investigate these properties for a wide range of constraints. We will briefly present the most remarkable results obtained in the maxentropic case. We characterize DCRLL sequences by three integer parameters (d, k, N), denoting that the runlengths occurring in these sequences are constrained between $d+1$ and $k+1$, and the charge or running digital sum (RDS) assumes N distinct values. Note that we consider sequences of symbols drawn from $\{-1, 1\}$, and that the constraints satisfy $0 \leq d < k \leq N-2$. In order to represent the (d, k, N) constraints we use the concept of runlength graphs described by Kerpez et al. [1], and we interpret a maxentropic DCRLL sequence as generated by a stationary Markov chain based on such a graph. This Markov chain description allows the evaluation of the power spectral density function $H(\omega)$, the sum variance $\sigma_z^2(d, k, N)$ (i.e. the variance of the RDS), the runlength distribution, and the average runlength of maxentropic DCRLL sequences [1].

II. The Main Numerical Results

The analysis of the sum variance $\sigma_z^2(d, k, N)$ in the practically interesting range of constraints, $(0 \leq d \leq 2, d < k \leq N-2, 9 \leq N < 30)$, reveals that it is in good approximation determined by N, and roughly independent of the d and k constraints. Hence, we can approximate $\sigma_z^2(d, k, N)$ by the known expression for the sum variance of maxentropic pure charge constrained sequences, i.e., $\sigma_z^2(d, k, N) \approx (1/12 - \pi^{-2}/2)(N+1)^2$ [2]. For the analyzed range of (d, k, N) constraints, this approximation is within 5% accuracy as long as $k - d > \lfloor N/2 \rfloor$. For k constraints only slightly larger than d, the true sum variances $\sigma_z^2(d, k, N)$ are somewhat less than the above approximation.

In the case of maxentropic pure charge constrained sequences, there is a simple relation between sum variance and low-frequency characteristic [2]. We are interested whether a corresponding relation exists for maxentropic DCRLL sequences. We express the low-frequency characteristic by a well-defined cut-off frequency. In order to provide a clear physical interpretation, we define the *cut-off frequency* ω_c of a maxentropic DCRLL sequence by $H(\omega_c) = H_0(d, k)/2$, where $H_0(d, k)$ denotes the DC-content of a maxentropic runlength-limited sequence with parameters (d, k) [2]. Indeed, for $N \gg 1$ we could find the relation $\omega_c \approx H_0(d, k)[2\sigma_z^2(d, k, N)]^{-1}$ between sum variance and cut-off frequency. For d constraints $0 \leq d \leq 2$, this approximation is within 10% accuracy as $N > 17$. We conclude that for N sufficiently large the sum variance $\sigma_z^2(d, k, N)$ is a useful criterion of the low-frequency characteristic of a maxentropic DCRLL sequence, a fact which again justifies the definition of the cut-off frequency ω_c.

As shown in [2], maxentropic pure charge constrained sequences have the fundamental property that the product of sum variance and redundancy is approximately constant. By introducing a refined redundancy definition, it turns out that for $N \gg 1$ a corresponding relation also holds for maxentropic DCRLL sequences. Let the *extra redundancy* be defined as $C(d, k, \infty) - C(d, k, N)$, where $C(d, k, N)$ denotes the capacity of the (d, k, N) constraint. In other words, the extra redundancy describes the increment in redundancy from the (d, k) runlength constraint to the (d, k, N) constraint. For maxentropic DCRLL sequences, we found that the product of sum variance and extra redundancy for $N \gg 1$ assumes a constant value which is determined by the d and k constraints. In the absence of a specific k constraint (i.e. $k = N-2$) and for d constraints $0 \leq d \leq 2$, for example, this sum variance-extra redundancy product appears to be constant for about $N > 20$.

III. A Neat Analytical Result

Kerpez et al. [1] present a closed-form expression for the Markov chain description of a maxentropic pure charge constrained sequence, where the constraint is represented by a runlength graph. Using this result, we are able to derive a closed-form expression for the runlength distribution of such a sequence. For $N \gg 1$, we can substitute the sums occurring in this expression by integrals which can be solved using some trigonometric manipulation. In this way, for $N \gg 1$ we approximately obtain the probability of occurrence of a run of length l in such a sequence by $Pr(l) \approx k_N(l)\lambda^{-l}$, where $l \in \{1, 2, \ldots, N-1\}$, $\log_2 \lambda$ denotes the capacity of the charge constraint (i.e. $\lambda = 2\cos[\pi(N+1)^{-1}]$), and $k_N(l) = (l-N+1)(N+1)^{-1}\cos[\pi(N-1+l)(N+1)^{-1}] + \pi^{-1}\sin[\pi(N-1-l)(N+1)^{-1}]$.

Acknowledgements

The author would like to thank K.A. Schouhamer Immink for useful discussions and suggestions.

References

[1] K.J. Kerpez, A. Gallopoulos, and C. Heegard, "Maximum Entropy Charge-Constrained Run-Length Codes," *IEEE Journal on Selected Areas in Comm.*, vol. 10, no. 1, pp. 242-252, January 1992.

[2] K.A. Schouhamer Immink, *Coding Techniques for Digital Recorders*, Prentice Hall International (UK) Ltd, 1991.

Coding for low complexity detection of multiple insertion/deletion errors

WA Clarke and HC Ferreira*

ESKOM NPTM&C Simmerpan, P O Box 107, Germiston, 1400, South Africa

Abstract - **In this paper two new methods for the detection of multiple insertion/deletions are presented. The first method recognises insertions/deletions in the previous symbol stream by extracting additional information from commonly used markers. A new coding method is also presented that relies on the number of transitions in a codeword to detect insertions/deletions in the previous codeword. This coding scheme also has certain spectral properties.**

I. INTRODUCTION

The insertion/deletion of symbols in a codeword result in a change in the length of the word. As a result the frame alignment is lost. One should make a distinction between the above case and the case of additive errors where only certain symbols in codewords are changed.

Recently [1], a coding algorithm was developed that generates codes with the ability to correct several insertions/deletions, assuming that the codeword boundaries were known.

Two approaches are presented in this paper. In section II a marker method is described that enables the receiver to detect insertions/deletions in the previous frame. This is the first step to correct insertions/deletions. In section III, a simple coding method is presented to detect insertions/deletions in every codeword, utilising the number of transitions in every codeword.

II. MARKERS

Markers (a known sequence of symbols) are used to delineate a stream of symbols into frames. In [2] a comprehensive overview on markers is presented. Additional information can be extracted from commonly used markers for the detection of insertions/deletions. If an insertion/deletion occurred in the preceding frame, the marker is shifted left or right. The decoder recognises a shifted version of the marker, as well as unkown adjacent symbols from the data stream, in the expected marker position.

Markers are chosen in such a way that the resulting sequence as described above are uniquely recognisable. All possible resulting sequences are stored in a lookup table which enables the decoder to detect a) that insertions/deletions occurred, and b) the number of shifts.

The functionality of the proposed scheme can be extended to include the detection of additive errors in the marker.

III. INSERTION/DELETION DETECTING CODE

A new insertion/deletion detecting code is introduced that relies on a constant number of transitions from 0 to 1 or 1 to 0 in the symbols of the codeword. Insertions/deletions in the preceding codeword are detected.

Each codeword consists of three sections: a head, middle and tail section. The middle section consists of a constant number of transitions while the head and tail sections act as buffers.

Insertions/deletions in the preceding codeword introduce shifts. The codewords are chosen in such a way that the transitions of the middle section increase or decrease with left and right shifts. The decoder counts the number of transitions in the middle section of the expected codeword. In this way the decoder recognises the occurrence of insertions/deletions in the previous symbol stream and can correct the frame alignment.

These codes are very flexible and good rates can be obtained. The implementation of these codes is easy and only a few simple logic gates are necessary to enable the detection of insertions/deletions. Lookup tables are not required for the detection of insertions/deletions as only transitions are monitored. These codes can be used to maintain frame synchronisation. If the synchronisation is lost, only a few codewords have to be examined to resynchronise as opposed to a number of frames in the case of only a marker being used once in a frame of a few hundred symbols.

As a result of the use of transitions, certain spectral density properties are obtained. The lower the number of transitions, the lower the peak energy content will be in the power spectral density of the code and vice versa. As a result, these codes can be useful for both insertion/deletion detection and spectral shaping.

IV. CONCLUSION

In this paper two methods were presented to detect the occurrence of insertions/deletions. Both methods are simple to decode and are of low complexity.

V. REFERENCES

[1] ASJ Helberg, 'Multiple Insertion/deletion correcting codes', Submitted to the *IEEE Transactions on Information Theory*.

[2] P Bylanski and DGW Ingram, *Digital Transmission Systems*, Chapter 5, Peter Peregrinus, Second Edition, England, 1988.

[3] WA Clarke and HC Ferreira, 'Multiple marker sequences for the detection of insertion/deletion errors', *Proceedings of the 4th Benelux-Japan Workshop on Coding and Information Theory*, Eindhoven, pp. 14, June 1994.

* Lab for Cybernetics, Rand Afrikaans University, P O Box 524, Aucklandpark, 2006, Johannesburg, South Africa.

Single-Error-Correcting Codes for Magnetic Recording

L.B. Levitin
College of Engineering
Boston University
Boston, MA 02215

F.S. Vainstein
Dept. Electrical Engineering
NC A&T State University
Greensboro, NC 27411

Abstract — **A construction is suggested of a code which corrects single bit-shift errors in (d,k) modulation codes. The codes are nearly optimal in redundancy. The encoding and decoding procedures are linear in the codeword length.**

Magnetic and magneto-optical data recording uses a transition from one direction of magnetization to another to represent 1, and an absence of transition to represent 0. Due to physical and technological reasons the number of zeroes between two successive transitions is limited by a minimum d and a maximum k. Codes that satisfy these constraints are known as (d,k) run-length-limited modulation codes [1, 2].

An important type of errors in magnetic recording is a shift of the border between two magnetic domains, i.e. a displacement of the position of 1 (which cannot be corrected by the usual write precompensation technique). These are so called bit-shift errors [3]. Usual error-correcting codes such as for the binary symmetric channel are not well-suited for this type of errors. This paper sugests constructions of codes correcting single diplacement errors.

Let n be the length of codewords in a (d,k) code. Then the maximum number of ones (nonzero components) in a codeword is $m = \lfloor n/d \rfloor$.

Consider first the case when a non-zero component can only be shifted by one position to the left or to the right. Denote by x_i the position of the i-th nonzero component $(1 \le x_i \le n, 1 \le m)$. Now on top of modulation (d,k)-constraints, we will require that a codeword should satisfy the following condition:

$$S_i = \sum_i i x_i = 0 \pmod{2m+1} \tag{1}$$

where the sum is taken over all nonzero components of a codeword. Generation of codewords which satisfy condition (1) can be conveniently combined with satisfying modulation constraints by modification of the inverse enumeration algorithm suggested by Fitingof [4].

The sum S_i is the syndrom. If $S_i \le m$, then $S_i = i$, where i is the number of the nonzero component shifted to the right. If $S_i \ge m+1$ then $2m+1-S_i = i$, where i is the number of the nonzero component shifted to the left. Thus, error correction is quite simple.

The code is nearly optimal. Indeed, since condition (1) and modulation (d,k) -constraints are independent, one can expect that the number of codewords which satisfy (1) is smaller than the size of the (d,k) modulation code by a factor of $2m+1$. But $2m+1$ is the maximum number of possible errors to be corrected (including the error-free case). The deviation from optimality is due to the fact that the actual number of possible errors in a given codeword can be smaller than the maximum, because of a smaller number of nonzero components.

Consider now a more general case, when one of the nonzero components can be displaced up to r positions to the right or to the left. Thus, the displacement g is:

$$-r \le g \le r \tag{2}$$

Construct a sequence of natural numbers (p_i), $1 \le i \le m$ in the following way:

1. $p_1 = 1$, $p_2 = r+1$
2. p_i is relatively prime with $p_1, p_2, \ldots, p_{i-1}$

The codewords should satisfy the following condition:

$$S_i = \sum_i p_i x_i = 0 \pmod{2rp_m + 1} \tag{3}$$

The encoding procedure is similar to that for the case $r = 1$. The error-correcting properties of the code are based on the following theorem.

Theorem 1 *For any two distinct single errors, i.e. for any two displasements: g of the i-th nonzero component and h of the j-th nonzero component, the syndromes are different:*

$$S_r(g,i) \ne S_r(h,j) \tag{4}$$
$$(-r \le g \le r,\ -r \le h \le r,\ 1 \le i \le m,\ 1 \le j \le m).$$

The correction is still simple:

1. Calculate $S'_r = \begin{cases} S_r & \text{if } S_r \le rp_m \\ S_r - 2rp_m - 1 & \text{if } S_r > rp_m \end{cases}$
2. Find the largest $g \le r$ such that $\frac{S'_r}{g}$ is an integer and $\frac{S'_r}{g} > r$.

Then $|\frac{S'_r}{g}| = p_i$, where i is the number of the displaced component, and $g \cdot \operatorname{sgn} S'_r$ is the displacement.

Since the enumeration and inverse enumeration algorithms for (d,k) decoding and encoding suggested in [4] have linear (in the length of codewords) complexity, it follows that the same is true for our error-correcting code.

References

[1] C.D. Mee and E.E. Daniel, Eds., *Magnetic Recordings*, vol 2: *Computer Data Storage*, MeGraw-Hill, New York, 1988.

[2] B.H Marcus, P.H. Sigel, and J.K. Wolf, "Finite-State Modulation Codes for Data Storage", *IEEE Journal on Selected Areas in Communications*, vol. 10, No.1, pp. 5-37, 1992.

[3] A.S. Hoagland and J.E. Monson, *Digital Magnetic Recording*, Wiley&Sons, New York, 1991.

[4] B. Fitingof, *Method and Apparatus for Providing Maximum Rate Modulation or Compression Encoding and Decoding*, U.S. Patent #5,099,237, Mar. 24, 1992.

Single-Track Gray Codes

Alain P. Hiltgen,
Crypto AG,
P.O. Box 474,
CH-6301, Zug,
Switzerland.

Kenneth G. Paterson[1],
Department of Mathematics,
Royal Holloway,
University of London,
Surrey TW20 0EX, U.K.

Marco Brandestini,
Brandestini Design Ltd.,
via Minigera 22,
CH-6926 Montagnola,
Switzerland.

Summary

A common use of Gray codes is in reducing quantisation errors in various types of analogue-to-digital conversion systems [1, 2]. As a typical example, a length n Gray code can be used to record the absolute angular positions of a rotating wheel by encoding the codewords on n concentrically arranged tracks. n reading heads, mounted radially across the tracks, suffice to recover the codewords and it is well known that quantisation errors are minimised by using a Gray encoding.

When high resolution is required, the need for a large number of concentric tracks results in encoders with large physical dimensions. This poses a problem in the design of small-scale or high-speed devices. We propose single-track Gray codes as a way of overcoming this problem. Let $W_0, W_1, \ldots, W_{p-1}$ be the codewords of a Gray code $\mathcal{C}$ and write $W_i = [w_i^0, w_i^1, \ldots, w_i^{n-1}]^T$. We call the sequence $w_0^j, w_1^j, \ldots, w_{p-1}^j$ *component sequence* j of $\mathcal{C}$.

Definition 1 *If for each $1 \leq j < n$, component sequence j of $\mathcal{C}$ is a cyclic shift by some k_j of component sequence 0, i.e. $w_0^j, w_1^j, \ldots, w_{p-1}^j = w_{k_j}^0, w_{k_j+1}^0, \ldots, w_{k_j+p-1}^0$ (where subscripts are reduced modulo p), then we say that $\mathcal{C}$ is a single-track Gray code.*

In a single-track Gray code, codeword W_i is actually equal to $[w_i^0, w_{i+k_1}^0, w_{i+k_2}^0, \ldots, w_{i+k_{n-1}}^0]^T$ and so, in the application above, the bits of any codeword can be obtained solely from a single track corresponding to component sequence 0. The n reading heads are then spaced around that single track at fixed relative positions $0, k_1, k_2, \ldots, k_{n-1}$. So, if a suitable single-track Gray code is available, the respective encoder can be made significantly smaller in size.

Necessary conditions on the parameters n and p of a single-track Gray code are easily established:

Lemma 2 *Suppose there exists a length n single-track Gray code with p codewords. Then p is an even multiple of n and $2n \leq p \leq 2^n$.*

We are interested in two problems. Firstly, for a given n, obtaining a single-track Gray code with as many codewords as possible, and secondly, for a given number of codewords (i.e. resolution), obtaining a code with the smallest possible length n (i.e. number of reading heads). Codes are easily obtained for $n = 1, 2, 3$. However, for larger n, the construction of codes poses an interesting combinatorial problem. Though not ruled out by the necessary conditions, there is in fact *no* length 4 code containing all 16 words. Thus the conditions of Lemma 2 are *not* sufficient. We have obtained good codes by hand for small n. The number of words in these codes and the corresponding bound from Lemma 2 are shown in the table below.

n	Number of codewords	Upper bound from Lemma 2
4	8	16
5	30	30
6	60	60
7	126	126
8	240	256
9	360	504

As an example, our length 5 single-track Gray code with 30 codewords is:

00111100011000000011111111100
00011000000001111111100001111
00000001111111110000111100011 0
01111111110000111100011000000 0
11110000111100011000000011111

The code for $n = 9$ is particularly useful, as it gives a one-degree resolution using the least possible number of reading heads.

Our other contribution is a general construction yielding codes for a large variety of parameters and leading to the following:

Theorem 3 *Suppose $n \geq 4$. Then there exists a length n single-track Gray code with nt codewords for every even t satisfying $2 \leq t \leq 2^{n-\lceil \sqrt{2(n-3)} \rceil - 1}$.*

These codes are not in general optimal with respect to the conditions of Lemma 2. We propose as open problems finding better or even optimal single-track Gray codes for larger n, and obtaining a stronger upper bound on p than that given by Lemma 2.

References

[1] E. Gilbert, "Gray Codes and Paths on the n-Cube." *Bell System Technical Journal* **37** (1958), 815–826.

[2] V. Klee, "The Use of Circuit Codes in Analog-to-Digital Conversion," in *Graph Theory and its Applications*, B. Harris, Ed., Academic Press, New York, 1970.

[1] Supported by a Royal Society ESEP Research Fellowship and by a Lloyd's of London Tercentenary Foundation Research Fellowship.

Optimizing the Encoder/Decoder Structures in a Discrete Communication System

A. K. Khandani[1]

Elect. & Comp. Eng. Dept., Univ. of Waterloo, Waterloo, Ontario, Canada, N2L 3G1
E-mail: khandani@shannon.uwaterloo.ca

Abstract — **The problem of optimizing the structure of the encoder/decoder pair in a discrete communication system (with an additive distortion measure) is expressed in terms of a Bilinear Programming Problem (BLP Problem). An efficient method, based on the simplex search in conjunction with the Generalized Upper Bounding Technique is presented for the solution. The special features of the problem are exploited to reduce the computational complexity of the proposed algorithm.**

I. Introduction

Consider a discrete communication system composed of a source $\boldsymbol{S}$, a channel $\boldsymbol{C}$, an encoder ξ and a decoder η. The source $\boldsymbol{S}$ is composed of N_s symbols s_i, $i=0,\ldots,N_s-1$. The symbol $s_i \in \boldsymbol{S}$ occurs with probability $P_s(i)$. A measure of distortion is defined between each pair of the source symbols. The distortion between the symbols $s_i, s_j \in \boldsymbol{S}$ is denoted as $D_s(i,j)$, $i,j=0,\ldots,N_s-1$. The channel $\boldsymbol{C}$ is composed of N_c symbols c_i, $i=0,\ldots,N_c-1$. The symbol $c_i \in \boldsymbol{C}$ occurs with probability $P_c(i)$ and has an energy of $E_c(i)$. This results in an average energy of $\sum_{i=0}^{N_c-1} P_c(i)E_c(i)$ at the channel input. The transition probabilities of the channel are denoted as $T_c(j|i)$.

The encoder provides a mapping, denoted as ξ, from the set of source symbols to the set of channel symbols such that the ith source symbol, $i=0,\ldots,N_s-1$, is mapped to the channel symbol indexed by $\xi(i) \in [0, N_c-1]$. Each source symbol is encoded to a specific channel symbol, however, (i) several source symbols may be encoded to the same channel symbol, and (ii) some of the channel symbols may not be used. The decoder provides a mapping, denoted as η, from the set of channel symbols to the set of source symbols such that the ith channel symbol, $i=0,\ldots,N_c-1$, is mapped to the source symbol indexed by $\eta(i) \in [0, N_s-1]$. Each channel symbol is decoded to a specific source symbol, however, several channel symbols may be decoded to the same source symbol.

Our objective is to optimize the two mappings, namely ξ, η, to minimize the average distortion between the encoder input and the decoder output. The introduced formulation optimizes the combined effects of source quantization and channel coding on the end-to-end distortion. Quantization of the source symbols occurs when several source symbols are encoded to the same channel symbol. Channel coding occurs when some of the channel symbols are not used at all. In the following, this optimization problem is formulated as a zero-one program.

[1]This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC).

II. Zero-one formulation of the problem

We assign an N_c dimensional binary vector to each symbol of the source at the channel input. The vector corresponding to the ith source symbol, $i=0,\ldots,N_s-1$, is denoted as $\mathbf{e}_i=[e_{ij}, j=0,\ldots,N_c-1]$. We impose the constraints that $e_{ij} \in \{0,1\}$ and $\sum_{j=0}^{N_c-1} e_{ij}=1$, $\forall i$. If the ith source symbol is encoded to the ℓth channel symbol, we set, $e_{ij}=1$, $j=\ell$ and $e_{ij}=0$, $j \neq \ell$. Similarly, we assign an N_s dimensional binary vector to each channel symbol at the decoder side. The vector corresponding to the jth channel symbol, $j=0,\ldots,N_c-1$, is denoted as $\mathbf{d}_j=[d_{ij}, i=0,\ldots,N_s-1]$. We impose the constraints that $d_{ij} \in \{0,1\}$ and $\sum_{i=0}^{N_s-1} d_{ij}=1$, $\forall j$. If the jth channel symbol is decoded to the ℓth source symbol, we set $d_{ij}=1$, $i=\ell$ and $d_{ij}=0$, $i \neq \ell$. Using these notations, the optimization problem is formulated as:

$$\begin{aligned}
\text{Minimize} \quad & \sum_{i=0}^{N_s-1}\sum_{j=0}^{N_c-1}\sum_{k=0}^{N_s-1}\sum_{l=0}^{N_c-1} P_s(i)T_c(l|j)D_s(i,k)e_{ij}d_{kl} \\
\text{Subject to:} \quad & \sum_{i=0}^{N_s-1}\sum_{j=0}^{N_c-1} P_s(i)E_c(j)e_{ij} \leq \overline{E} \\
& e_{ij} \in \{0,1\} \quad \text{and} \quad \sum_{j=0}^{N_c-1} e_{ij}=1, \quad \forall i \\
& d_{ij} \in \{0,1\} \quad \text{and} \quad \sum_{i=0}^{N_s-1} d_{ij}=1, \quad \forall j
\end{aligned} \tag{1}$$

This optimization problem is transformed into a *Bilinear Programming Problem* (BLP Problem) [1]. The problem has some special features which substantially facilitates its solution. These features are: (i) Existence of the Generalized Upper Bounding (GUB) constraints for both encoder and decoder. (ii) The encoder structure has only one extra constraint in addition to the GUB's, namely the energy constrain. (iii) The decoder constraints are all GUB's and consequently the linear program involved in the optimization of the decoder is decomposable. Using these features, an efficient method based on a variant of the simplex search is presented for the solution.

References

[1] F. A. Al-Khayyal, "Jointly Constrained Bilinear Programs and Related Problems: An Overview," *Computers Math. Applic.*, vol. 19, no. 11, pp. 53—62, 1990.

The effect of space diversity on coded modulation for the fading channel

J. Ventura-Traveset, G. Caire, and E. Biglieri[1]

J. V.-T. is with European Space Agency / ESTEC, P.O. Box 299, 2200 AG Noordwijk (The Netherlands)

E. B. and G. C. are with Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy)

fax: +39 11 5644099 • e-mail: <name>@polito.it

Abstract — **We address the problem of designing a coded modulation scheme for the fading channel when space diversity is used. We focus on the fact that a channel affected by fading can be asymptotically turned into an additive white Gaussian noise (AWGN) channel by increasing the number of diversity branches, thus turning coded-modulation schemes designed for the AWGN channel into efficient codes over the fading channel.**

I. Introduction

The severe performance degradation effects associated with flat fading in radio channels are well known. Similarly well known is the fact that when coping with fading an alternative option to increased power is the use of multiple-receiver techniques categorized under the name of diversity. Recently, coded modulation has been regarded as a way of introducing time diversity. Actually, the effect of increasing the Hamming distance between pairs of possible symbol sequences transmitted over the flat Rayleigh-fading channel is the same as induced by increasing the number of branches in space diversity. One problem with this approach is that the design criteria for coded modulation schemes in fading channels differ from the standard minimum-Euclidean-distance criterion valid for the AWGN case. Consequently, a code optimal for the AWGN channel may perform poorly on a fading channel and vice versa.

We study the synergy of space diversity and code diversity. In partucular, we focus our analysis on the fact that antenna diversity and maximal-ratio combining have the effect of turning the equivalent transmission channel into an AWGN channel. A structure of a receiver with constant total gain is advocated. With it, when the space-diversity order is M the energy per diversity branch is decreased by a factor of M, so that the average signal-to-noise ratio at the decoder input remains the same irrespective of the diversity order. In practical terms, we might think of an antenna array whose number of elements is increased without increasing the total area, so that the equivalent gain of the antenna is kept constant. With this receiver, at no additional cost in terms of antenna size, an optimal code for the AWGN channel can achieve (asymptotically as the number of diversity branches increases) the same optimal performance on a fading channel, irrespective of the fading parameters. This asymptotic performance is approached with only a few, highly correlated diversity branches.

The following results were obtained:

- Bounds on the bit error probability of a coded modulated system with diversity, including branch correlation.
- The cut-off rate of the diversity channel.
- Simulation results based on simple coded modulation schemes for 8-PSK with several detection strategies.
- Rate of convergence of the fading channel to an AWGN channel as the number of diversity branches increases.

In the following we describe the latter results.

II. Convergence to AWGN channel

With antenna diversity, coherent detection and perfect channel-state information the convergence to AWGN channel is very quick.

We observe that the divergence of the channel with diversity from a channel without fading is due to the combination of two factors, namely, divergence from Gaussianity and a larger value of the noise variance. While the convergence to a channel in which fading is simply wiped out is important, the sheer convergence of the total disturbance to a normal distribution (even with a slightly larger variance) implies that coding schemes that have been optimized for a Gaussian channel will perform closer and closer to optimality. Convergence can be studied by examining the Kullback-Leibler distance of the probability density function $f(x)$ of the total disturbance (fading plus noise) from a normal distribution with a variance equal to that of the noise, which we denote here by $g(x)$, and from a normal distribution with larger variance, denoted $g'(x)$. The results obtained are reported in Table 1.

M	$D(f \parallel g)$	$D(f \parallel g')$
2	0.474	0.167
3	0.154	0.060
4	0.076	0.031
5	0.045	0.018
6	0.030	0.012
7	0.021	0.009
8	0.016	0.007
9	0.012	0.005
10	0.010	0.004

Table 1: Kullback-Leibler distance of distributions $f(x)$ and $g(x)$ and of distributions $f(x)$ and $g'(x)$.

[1]This research was sponsored in part by the Human-Capital and Mobility Program of the European Union.

MULTILEVEL CONCATENATED CODED MODULATION SCHEMES FOR THE SHADOWED MOBILE SATELLITE COMMUNICATION CHANNEL

Do Jun Rhee and Shu Lin[1]
Department of Electrical Engineering,
University of Hawaii at Manoa
Honolulu, Hawaii 96822, U.S.A.

Abstract — **This paper presents a bandwidth efficient multilevel concatenated coded modulation scheme for reliable data transmission over the shadowed mobile satellite communication (MSAT) channel. In this scheme, bandwidth efficient block modulation codes are used as the inner codes and Reed-Solomon codes of various error correcting capabilities are used as the outer codes. The inner and outer codes are concatenated in multiple levels. A systematic method for constructing multilevel concatenated modulation codes is presented and a multistage closest coset decoding for these codes is proposed. Specific multilevel concatenated 8-PSK modulation codes have been constructed. These codes are designed to remove the error floor phenomenon or lower the bit-error rate of the error floor caused by the large Doppler frequency shift due to the motion of vehicles. Simulation results show that these codes perform very well and achieve large coding gains over the uncoded reference modulation systems.**

I. Introduction

In this paper, we propose and investigate multilevel concatenated coded modulation schemes for shadowed MSAT channel. A statistical model for the shadowed mobile satellite channel has been devised by Loo [1]. This model has been used by other researchers [2, 3] to study error performances of coded modulation schemes over the shadowed mobile satellite communication channel. In the Loo's model, there are three different kinds of shadowing, i.e., light, average and heavy. The corresponding Rician factors are 6.16, 5.46 and -19.33 dB, respectively. Therefore, in the heavy shadowed MSAT channel which is statistically close to the Rayleigh fading channel, a coded modulation system suffers very severe distortion due to randomly changing phase and the multipath fading. Especially, if the Doppler frequency shift is large due to the motion of vehicle, a coded modulation system faces the error floor phenomenon.

II. Multilevel Concatenated BCM Schemes

Coded modulation in conjunction with concatenation is a powerful technique for achieving **high reliability**, large coding **gain**, and **high spectral efficiency** with **reduced decoding complexity**. This combination of coded modulation and concatenation is known as **concatenated coded modulation** [4]. Error performance of the single-level concatenated TCM and BCM schemes for the Rayleigh fading channel was investigated by Vucetic and Lin in 1991 [5]. All these studies showed that by properly choosing the inner codes and outer codes, large coding gains and high spectral efficiency could be achieved with reduced decoding complexity.

However, a major shortcoming of a single-level concatenated coded system with multilevel block modulation code as the inner code is that the outer code corrects all the output bits of the inner code decoder to the same degree. Since a multilevel modulation code is constructed from component codes with different distance profiles, multistage decoding results in different bit-error probabilities for different component codes at the output of the inner code decoder. As a result, the overall error performance of a single-level concatenated coded modulation system is dominated by the worst bit-error probability of the component code of the modulation inner code. To improve the overall error performance, it is necessary to provide different levels of error protection for different inner component codes in a concatenated coded modulation system. One approach to this improvement is to use multilevel concatenation with multiple outer codes to provide different levels of error protection for different inner component codes. Multilevel concatenation provides the flexibility of choosing outer codes with different error correcting capabilities and furthermore improves the spectral efficiency over the single-level concatenation scheme.

Simulation results show that these codes achieve very impressive real coding gains over the uncoded reference system and single-level concatenated BCM codes using the same inner codes.

[1]This research was supported by NSF Grants NCR-91-1540 and NCR 94-15374 and NASA Grant NAG 5-931.

References

[1] C. Loo, "A statistical model for a land mobile satellite link,"in it Links, for the future (ICC'84), Science, Systems, and Services for Communication, P. Dewilde and C. A. May, Ed. New York:IEEE/Elsevier, North-Holl and, 1984.

[2] P. J. McLane, P. H. Wittke, P. H0, and C. Loo, "PSK and DPSK trellis codes for fast fading, shadowed mobile satellite communication channels," in *Pro. 1987 Int. Conf. Commun.*, Seattle, WA, June 7-10, 1987, pp. 21.1.1-21.1.6.

[3] S. H. Jamali and T. Le-Ngoc, "Performance Comparison of Different Decoding Strategies for a Bandwidth-Efficient Block-Coded Scheme on Mobile Radio Channels," *IEEE Trans. on Veh. Technol.*, Vol. 41, No. 4, pp. 505-51 5, Nov. 1993.

[4] T. Kasami, T. Takata, T. Fujiwara, and S. Lin, "A Concatenated Coded Modulation Scheme for Error Control," *IEEE Trans. on Communications*, Vol. COM-38, No. 6, pp. 752-763, June 1990.

[5] B. Vucetic and S. Lin, "Block Coded Modulation and Concatenated Schemes for Error Control On Fading Channels," **Journal of Discrete Applied Mathematics**, No. 33, pp. 257-269, 1991.

A Hidden Markov Model (HMM) –Based MAP Receiver For Nakagami Fading Channels

Hongwei Kong and Ed Shwedyk
Dept. Elec. Comp. Eng., University of Manitoba, Winnipeg, MB R3T 5V6, Canada
email: hongwei@ee.umanitoba.ca / shwedyk@ee.umanitoba.ca

I. INTRODUCTION

Signalling over Rayleigh fading channels can be classed as a general Gaussian problem. Optimal linear filtering can then be applied to jointly estimate the channel and detect the information sequence [1]. For fading channels with non–Gaussian distributions, optimal linear filtering does not necessarily yield the best channel estimates. To exploit the channel memory, a first order finite Markov chain model (HMM) that statistically characterizes the Nakagami–m fading process is used to aid the channel estimation. Based on this, a maximum a posteriori (MAP) receiver using coherent detection is presented for binary PAM signals.

II.SYSTEM MODEL AND THE BRANCH METRIC

The system model is shown in Fig. 1 where $g(t)$ is the multiplicative Nakagami fading process. A first order finite state Markov chain model for $g(t)$ can be derived using the procedure described in [2].

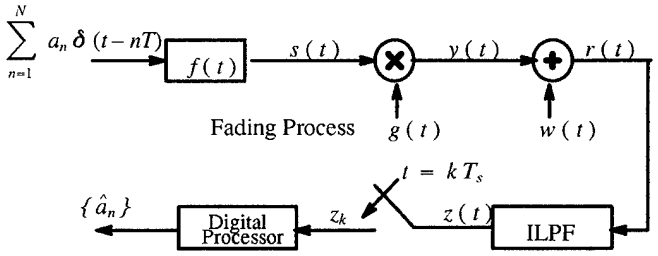


Fig. 1 System Model

The shaping filter, $f(t)$, is selected so that when the received signal is sampled with interval T_s, no intersymbol interference will occur. A trellis can be set up for this receiver where its states are the states of the Markov chain model. The branch metric for the trellis search is:

$$[\, z_k - g_k b_k f(0)\,]^2 - 2\sigma_k^2 \ln \Pr(g_k \mid g_{k-1})$$

where $\{b_k\}$ is the equivalent information sequence and σ_k^2 is the variance of the noise that accounts for both the additive white Gaussian noise and the modelling error of the fading process. The last term accounts for the state transition probability of the Markov chain.

III. SIMULATION RESULTS

Simulation has been done for binary PAM with coherent detection. The Nakagami fading process is generated from a correlated Gaussian process which in turn is generated by passing a white Gaussian process through a second order low pass Butterworth filter whose cutoff frequency determines the rate of fading. An 8–state Markov chain model is used to represent the Nakagami fading process.

Fig. 2 shows the error performance for the MAP receiver for m=0.5 and 5.0. The bandwidth of the Butterworth filter is chosen to be 0.1Hz, which corresponds to fast fading. For comparison, the error performance for receivers where the LMS algorithm is used to estimate the channel is also given. Fig. 3 shows the error performance for two different fading rates with m=2.0. One can observe from the figures that the

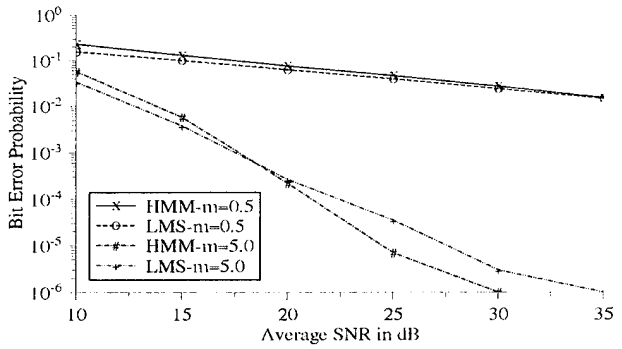


Fig. 2 Error Performance for Fast Fading

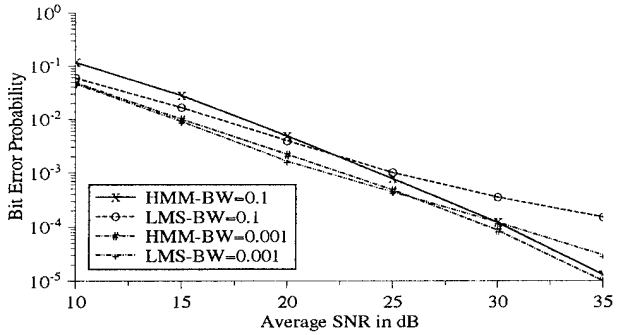


Fig . 3 Error Performance for Different Fading Rates, m=2.0

MAP receivers using the Markov chain model give somewhat better error performance for medium to high average SNR than receivers which employ optimal linear filtering to estimate the channel. Also the error performance of the MAP receiver is not sensitive to the fading rates for a fixed value of m. Simulations have also been done for m=1.0, 10.0 and 20.0, under different fading rates. The improvement of the error performance using the Markov model becomes more significant as m and/or the fading rates increase. The method is readily extended to frequency selective fading channels with non–Gaussian distributions whereas MLSE receiver proposed in [1] is difficult or impossible to implement.

REFERENCES

[1] Q.Dai and E.Shwedyk, "Detection of Bandlimited Signals over Frequency Selective Rayleigh Fading Channels", *IEEE Trans Commun.*, Vol 42, NO. 2/3/4, pp 941–950, Feb/Mar/Apr., 1994.

[2] H.Kong and E.Shwedyk, "Sequence Estimation for Frequency Nonselective Channels Using A Hidden Markov Model", pp 1251–1253, *44th Vehicular Tech. Conf*, Stockholm, June, 1994.

Research Supported By an NSERC Operating Grant

Polynomial representation of burst error statistics

Ludwig Kittel
FernUniversität in Hagen, D-58084 Hagen, Germany

Abstract - **In this paper, a burst error process is characterized by three variables x, y, z related to a two-state Gilbert-Elliott-Channel with fifty percent bit error rate in the bad state. The variables x and y describe the Markov process of the model. The variable z reflects the mean bit error rate. On such a channel, the probability of any single error sequence and hence any collection thereof can be represented by polynomials in x, y, z which extends the one-variable description in z applicable for statistically independent errors.**

I. Burst Error Statistics

On a binary channel, an n-bit error sequence $w = e_1 e_2 \cdots e_n, e_\nu \in \{0,1\}$, can be considered as elementary error event. There are $N = 2^n$ distinct elementary events $w_i, i = 0, 1, \cdots, N-1$; each occurring with probability $P_i = P(w_i) = Prob(w_i)$. A composite error event E is given by the union of constituting elementary error events w_i characterized by an appropriate index set I_E, i.e.

$$E = \bigcup_{i \in I_E} w_i \ , \quad P_E = Prob(E) = \sum_{i \in I_E} Prob(w_i). \qquad (1)$$

Pertaining burst error statistics P_E are, among others, the error weight distribution $P(m,n)$, and the error correlation function $R(\tau)$. $P(m,n)$ is the probability of m errors in a block of n bits; $R(\tau) = Prob(1e^{\tau-1}1)$ is the probability of two errors occurring at a distance τ.

II. Burst Error Model

The burst error process is modelled by a two-state Gilbert-Elliott-Channel (GEC), characterized by the bit error rates p_G and p_B associated to the good state G and the bad state B, and the state transition probabilities $P = Prob(B|G)$ and $Q = Prob(G|B)$, resp. The mean bit error rate is $\overline{p}_b = p_G \frac{Q}{P+Q} + p_B \frac{P}{P+Q}$. A reduced GEC with $p_B = 0.5$ will be applied. The three remaining parameters p_G, P, Q are expressed by $x = P/Q, y = 1-(P+Q)$, and $z = 1-2\overline{p}_b$. See [2] for physical interpretation. Describing matrices are

$$\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_1 = \frac{1}{1+x}\begin{bmatrix} 1+xy & x-xy \\ 1-y & x+y \end{bmatrix}, \qquad (2)$$

$$\delta = \mathbf{D}_0 - \mathbf{D}_1 = z\begin{bmatrix} 1+xy & x-xy \\ 0 & 0 \end{bmatrix}. \qquad (3)$$

The stationary state distribution is $\boldsymbol{\sigma}_0 = \frac{1}{1+x}[1, x]$.

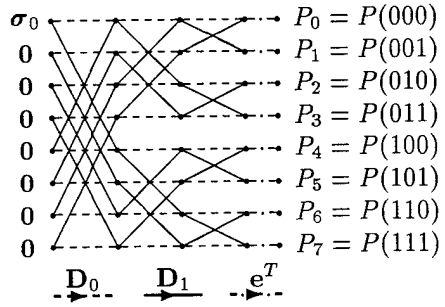


Fig. 1: Trellis for evaluating $P(w_i)$

III. Polynomial Representations

The product formalism of stochastic automata theory

$$P(w) = \boldsymbol{\sigma}_0 \mathbf{D}_{e_1} \mathbf{D}_{e_2} \cdots \mathbf{D}_{e_n} \mathbf{e}^T, \mathbf{e}^T = (1, 1, \cdots, 1)^T \quad (4)$$

can be used to show by complete induction that P(w) is indeed a polynomial in x, y, z. Generalizing (4) yields

$$\begin{aligned} R(\tau) &= \boldsymbol{\sigma}_0 \mathbf{D}_1 \mathbf{D}^{\tau-1} \mathbf{D}_1 \mathbf{e}^T, \qquad (5) \\ &= \frac{1}{4}[1 + 2z + (1+xy^\tau)z^2]. \end{aligned}$$

Applying modal analysis [1,2], the probability vector $\mathbf{P} = [P(w_i)]$ can be expressed via Walsh-Hadamard-Transformation of the spectral coefficient vector $\mathbf{Q} = [Q(w_i)]$, where $Q_i = Q(w_i)$ are simple polynomials in x, y, z.

$$\mathbf{P} = 2^{-n}\mathbf{Q}\mathbf{V}_n, \mathbf{V}_n = \begin{bmatrix} \mathbf{V}_{n-1} & \mathbf{V}_{n-1} \\ \mathbf{V}_{n-1} & -\mathbf{V}_{n-1} \end{bmatrix}, \mathbf{V}_0 = [1] \quad (6)$$

For $n = 3$, evaluation trellises are shown in Fig. 1, 2. As the Hadamard matrix $\mathbf{V}_n$ consists of entries $+1$ and -1, resp., $P(w_i)$ and hence P_E consist of appropriate aggregats of $Q(w_i)$, e.g.

$$\begin{aligned} P(2,3) &= \frac{1}{8}[3 - 3z - 2(1+xy)z^2 \qquad (7) \\ &\quad -(1+xy^2)z^2 + 3(1+xy)^2 z^3]. \end{aligned}$$

References

[1] Kittel, L.: "Modal analysis of block coded transmission systems with application to error control performance evaluation", ITG-Fachbericht 107, VDE-Verlag GmbH, Berlin 1989, pp.281-288

[2] Kittel, L. and Zepernick, H.-J.: "Generalized weight polynomials for linear binary block codes used on a burst error channel", ISITA'90, Honolulu, Nov. 90, pp.175-178

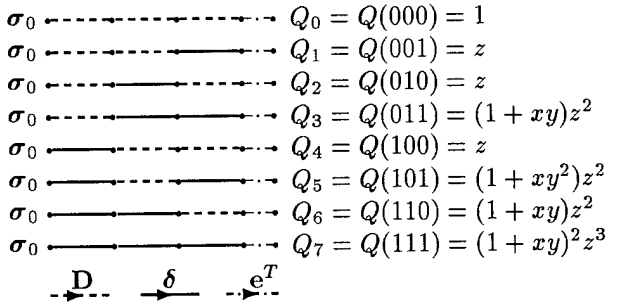


Fig. 2: Diagonalized trellis for evaluating $Q(w_i)$

Coherent Detection for Transmission over Severely Time and Frequency Dispersive Multipath Fading Channels

Elie Bejjani *[1], *Member, IEEE*, Jean-Claude Belfiore **, *Member, IEEE* and Philippe Leclair *, *Member, IEEE*
(*) SETICS, 194 rue de Tolbiac, 75013 Paris- France
(**) ENST, Dép. Comm., 46 rue Barrault, 75634 Paris cedex 13- France

Abstract — **We propose a new technique for coherent transmission over multipath Rayleigh fading channels, based on the use of one special case of time-frequency well-localized orthonormal functions, namely the Prolate Spheroidal Wave Functions (PSWF). Acceptable SER performances are obtained until values of about 0.1 of the channel's spread factor.**

I. Introduction

Coherent signaling over very dispersive Rayleigh fading channels is quite a challenging task. A classical rule of thumb to respect in such cases is to choose a signaling symbol time interval T verifying $T_m \ll T \ll 1/B_d$, where T_m denotes the time spread due to multipath propagation, and B_d denotes the Doppler spread bandwidth due to individual path's envelope fading. This is indeed possible when the channel's spread factor $L = T_m B_d$ is very small ($L \leq 0.01$). Excellent results have been achieved in such situations by the use of pilot symbols in association with coded modulation [1]-[2]. When L approaches unity, any attempt to make $T \ll 1/B_d$ will result in severe multipath spreading.
In our work we consider coherent signaling over channels with spread factors $T_m B_d \leq 0.1$. Our technique permit coherent detection in situations traditionally reserved to non-coherent reception, with the evident benefit of higher spectral efficiency.

II. Time-Frequency Orthonormal Bases

Time-fequency localization operators are of interest for many applications. A well known example of such operators is the one presented by Slepian and Pollak [3]. In an extensive serie of articles ([3] and other) they studied the properties of signal band- and time-limiting operators on the $[-T/2, T/2] \times [-W, W]$ rectangle. They demonstrated that the orthonormal family of Prolate Spheroidal Wave Functions (PSWFs) is a complete basis of singular functions for the above-mentioned operator.
Another example of such operators is the case of the Hermite polynomial functions which are the eigenfunctions of the projection operator on disks of the time-frequency plane $(t^2 + \omega^2 \leq R^2)$ [4].

III. Simultaneous Data and Pilot Symbols

In this section we investigate the performances of a transmission scheme well fitted to severely dispersive Rayleigh fading channels (L ranging from about 0.01 to 0.1).
When the spread factor of the channel exceeds 0.01, solutions based on the transmission of frames of N symbols with the first symbol being the pilot and the $N-1$ remaining symbols carrying the data [2] do not work. In our approach we exploit the orthogonality of the PSWFs in order to transmit in the same time interval T both pilot and information symbols. This is accomplished by simultaneous transmission of different orthogonal PSWFs, one among these carrying the pilot symbol. Fig.1 shows computer simulated performances of the proposed technique for several values of the spread factor L. It confirms that for $L < 0.01$, the performance of our technique is comparable with that obtained in [2]. Moreover, we notice that the effect of the extending multipath spread is the presence of a floor of the SER.
When the channel's spread factor exceeds 0.01, the advantage of our method is the graceful degradation of the performances in contrast with the impossibility of implementing the method described in [2].
Results for the QPSK modulation (not shown) are only a few percent worse than those of the BPSK.

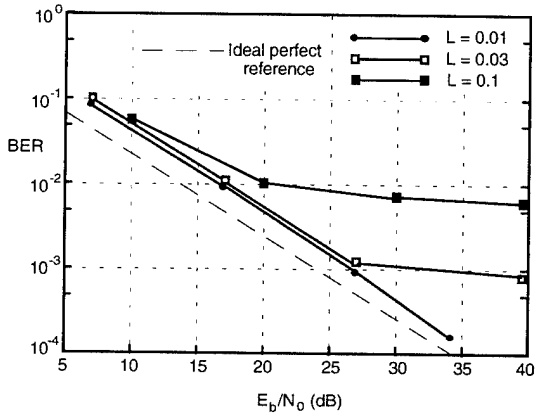


Fig. 1: SER performance for BPSK.

IV. Conclusion

We showed the feasibility of coherent detection for multipath fading channels with a spread factor reaching 0.1. It is possible to achieve coherent transmission with even higher spread factors, but on condition that additional processing is used. We are presently working in this direction.

[1]The work of E. Bejjani is supported by SETICS (Société d'Etudes en TéléInformatique et Communications Systèmes), within a PhD thesis contract in association with ENST (Ecole Nationale Supérieure des Télécommunications).

References

[1] K. Boullé and J. C. Belfiore, "Modulation schemes designed for the rayleigh channel", *CISS'92*, Princeton (USA), March 1992.

[2] J. K. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels", *IEEE Trans. Commun.*, vol. 40, No. 4, pp. 686–693, November 1991.

[3] D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier analysis and uncertainty- I", *Bell Syst. Tech. J.*, vol. 40, pp. 43–64, January 1961.

[4] I. Daubechies, "Time-Frequency localization operators: A geometric phase space approach", *IEEE Trans. Inf. Theory*, vol. 34, No. 4, pp. 605–612, July 1988.

Optimised multistage coded modulation design for Rayleigh fading channels

A. G. Burr

Dept. of Electronics, University of York, York, U.K.

I. Introduction

It has been known for some time [1,2] that coded modulation schemes such as Ungerboeck's [3] which are optimised for the Gaussian channel do not perform well on fading channels, and especially on the Rayleigh fading channel. Ungerboeck's codes maximise minimum Euclidean distance between coded sequences; Divsalar and Simon identified a number of parameters that should be maximised in preference, notably the minimum Hamming distance and the product distance. More recently it has been suggested [4] that the framework of multilevel coded modulation (MCM) forms a suitable basis for the design of such codes, since it enables the Hamming distance readily to be maximised. This has the further advantage [5] that decoders may be implemented using readily-available ASICs for binary convolutional codes.

In most of this work the aim has been to optimise asymptotic performance at high signal to noise ratio (SNR), and the choice of code parameters has been made accordingly. However, it is well-known (e.g. [6]) that this may not optimise performance at practical SNR. This paper presents a design technique to minimise required SNR performance for a specific target bit error ratio (BER). It also describes a new simplified bounding technique for the BER of MCM on a Rayleigh fading channel, which avoids the use of Chernoff bounds.

II. Bounds on BER of MCM on a Rayleigh channel

The principle of multistage decoding of MCM is to decode each partition of the signalling constellation separately, treating the remaining partitions as uncoded. This is of course sub-optimum. It allows us, however, to treat each stage of the decoding process as binary. We may then use analytical expressions for the BER of binary signalling on a Rayleigh fading channel [7]. We treat a binary code with minimum free distance d as binary signalling with d branch diversity, and use a union bounding technique similar to that described in [8]. From [7] p. 474, the BER of binary signalling with d-branch diversity is:

$$P(d,E_c/N_0)=\left(\frac{1-\mu}{2}\right)^d \sum_{k=0}^{d-1} {}^{d-1+k}C_k \left(\frac{1+\mu}{2}\right)^k \quad ..(1)$$

where $\mu=\sqrt{(E_c/N_0)/(1+E_c/N_0)}$

Following [6], we define the stage BER P_i as the error probability in all stages due to errors at the i^{th} stage, thus including error propagation, allowed for in the factor ε_i. Note that, unlike [4], we do not assume interleaving between stages. Then following [8] we calculate an estimate of P_i taking into account erroneous paths up to Hamming distance d_{max}. Suitable values for d_{max} are found by comparison with simulations.

$$P_i \approx \varepsilon_i \sum_{d=d_{free}}^{d_{max}} a_i^d A_{d,i} P\left(d, \frac{R_i E_b}{N_0}\frac{\Delta_i^2}{4}\right) \quad ..(2)$$

where $\varepsilon_i = 1+R_{i+1}/2R_i, i=1,2; \varepsilon_3=1$

In an MCM decoder the error-weighted distance spectrum A_d of the code must be multipled by the factor $a_i{}^d$, where $\{a_i, i=1..3\}=\{2,2,1\}$ is the number of neighbouring points in the signalling constellation partition at each level.. $\{\Delta_i, i=1..3\}=\{2\sin(\pi/8), \sqrt{2}, 2\}$ is the partition minimum distance at each stage.

III. Design and performance of optimum codes

Using this technique we calculate stage BERs for a given scheme. The overall BER is then the sum of these. It has been noted [4] that many MCM codes have overall BER dominated by one stage of decoding. Here we apply the principle of equalising stage BERs at a given SNR, hence optimising the codes for this SNR.

This method has been applied to optimise 8-PSK codes for BER 10^{-3} and 10^{-6}. Codes with encoder memory 6 and rates $\{R_i, i=1..3\}=\{3/8, 3/4, 7/8\}$ and $\{1/2, 2/3, 5/6\}$ respectively were selected by means of an iterative procedure which equalised stage BERs at the required SNR. For code rates other than $1/k$, punctured codes are used.

Fig. 2. compares the overall BERs of these codes with a code with equal Hamming distances on each level [4]. It can be seen that there is a significant performance improvement over the equal Hamming distance code, of about 2.3 dB for BER 10^{-6}. This code also improves on the Schlegel and Costello code with the same memory [2] by over 4 dB at this BER. However, asymptotically the codes described here have much poorer performance.

IV. Conclusions

We have presented codes for the Rayleigh fading channel optimised for given finite error rate that achieve significant improvements in coding gain on previously-described codes. These codes are based on multilevel coded modulation, and hence can be decoded using multistage decoding, using readily-available Viterbi decoders. The use of punctured codes also makes for a very flexible structure, in which codes of different overall rates, and optimised for different BERs, may readily be implemented.

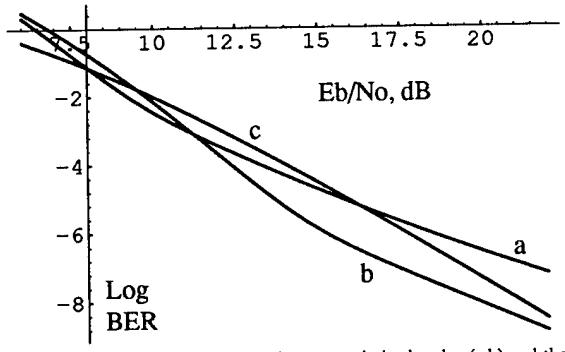


Fig. 2. Comparison of overall BER for the two optimised codes (a,b) and the equal Hamming distance code (c)

References

1. Divsalar, D. and Simon, M. K. "The design of trellis coded MPSK for fading channels: performance criteria" *IEEE Trans. Commun*, vol. 36, pp. 1004-1012, September 1988
2. Schlegel, C. and Costello, D. J. "Bandwidth efficient coding for fading channels: code construction and performance analysis" *IEEE J. Select. Areas Commun*, vol. 7, pp 1356-1368, December 1989
3. Ungerboeck, G. "Channel coding with multilevel/phase signals" *IEEE Trans. Inform. Theory*, vol. 28, pp 55-67, January 1982
4. Seshadri, N. and Sundberg, C-E. W. "Multilevel coded modulations for fading channels" *Proc. 5th Int. Tirrenia Workshop* (Elsevier, 1992), pp. 305- 316
5. Viterbi, A. J. et al "A pragmatic approach to trellis-coded modulation" *IEEE Communications Magazine*, July 1989
6. Burr A. G. and Lunn T. J: "Code optimisation for finite error rate" *Proc. IEEE Symposium on Information Theory*, San Antonio, Texas,January 1993, p 67
7. Proakis, J. G. "Digital Communications" McGraw-Hill, 1983
8. Burr, A. G. "Bounds and approximations for the bit error probability of convolutional codes" Electronics Letters, vol. 29, July 1993, pp 1287-88

Distributed Reception of Fading Signals in Noise

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

Abstract — **A multiple antenna diversity scheme is investigated for digital wireless communications. In this scheme the antenna observations are immediately quantized and only the quantized values are sent to a fusion center to decide which symbol was transmitted. The case where fine quantization is impractical is considered, so that distributed detection principles apply. The optimum reception scheme is described for the case where frequency shift keying is employed. Multiple bit quantization schemes are considered for cases where the observations at each antenna are influenced by slow Rayleigh fading and Gaussian additive noise. Some numerical results are provided.**

I. Introduction

There is significant interest in using wireless communication systems in environments where severe multipath fading and co-channel interference is present, which can limit system performance [1]. To mitigate the effects of multipath fading and co-channel interference, diversity techniques using multiple receive and transmit antennas have been proposed [1] and it has been found that the performance improvements obtained by using these schemes can be significant.

There appears to be a trend towards increasing the portion of wireless receivers that are implemented using digital technology in many applications. Recent improvements in electronic technology indicate that all digital receivers are becoming practical at many frequencies of interest and further improvements in the speed of analog-to-digital converters are expected to continue this trend. These facts indicate that multiple antenna diversity schemes that combine quantized samples should be considered, as illustrated in Figure 1.

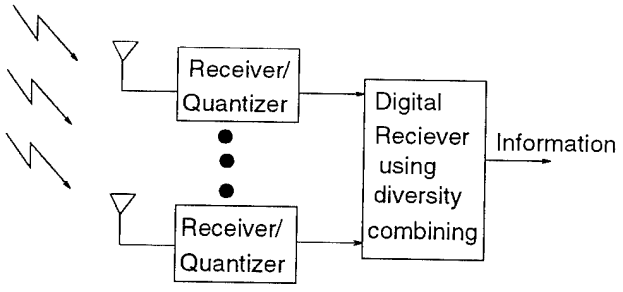


Figure 1: Distributed diversity combining.

In Figure 1, each individual receiver makes a multi-bit decision about which symbol was sent based only on the observations available at the co-located antenna. These decisions are then transmitted to a single location where a final decision is made. This is equivalent to a distributed signal detection problem. Two studies of diversity schemes based on combining single bit decisions made at several antennas have been reported [2, 3]. More recently, we investigated the optimum design of multi-bit decision schemes. If the quantizations produce samples with enough bits of resolution then the entire scheme will closely resemble the diversity schemes considered for analog receiver implementations [1], which includes the majority of research in this area. However, based on our recent results, it appears that it is not always necessary to use such high resolution quantizations. Using course quantizations, with only a few bits resolution, could reduce cost and complexity considerably. Our recent work indicates that course quantizations can sometimes be used without noticeable loss in performance provided one uses the proper quantizer designs.

We considered cases with independent fading (and noise) from antenna to antenna, a case of considerable interest [1]. Further, we considered a communications system where noncoherent binary frequency shift keying (FSK) is to be employed. Assume that N receivers, each with an associated antenna, are to be employed to achieve a diversity gain. A nonselective fading channel is considered where the fading is assumed to be slow enough so that it can be assumed constant over several bit periods. In our explicit examples, Rayleigh fading is assumed. The observations at each receiver are assumed to include additive zero-mean Gaussian noise.

Each of the receivers will generate a multiple bit decision and a single final decision will be made by fusing the decisions from the individual receivers. Assume that synchronization between the individual receiver decisions has been achieved, so that each set of receiver decisions correspond to the same transmitted digit. We assume that an accurate estimate of the signal-to-noise ratio is obtained for the observations available at each receiver. Here we consider the case where these estimates can be assumed to be correct, as a first approximation.

We have outlined the optimum design of such a system and we compared the performance of this system to a system which uses infinite precision. Our results indicate that using only two or three bits in the individual decisions does not sacrifice much performance, while this can simplify receiver design and construction. This appears to be an important result which could be used to reduce the implementation cost of wireless receivers. Due to the expansion in this industry, we believe these results could have significant impact.

[0]This material is based upon work supported by the National Science Foundation under Grant No. MIP-9211298

References

[1] N. Seshadri, C. W. Sundberg and V. Weerackody, "Advanced Techniques for Modulation, Error Correction, Channel Equalization, and Diversity," *AT & T Technical Journal*, vol. 72, No. 4, pp. 48-63, Jul. 1993.

[2] A. D. Kot and C. Leung, "Optimal Partial Decision Combining in Diversity Systems," *IEEE Transactions on Communications*, pp. 981-991, July 1990.

[3] R. Sannegowda and V. Aalo, "Performance of Partial Decision Combining Diversity Schemes in Nakagami Fading," *27th Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, NJ, March 1994.

Performance of Trellis Coded Direct-Sequence Spread-Spectrum with Noncoherent Reception in a Fading Environment

Victor W, Cheng, Wayne E. Stark[1]
The University of Michigan
Ann Arbor, Michigan 48105
vicwk@umich.edu, stark@eecs.umich.edu

Abstract — **In this paper we consider a different coding scheme for direct-sequence spread-spectrum(DS-SS). The Nordstrom-Robinson(NR) code, a nonlinear code that has large distance for a given rate, used in conjunction with a trellis-code [2] version is examined. A bound is developed on the error probability for this trellis coded Nordstrom-Robinson (TCNR) code with noncoherent reception over a frequency-nonselective Rayleigh or Rician fading channel with additive white Gaussian noise. This bound is tighter than a standard union bound. Our results indicate that the standard union bound can be significantly different from the more accurate results obtained from the improved union bound.**

I. Introduction and System Model

In a conventional DS-SS communication system a single data bit is transmitted using a pseudo-random sequence or its negative and binary phase shift keying. The number of information bits per channel chip is a measure of the rate of the system when it is used in an environment with multiple-access interference or multipath fading, which limits the maximum data rate capability. An error-correcting code such as convolutional code or block code can be used to provide additional protection, usually at the expense of data rate. It is also important to consider the number of nearest neighbors codewords, which affect error probability. A method to reduce the number of nearest neighbors without sacrificing data rate is to use a combination of an orthogonal code with a trellis at the expense of complexity. In this paper we wish to explore a coding scheme to achieve higher data rate and lower error probability. This coding scheme was first introduced in [1] and analyzed for coherent reception with multiple-access interference. A nonlinear Nordstrom-Robinson (NR) code can also be modified and used with noncoherent detection. This code has good distance and rate performance, and can be efficiently decoded with a soft decision algorithm.

If an orthogonal code has 16 codewords of length 16 with minimum distance 8, the data rate is 4/16 (4 information bits over 16 channel chips). Starting with this code, we can, by adding selected orthogonal cosets to the original code, increase the number of codewords up to 128 with the minimum distance slightly decreasing to 6. By doing so we get the nonlinear Nordstrom-Robinson code, which is composed of 8 cosets, each of 16 orthogonal codewords. The NR code has the geometric uniformity property; i.e., its distance distribution and its weight distribution are identical. This property greatly simplifies the analysis and simulation because the conditional error probability does not depend on which particular codeword is transmitted. When combined with a 4-state trellis, this trellis coded NR(TCNR) code, an example of finite-state codes, can transmit 6 information bits in every 16 channel chips. Thus we have decreased the minimum distance by 25% while having increased the rate by 50% to 6/16.

In this paper we examine the performance of this 4-state TCNR code over a frequency-nonselective Rayleigh or Rician fading channel with additive white Gaussian noise. Noncoherent reception is assumed, and the codewords are assumed to be interleaved at every 16 chips. An upper and a lower bound on the error probability have been derived. Also, the error performance of the TCNR code and a conventional DS-SS code with the same data rate, 6/16, is compared.

II. Numerical Results and Conclusions

The upper bound we derive for the TCNR code is tighter than the standard union bound. This is because in the TCNR code the minimum distance error events are from codewords within the orthogonal coset, whose error probability can be calculated exactly when the channel is assumed nonselective Rayleigh or Rician fading and thus the orthogonality within each coset is preserved. By taking this minimum distance error and then upper bounding all the remaining error events, we get an improved union bound. Numerical results imply that, at high signal-to-noise ratio (SNR), this upper bound tends to merge with the error probability from minimum distance codewords only, which is our lower bound. It is also shown that at high SNR, the TCNR code has better error performance than the conventional DS-SS code with the same data rate. For example, our results indicate that, compared with the conventional DS-SS code with the same data rate, there is approximately a 4-dB gain in E_b/N_o at high SNR for TCNR code over a Rayleigh fading channel.

References

[1] W. E. Stark and J. S. Lehnert, "Coding alternatives for direct-sequence spread-spectrum multiple access communications", *Thirty-Second Annual Allerton Conference of Communication, Control, and Computing*, Oct, 1994.

[2] F.Pollara, R.J. McEliece, and K. Abdel-Ghaffer, "Finite state codes", *IEEE Transactions on Information Theory*, vol. 34, pp. 1083-89, Sep, 1988.

[1]This paper was partially supported by the National Science Foundation under Grant NCR-9115969

Reliable Communication over the Rayleigh Fading Channel with I-Q TCM

Saud A. Al-Semari and Thomas E. Fuja[1]

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

Abstract — **I-Q TCM is a form of coded modulation in which two independent encoders select the in-phase and quadrature components of the transmitted signal. This design approach results in a significant increase in minimum time diversity when compared with comparable "traditional" TCM schemes. I-Q TCM schemes of varying complexity are presented; it is shown that the coding gains of moderately complex systems are very close to what is expected from the cutoff rate limit.**

I. Introduction

The design of trellis-coded modulation (TCM) schemes for mitigating the effects of Rayleigh-distributed flat fading has received considerable attention. It has been pointed out that the effective time diversity of the code (i.e., its symbol-wise minimum Hamming distance) is the main design criterion to optimize trellis codes for such channels [1]. TCM schemes optimized for the Rayleigh fading channel were presented in [2] and [3]. Most of these coding schemes use the "traditional" Ungerboeck approach – i.e., they involve doubling the constellation size over what is required for uncoded transmission and the use of a rate $k/(k+1)$ encoder to describe valid symbol sequences. However, if a rate $k/(k+1)$ code is used, the achievable minimum time diversity, L, is upper bounded by $L \leq \lfloor \nu/k \rfloor + 1$, where ν is the number of memory elements in the encoder. Therefore, most of the results obtained are far short of the cutoff rate (R_o) limit. To achieve a higher degree of minimum time diversity, we propose the use of I-Q TCM.

The basic idea of I-Q TCM is to use two independent encoders in parallel to select the in-phase and quadrature components of the transmitted sequence; this approach was used by Ho *et. al* to demonstrate the feasibility of dense constellations for fading channels. Using this approach, L is upper bounded by $L \leq \lfloor 2\nu/k \rfloor + 1$. Furthermore, no additional decoding complexity is required; complexity here is measured by the number of paths, excluding parallel ones, emanating from a given state times the number of states, per information bit. Although the proposed codes have two parallel encoders, they have the same complexity as a code with a single encoder with the same number of states because the number of bits entering each encoder is reduced and independent decoding is performed on the two decoders.

II. Designed Codes

Codes with bandwidth efficiencies of 1,2, and 3 bits/s/Hz and different constraint lengths were designed. If the bandwidth efficiency is not an even number, then the encoder operates every two signaling intervals, producing 4-dimensional coded signals.

[1]This work was supported in part by National Science Foundation grant NCR-8957623; also by the NSF Engineering Research Centers Program, CDR-8803012.

I-Q TCM schemes with a bandwidth efficiency of 1 b/s/Hz were designed using QPSK modulation; coding gains of 2-3 dB are achieved with respect to the traditionally designed schemes of equal complexity. Moreover, for the I-Q QPSK 64-state code, a BER of 10^{-5} can be achieved at $E_b/N_o \approx 7.5$ dB. This is only 2 dB from the cutoff rate limit.

Codes with a bandwidth efficiency of 2 b/s/Hz based on 16-QAM were also designed. They provided coding gains of ~ 4.5 dB over 8-PSK schemes [2]. Fig. 1 shows the simulated BER of the proposed codes and the codes from [2] for $\nu = 3, 4, 6$. Note that a BER of 10^{-5} can be achieved at $E_b/N_o \approx 10.5$ dB using the 64-state code. This is only 2.5 dB from the cutoff rate limit for 16-QAM signaling.

In this talk, both simulation and analytical results regarding the BER performance of the proposed codes will be presented. In addition, the effect of a non-uniform signal constellation and space diversity reception will be considered.

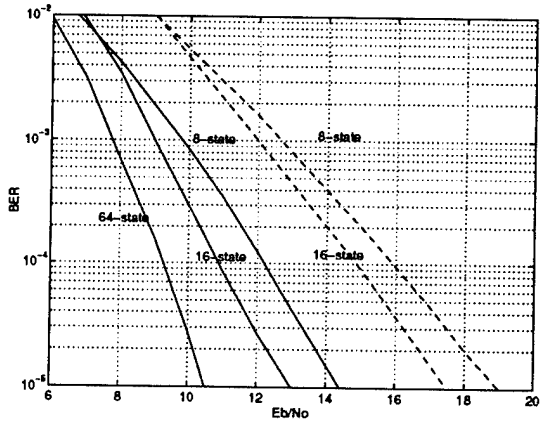


Figure 1: Comparison of proposed codes and codes from [2] for a bandwidth efficiency of 2 bits/s/Hz. The solid lines represent the proposed codes; the dashed lines represent codes from [2].

References

[1] Divsalar, D. and Simon, M., "The design of trellis coded MPSK for fading channels : Performance criteria," *IEEE Transactions on Communications,* September 1988, pp.1004-1012.

[2] Schlegel, C. and Costello, D. J., Jr., "Bandwidth efficient coding for fading channels: code construction and performance analysis," *IEEE Journal of Selected Areas in Communications,* December 1989, pp. 1356-1368.

[3] H. Jamali, and T. Le-Ngoc, Coded-Modulation Techniques for Fading Channels, Kluwer Academic Publishers, Boston, 1994.

[4] Ho, Paul K., Cavers, James K., and Varaldi, Jean L. "The effects of constellation density on trellis-coded modulation in fading channels," *IEEE Transactions on Vehicular Technology,* August 1993, pp. 318-325.

A Highly Adaptive High-Speed Wireless Transceiver

Gregory J. Pottie[1]
Electrical Engineering Dept.
UCLA, 405 Hilgard Ave.
Los Angeles CA 90095
pottie@icsl.ucla.edu

Abstract -- **In personal communications systems, users contend for the resources of frequency and time, with re-use determined by spatial separation, power allocation, antenna beam patterns, data rate, and the required signal to interference ratio for reliable operation. We describe a highly adaptive system with distributed control, i.e., each link is optimized independently, with coupling only via the mutual interference.**

I. SUMMARY

A high-performance experimental radio transceiver is under development at UCLA. It will include frequency hopping, variable bit and power allocation, channel coding, adaptive equalization, coherent M-QAM signaling, rapid channel probing, and adaptive transmitter and receiver antenna arrays. Most control functions will be distributed, requiring at most feedback between a transmitter and its intended receiver. Each link independently attempts to achieve the highest possible signal to interference ratio (SIR). The challenge is to design a set of adaptive algorithms that will interact in a stable fashion, while increasing the robustness and throughput of the network. We outline the major adaptive subsystems below.

In a radio network, frequency and time slots (channels) may be re-used at some distance due to propagation losses. In dynamic power and channel allocation (DPCA) algorithms, channels and transmitter powers are assigned to users so that all members of the network meet their own SIR requirement for reliable communication. A distributed DCPA algorithm has been developed, with the property that active users are protected from being dropped, at the cost of slightly reduced throughput relative to centralized control. In essence, new users may increase their power less aggressively than active users, and drop out when making little progress in their SIR. Convergence can be improved by probing the channel. The combination of the present SIR at a given power level, and the "resistance" of the system in the form of increased interference to each power increment are used in making a prediction of the final SIR.

In a frequency hopped system, we access many channels, and probe to predict the final SIR in each. We may then assign bits and power to maximize the throughput for this expected SIR distribution. Channel coding with interleaving across the frequency slots serves to realize the frequency diversity, provides coding gain, and some smoothing of small SIR estimation errors. We may arrange the frequency hops to be synchronous among cells, so that the same set of interferers is encountered on each hop. DPCA then reduces to the single channel form. A second option is to randomize the hopping patterns among the different cells. A combination of bit allocation and coding then produces a hybrid mix of interference averaging and interference avoidance, since we may choose not to allocate any power to those hops with large resistance. Simulations for the simpler case of choosing M out of N hops with equal bit allocation reveals that network throughput is very similar to the first option. However, this procedure is more robust with respect to channel variations since the set of channels occupied can be slowly changed, with the effect on any other being small since there is mutual interference in only one hop. We are also investigating hybrid fixed assignment/DCPA schemes, which alleviate certain difficulties that arise in admission and handoff.

Antenna arrays may also be used to suppress multipath and reduce interference. We propose to adapt both transmitter and receiver arrays using least squares techniques. Switching between sets of fixed beam patterns is not feasible for indoor systems, since we must gain some compromise benefit between diversity combining and interference cancellation, and the multipath has a very wide angular spread. Additionally, the human body interacts with the terminal to change the beam pattern. Another interaction is that between different pairs of communicating users. As the transmitter pattern of one array changes, so do the receiver patterns for all users in the vicinity. This in turn affects their transmitter patterns, as the latter may only be adapted based on the received signals. The antenna patterns must also react to changes in the power levels and/or channel assignments of the other users in the network. Thus, for indoor applications the interaction of these adaptive loops may be the dominant factor in the channel dynamics, rather than motion of the radios. We have investigated the dynamics of an adaptive antenna array with a variety of equalizer and transmitter adaptation options, with the conclusion that the ordinary LMS algorithm should be adequate. The imposition of orthogonality among channels within a cell together with the minimum SIR requirement for links to be declared feasible serve to decouple the users. The antenna arrays should be adapted on a time scale faster than power control, since the antenna gain affects the perceived path gains between users.

1. Supported by ARPA contract JFBI94-222/J4C942220

Adaptive Forward Error Control Schemes in Channels with Side Information at the Transmitter.

J. Larrea-Arrieta and D.J. Tait
Communication Research Group, Division of Electrical Engineering,
School of Engineering, University of Manchester, Dover St., M13 9PL, UK.
E-mail: larrea@comms.ee.man.ac.uk

Abstract - **The advisability of using adaptive strategies in channels with side information at the transmitter is considered. Different adaptive strategies are defined for block codes (BC) and punctured convolutional codes (PCC) and compared on throughput and bit error probability after decoding.**

I. INTRODUCTION.

Shannon [1] studied some communication systems with side information available to the transmitter, proving the positive effect of the side information on the achievable capacity. Nevertheless, it is not obvious how to take advantage of the side information in practice. The authors have been studying a system in which this side information is a true indication of the channel state and consequently, the transmitter adapts the parameters of an error control scheme in order to obtain the desired error rate, whilst keeping throughput, complexity and delay at acceptable levels.

Many different schemes can be proposed. A gross distinction between competing schemes is based on whether or not the side information is embedded in the transmission. In a previous paper [2] both possibilities were analysed and the scheme that did not transmit the side information was identified as superior.

It is the aim of this abstract to present some improvements on the previously presented adaptive schemes (all of them based on BC), and also to introduce some new approaches using PCC.

II. MODEL DESCRIPTION.

The transmitter strategy is simply to use different codes for different channel states. Because the block lengths are not the same for all codes a special metric is adopted at the receiver [3]. This metric depends on the joint probability of the message **m** and the received sequence **y** $P(\mathbf{m},\mathbf{y})$; we can write:

$$P(\mathbf{m},\mathbf{y}) = P(\mathbf{m})P(\mathbf{x}_m|\mathbf{m})P(\mathbf{y}|\mathbf{x}_m)$$

The main problem observed in the previous work was the possibility of losing synchronisation. Lack of synchronisation is detected when l blocks are found in error in m consecutive blocks. However, in the previous approach the data was sent directly to the data sink, whereas now it is kept until m blocks are decoded. At this point, provided that l blocks are not in error, the first data block is sent to the data sink. Otherwise re-synchronisation is achieved by moving a decision window bit by bit. This technique reduces the decoded error rate although it increases the delay and the complexity. Another possible scheme is developed using a tree structure. Here, a stack-like algorithm is implemented using the previously defined metric. The advantage of this scheme is that it achieves synchronisation automatically, provided the buffer does not overflow. The disadvantage is that a poor metric can easily lead to buffer overflow.

Finally a completely different scheme is proposed, this time using PCC. The main advantage of this scheme is that rate changes can be more gradual since they only involve a change in the puncturing matrix. Two techniques of this type are examined depending on whether the convolutional encoder is flushed after each block. Clearly, when the encoder is flushed the error performance is better and the complexity is lower but the throughput is less good. Without flushing the throughput increases but the decision technique is more troublesome and incorrect decoding can result more easily.

Further analysis allows the complexity of each scheme to be calculated and compared. However, since the main source of complexity resides in the decoder and this depends on the algorithm used, a completely fair comparison is very difficult. Nonetheless, the scheme with constant block length promises the least complexity, followed, quite closely, by the flushed PCC scheme. On the other hand, the schemes using either a constant number of information bits or a non-flushed PCC are quite complex due to the occasional necessity for data re-decoding or backtracking respectively.

III. RESULTS AND CONCLUSION.

The schemes presented are designed to achieve an error rate under 10^{-5} while working between 1 and 12 dB (E_s/N_o). The codes used are tabulated here:

BC:

Constant block length:

Scheme1.- (n,k,t)=(63,11,12), (63,24,7), (63,30,6) and (63,51,2).

Constant number of information bits:

(n,k,t)=(63,11,12), (31,11,5), (23,11,3) and (15,11,1)

Scheme2.- Buffer technique. Scheme3.- Tree technique.

PCC: (Buffer length 96)

Original code (n,k,K)=(4,1,5) and punctured to rate 1/3, 1/2 and 7/8.

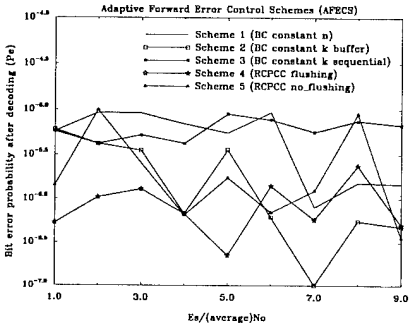
Scheme4.- Flushing. Scheme5.- Without flushing.

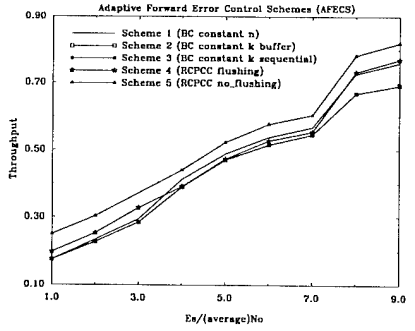
REFERENCES:

[1] Shannon, C.E.: "Channels with Side Information at the transmitter", IBM J. Res. Develop., Vol. 2, pp289-293, 1958.

[2] Larrea-Arrieta, J. and Tait, D.J.: "Comparison of Adaptive Error Control Schemes in Channels without Feedback", EUROCODE94, l'Abbaye de La Bussiere sur Ouche, Cote d'Or, France, 1994.

[3] Massey, J.L.: "Variable-Length Codes and Fano Metric", IEEE Trans. Inf. Theory, Vol. IT-18, pp196-198, 1972.





Novel Scarce-State-Transition Syndrome-Former Error-Trellis Decoding of (n, n-1) Convolutional Codes

L. H. Charles Lee*, David J. Tait**, Patrick G. Farrell**, and Paul S. C. Leung***

* School of Mathematics, Physics, Computing and Electronics, Macquarie University, Sydney, NSW 2109, Australia

** Department of Electrical Engineering, The University, Dover St., Manchester, M13 9PL, United Kingdom

*** School of Computer and Information Science, University of South Australia, Whyalla, SA 5608, Australia

Abstract - **A novel maximum-likelihood hard- and 8-level soft-decision scarce-state-transition (SST) type syndrome-former error-trellis decoding system of (n, n-1) convolutional codes with coherent BPSK signals for additive white Gaussian noise channels is proposed. The proposed system retains the same number of binary comparisons as the syndrome-former trellis decoding method of Yamada *et al.* [2]. Like the original SST-type register-exchange Viterbi decoding system [4], the proposed system also has the same advantage of drawing less power when implemented on CMOS LSI chips. A combination of the two techniques results a less complex and low power consumption decoding system.**

SUMMARY

In Viterbi algorithm decoding of (n, k) convolutional codes, the decoder carries out (2^k-1)-ary comparisons at each node of the encoder trellis [1]. The implementation of the Viterbi decoder becomes impractical for high-rate, powerful codes as the number of operations and memory path histories increase. In a 1983 paper, Yamada *et al.* [2] proposed a maximum-likelihood decoding system for rate-(n-1)/n convolutional codes, and the system performance was studied by Lee and Farrell [3]. The decoding system applies the Viterbi algorithm to the syndrome-former trellis of the code. Apparently, the number of trellis states is doubled, but the number of comparisons at each node is reduced to a binary comparison. Recently, Kubota *et al.* [4] proposed scarce-state-transition (SST) register-exchange (information bits are associated with surviving paths) Viterbi decoding system of reduced states, implemented on CMOS VLSI chips and consumed less power in the low bit-error-rate (BER) operating region when compared with a hypothetical register-exchange type of Viterbi decoder. A power consumption reduction of 40% at a bit error rate of 0.0001 can be achieved when operating at an information rate of 25 Mbit/s [4], and the measured power consumption with increasing channel noise was also reported in [4]. In this paper, we proposed a new maximum-likelihood SST-type trellis decoding system for rate-(n-1)/n convolutional codes, called the SST-type syndrome-former error-trellis decoding system. Our decoding system differs from the error-trellis syndrome decoding technique proposed by Reed *et al.* [5]. In their paper, the trellis is constrained and drawn from a k-input, (n-k)-output regulator circuit of a rate-k/n convolutional code and is only applicable to the class of systematic codes whereas our syndrome-former error-trellis is drawn from the n-input, single-output syndrome-former circuit of a rate-(n-1)/n systematic or non-systematic convolutional code. The new system is similar to the SST-type Viterbi decoding system [4] in that it has the advantage of drawing less power when implemented on CMOS chips and operated in a low BER condition. Like the Yamada decoding system [2], the new system has also retained a binary comparison at each trellis node and significantly reduces the decoding complexity. A combination of the two techniques results a less complex and low power consumption decoding system.

The simulated bit error probability performance of the proposed hard- and 8-level soft-decision decoding system, shown in Figure 1, for additive white Gaussian channels is presented. Furthermore, the implementation complexity of the new decoding system is compared with the SST-type register-exchange Viterbi decoding system.

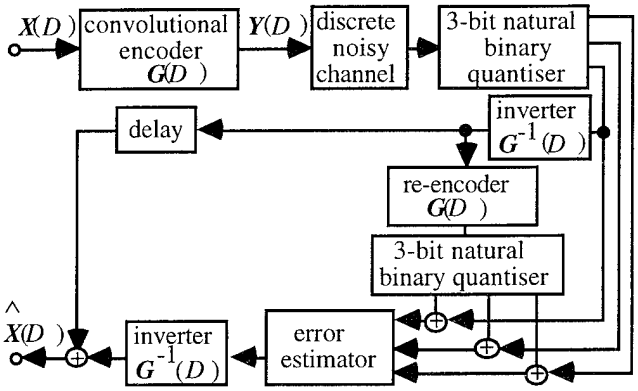


Fig. 1 Model of an eight-level soft-decision SST-type syndrome-former error-trellis decoding system

REFERENCES

[1] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", *IEEE Trans.*, vol. IT-13, pp. 260-269, 1967.

[2] T. Yamada, H. Harashima, and H. Miyakawa, "A new maximum likelihood decoding of high rate convolutional codes using a trellis", *Trans. Inst. Electron. & Commun. Eng. Japan*, 66A, pp. 11-16, 1983.

[3] L. H. C. Lee, and P. G. Farrell, "Error performance of maximum-likelihood trellis decoding of (n, n-1) convolutional codes : A simulation study", *IEE Proc.-F*, vol. 134, no. 7, pp. 1673-680, 1987.

[4] S. Kubota, S. Kato, and T. Ishitanil "Novel Viterbi decoder VLSI implementation and its performance", *IEEE Trans.*, vol. COM-41, pp. 1170-1178, 1993.

[5] I. S. Reed, and T. K. Turong, "Error-trellis syndrome decoding techniques for convolutional codes", *IEE Proc.-F*, vol. 123, no. 2, pp. 77-83, 1985.

Construction of Trellis Codes at High Spectral Efficiencies for Use with Sequential Decoding[1]

Fu-Quan Wang
VoCAL Technologies, Ltd.
1576 Sweet Home Road
Buffalo, New York 14228

Daniel J. Costello, Jr.
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Abstract - Sequential decoding of trellis codes at high spectral efficiencies is investigated and large constraint length trellis codes for two dimensional and four dimensional constellations are constructed for use with sequential decoding. It is shown that the channel cut-off rate bound can be achieved using constraint lengths between 16 and 19 with sequential decoding at a bit error rate of 10^{-5} - 10^{-6}.

I. INTRODUCTION

Recently, it has been shown that sequential decoding is a good alternative to Viterbi decoding for trellis codes and significant coding gains can be achieved using sequential decoding with large constraint length trellis codes compared to Viterbi decoding with small constraint lengths[1]-[3]. The channel cut-off rate R_O is the maximum rate at which the average number of computations for sequential decoding is bounded. Thus, R_O is regarded as the maximum rate for which reliable communication can be achieved with reasonable complexity. Trellis codes for 8-PSK and 16-QAM constellations with large constraint lengths were constructed for use with sequential decoding in [3,4]. For these constellations, it was shown that the channel cut-off rate bound can be achieved using large constraint length codes with sequential decoding at a bit error rate (BER) of 10^{-5} - 10^{-6} on Additive White Gaussian Noise (AWGN) channels[3]. In this paper, we discuss the construction of trellis codes at higher spectral efficiencies for use with sequential decoding.

II. SEQUENTIAL DECODING AND THE FANO METRIC

The calculation of the Fano metric at high spectral efficiencies and for multidimensional signal constellations is discussed. We show that the computation of the Fano metric for multidimensional signals can be decomposed into a simpler calculation for the constituent two dimensional signals, and thus that the computational complexity of decoding a multidimensional trellis code using sequential decoding is comparable to decoding a two dimensional trellis code. Simulation results show that the computational distribution for sequential decoding of multidimensional trellis codes at high spectral efficiencies can be very well approximated by a Pareto distribution. This implies that the code construction criteria for trellis codes with sequential decoding derived for small spectral efficiencies can also be applied to the construction of trellis codes at high spectral efficiencies.

III. CODE CONSTRUCTION

The Random Search (RS) algorithm proposed in [3] is investigated and modified to construct trellis codes at high spectral efficiencies. This work was motivated by the random coding principle that an arbitrary selection of code symbols will produce a good code with high probability. In the code construction algorithm, the sequential decoding performance was used as the criterion for selecting good codes. Thus the algorithm works well as long as the performance of the code can be evaluated using sequential decoding.

In practice, rotational invariance is a desirable property. It allows the decoder to synchronize quickly at startup or after a phase slip. In [5], a simple method was proposed to check the rotational invariance of a given code. It was shown that rotationally invariant trellis codes with large constraint lengths can be found in a systematic way. In the modified RS (MRS) algorithm, this method is used to insure that rotationally invariant trellis codes are found.

IV. RESULTS AND DISCUSSIONS

The MRS algorithm was used to construct two dimensional and four dimensional trellis codes. 180^O rotationally invariant linear trellis codes for two dimensional constellations with constraint lengths 16-19 were obtained. Simulation results show that the cut-off rate bound can be achieved using sequential decoding with a constraint length 16 code at a BER of 10^{-5} and with a constraint length 19 code at a BER of 10^{-6}. Similarly, 180^O rotationally invariant linear trellis codes and 90^O rotationally invariant nonlinear trellis codes for four dimensional constellations were found using the MRS algorithm. The partitioning and labeling of the four dimensional constellations are the same as Wei's [6]. It was also shown that the channel cut-off rate bound can be achieved using sequential decoding with four dimensional codes using constraint lengths between 16 and 19 at BER's of 10^{-5} - 10^{-6}.

ACKNOWLEDGMENT

The authors would like to thank Dr. G. D. Forney, Jr. for his continuing interest in this work. The comments of Dr. Lance C. Perez are also greatly appreciated.

REFERENCES

[1] G. J. Pottie and D. P. Taylor, "A Comparison of Reduced Complexity Decoding Algorithms for Trellis Codes", *IEEE J. Select. Areas Commun.*, pp. 1369-1380, Dec. 1989.

[2] F. Q. Wang and D. J. Costello,Jr., "Erasurefree Sequential Decoding of Trellis Codes", *IEEE Trans. Inform. Theory*, pp. 1803-1817, Nov. 1994.

[3] F. Q. Wang and D. J. Costello, Jr., "Probabilistic Construction of Large Constraint Length Trellis Codes for Sequential Decoding", *IEEE Trans. Commun.*, Aug. 1995.

[4] J. Porath and T. Aulin, "Algorithmic Construction of Trellis Codes", *IEEE Trans. Commun.*, pp. 649-654, May 1993.

[5] F. Q. Wang and D. J. Costello,Jr., "New Rotationally Invariant Four Dimensional Trellis Codes", *IEEE Trans. Inform. Theory*, to appear.

[6] L. F. Wei, "Trellis Coded Modulation with Multidimensional Constellations", *IEEE Trans. Inform. Theory*, pp. 483-501, Jul. 1987.

[1] This work was supported by NASA grant NAG 5-557 and NSF grant NCR 89-03429

Real Number Convolutional Code Correction and Reliability Calculations in Fault-Tolerant Systems

Robert Redinbo
Department of Electrical
and Computer Engineering
University of California
Davis, CA 95616 USA

An efficient fault-detecting methodology, algorithm-based fault tolerance, may be extended to include error correction of the output data in a protected linear processing system by coupling a high-rate real convolutional code with a smoothed Kalman recursive estimation technique [1]. A completely protected fault-tolerant linear processing system involving error correction is shown in Figure 1 where it is guaranteed that no miscorrected data leave the configuration if at most one box-surrounded subsystem fails at a time.. The real convolutional code dictates the comparable parity streams computed in two ways, forming the syndrome stream that is passed to the Fixed-Lag Corrector when values exceed threshold settings. The block processing and down sampling features of the convolutional code permit the overhead area to be from 20-50% of the main processing area.

The reliability function of the protected system is calculated when failures are assumed to arrive according to a Poisson process with uniform rate per unit area. Arrivals in the main processing part are assumed independent of those in the protection overhead parts leading to respective arrival rates **a** and **b** as shown in Figure 1. The reliability levels are computed using iterated integrals over appropriate regions and conditional probability expansions. The guard space of the convolutional code is described by parameter **c**. Reasonable lower bounds on the reliability levels which depend only on the arrival rates **a** and **b** and guard parameter **c** are established by bounding individual conditional events.

$$R(t) \geq e^{-(a+b)t}\left\{ 1 + bt + \sum_{n=1}^{\lfloor t/c \rfloor} a^n \int_{\xi_1=0}^{t-nc} \int_{\xi_2=\xi_1+c}^{t-(n-1)c} \cdots \int_{\xi_i=\xi_{i-1}+c}^{t-(n-i+1)c} \cdots \int_{\xi_n=\xi_{n-1}+c}^{t-c} \left[1 + b\xi_1\right] \cdot \left\{ \prod_{i=2}^{n} \left[1 + b(\xi_i - \xi_{i-1} - c)\right] \right\} \cdot \left[1 + b(t - \xi_n - c)\right] \, d\xi_n \, d\xi_{n-1} \cdots d\xi_1 \right\}$$

These reliability bounds are easily calculated employing a standard computer algebra package on a workstation. Typical results are shown in Figure 2 for a real code based on a binary rate 5/6 Berlekamp-Preparata-Massey burst correcting code. The failure intensities (FITS) of rates **a** and **b** are indicated and the guard space parameter **c** is related to a 30ns clocking period. The results for a coded versus uncoded system are displayed separately because of the large differences in scales of the reliability logarithms. There is a dramatic improvement in levels due to correction even when the additional area of the protection overhead is included.

This research was supported by grant NSF MIP-92 15957.

[1] G. R. Redinbo."Real Convolutional Codes, Time-Varying Errors and Kalman Error-Correction for Fault Tolerance and Communications Systems Applications," *IEEE International Symposium on Information Theory,* Trondheim, Norway, pg 162, June, 1994.

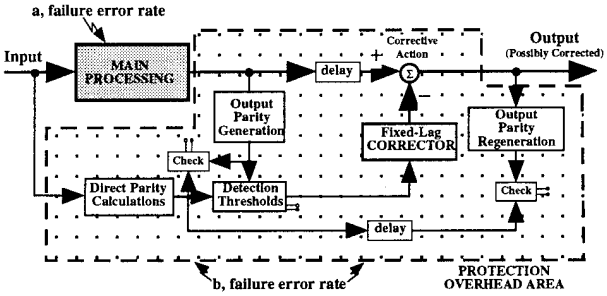
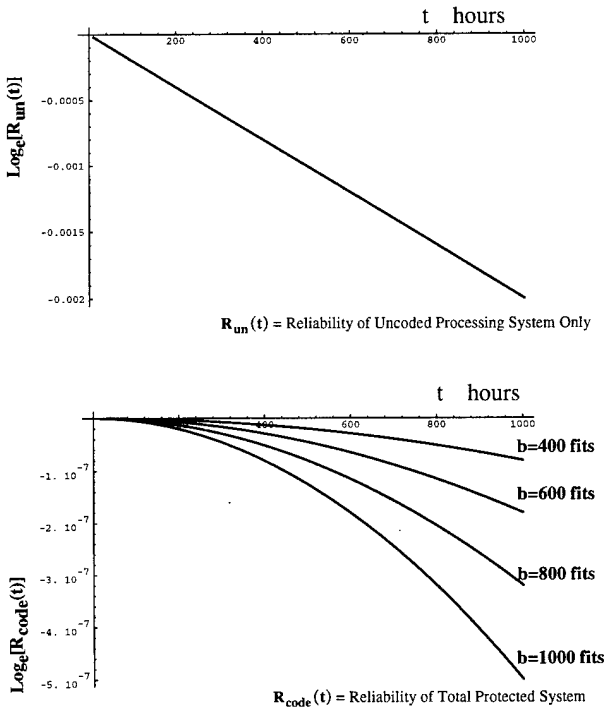


Figure 1 : Protected, Correcting Processing System



$R_{un}(t)$ = Reliability of Uncoded Processing System Only

$R_{code}(t)$ = Reliability of Total Protected System

PARAMETERS:

a = $2\ 10^{-6}$ = 2000 fits ; Processing Section Failure Rate
b = 400, 600, 800, 1000 fits ; Corrector Section Failure Rate
T = 30ns. ; Processing Sample Period

Figure 2 : Example Comparisons of Reliabilities for Coded and Uncoded Systems

Bidirectional Viterbi Decoding Algorithm with Repeat Request and Estimation of Unreliable Region

Masato Tajima

Dept. of Elect. & Comp. Sci., Faculty of Eng., Toyama University, 3190 Gofuku, Toyama 930, JAPAN

Abstract -- A bidirectional Viterbi decoding algorithm for framed information with repeat request which is an extended version of the Yamamoto-Itoh scheme is presented. A method to estimate the unreliable region in a received frame using the proposed algorithm is also presented.

Ⅰ. BIDIRECTIONAL VITERBI DECODING ALGORITHM WITH REPEAT REQUEST

Yamamoto and Itoh proposed a convolutionally coded ARQ scheme with Viterbi decoding in order to improve the reliability of convolutional coding/Viterbi decoding [1]. In this scheme, all survivors are labeled either GOOD or BAD and retransmission is requested if all survivors are labeled BAD. On the other hand, it is known that a string of received symbols corresponding to a frame of information bits augmented by tail bits can be decoded from both directions simultaneously [2], [3]. Taking these facts into consideration, we propose a bidirectional Viterbi decoding algorithm for framed information with repeat request.

In this scheme, a received frame is accepted if the forward and reverse decoders can meet without declaring retransmission in the course of decoding. The ML path is decoded by a trace-back method or some equivalent one, after determining the node x_0 on the ML path at the point of junction. The proposed scheme is most efficiently applied for the case where one of the two decoders (let this decoder be the forward one) stops at some node level t_0 declaring retransmission, while the other decoder (i.e., the reverse decoder) can proceed up to level t_0. In this case, only a part of the frame (i.e., $[0, t_0]$) is needed to retransmit. For the retransmitted data, the two decoders resume decoding from both directions (Note : the reverse decoder can continue decoding operations using the survivors and their metrics computed till then). If the same situation happens after the first retransmission, partial retransmission is requested again and the procedure is repeated until the two decoders can join. It is derived analytically that the averaged quantity of retransmission per frame in the proposed scheme is approximated by $(L'/4)p_X$ (L' : length of a coded frame, p_X : probability of retransmission per frame).

Ⅱ. ESTIMATION OF UNRELIABLE REGION

When retransmission is requested at node level t_0 in the Yamamoto-Itoh scheme, we know that some noisy region has started before t_0 in the corresponding trellis. That is, the noisy region in the received data has been roughly estimated in "one" direction. Making use of this fact, we show that the unreliable region is estimated as an "interval" by using a bidirectional Viterbi decoding algorithm with repeat request, especially in the case where the two decoders, one with a flag of retransmission and the other without it, can join in the course of decoding. In such a case, by tracing the ML path forward and backward, we can find the node $x_1^{\#}$ (level $t_1^{\#}$) at which the label of the ML path turns BAD for the first time and the node $x_i^{\#\#}$ (level $t_i^{\#\#}$) at which the second best path for node $x_i^{\#}$ has diverged from the ML path in forward (i=1) and reverse (i=2) decoding. Then the interval $[t_1^{\#\#}, t_2^{\#\#}]$ is regarded as an unreliable region. The relation between $t_i^{\#}$ and $t_i^{\#\#}$ is given by the following lemma :

<Lemma> $t_1^{\#\#} \leqq t_2^{\#}$ and $t_1^{\#} \leqq t_2^{\#\#}$.

In order to realize the above idea, we incorporate a new bidirectional Viterbi decoding scheme into the proposed algorithm. In this scheme, only the metrics of the survivors are remembered until the two decoders join and after that they serve as preliminary computations for determining the nodes on the ML path. It is shown that the scheme is very convenient for tracing the ML path forward or backward.

REFERENCES

[1] H. Yamamoto and K. Itoh, " Viterbi Decoding Algorithm for Convolutional Codes with Repeat Request, " IEEE Trans. Inform.Theory, vol.IT-26, no.5, pp.540-547, 1980.

[2] J. Belzile and D. Haccoun, " Bidirectional Breadth-First Algorithms for the Decoding of Convolutional Codes, " IEEE trans. Commun., vol.41, no.2, pp.370-380, 1993.

[3] M. Tajima, " Bidirectinal Viterbi Decoding Algorithm and Its Applications, " Technical Report of IEICE, IT93-110, pp.25-30, 1994 (in Japanese).

Reduced Complexity Algebraic Type Viterbi Decoding of q-ary Convolutional Codes

Kamil Sh. Zigangirov

Department of Telecommunication Theory, University of Lund, Box 118, S-221 00 Lund, Sweden

Abstract — **A reduced complexity algebraic type algorithm is described for decoding of convolutional codes over $GF(q)$, $q > 2$. It is founded on the same principles as algebraic-sequential decoding [1, 2]. It is proved that for large q, the algorithm has better complexity-reliability tradeoff than the conventional Viterbi algorithm.**

SUMMARY

Let us consider a discrete symmetric memoryless channel (DSMC) with input and output alphabets $A = \{0, 1, \ldots, q-1\}$, where $q > 2$ is a prime or a power of a prime. By definition of a DSMC, each output symbol of the channel depends only on the corresponding input. The conditional probability p_{ij} of receiving symbol j, $j \in A$, provided that the symbol i, $i \in A$, has been transmitted, is given

$$p_{ij} = \begin{cases} 1-\varepsilon, & \text{if } i = j \\ \varepsilon/(q-1), & \text{otherwise} \end{cases}$$

Let $\mathbf{v} = (\mathbf{v}_0, \mathbf{v}_1, \ldots) = (v_{01}, v_{02}, \ldots v_{0c}, v_{11}, v_{12}, \ldots, v_{1c}, \ldots)$ be the output (code) sequences at the output of the convolutional rate $R = b/c$ memory m encoder, $\mathbf{u} = (\mathbf{u}_0, \mathbf{u}_1, \ldots) = (u_{01}, u_{02}, \ldots, u_{0b}, u_{11}, u_{12}, \ldots, u_{1b}, \ldots)$ be the input (data) sequence. Then $\mathbf{v} = \mathbf{u}G$, where G is a semi-infinite generator matrix having $b \times c$ submatrices as elements. All elements of $\mathbf{v}$, $\mathbf{u}$ and G are elements in $GF(q)$ and all operations are performed over $GF(q)$. Let $\mathbf{r} = (\mathbf{r}_0, \mathbf{r}_1, \ldots) = (r_{01}, r_{02}, \ldots, r_{0c}, r_{11}, r_{12}, \ldots, r_{1c}, \ldots)$ be the received sequence, $r_{ij} \in GF(q)$.

We introduce the binary error locator sequence $\mathbf{l} = (\mathbf{l}_0, \mathbf{l}_1, \ldots) = (\ell_{01}, \ell_{02}, \ldots, \ell_{0c}, \ell_{11}, \ell_{12}, \ldots \ell_{1c}, \ldots)$, $\ell_{ij} \in \{c, e\}$, where $\ell_{ij} = c$ ("correct") if r_{ij} is received correctly and $\ell_{ij} = e$ ("erroneous") otherwise. A sequence $\mathbf{l}$ is considered as survived, if there exists a code sequence $\mathbf{v}$, which symbols coincide with the symbols of $\mathbf{r}$ in the positions where $\mathbf{l}$ have symbol c and not in the other positions. If the decoder knows the error locator sequence, it can correctly decode the information sequence, if it can do maximum-likelihood (Viterbi) decoding.

The set of survived error locator sequences can be represented as a set of paths in a binary error locator tree. The decoding algorithm can then be treated as a search algorithm in the error locator tree. We consider a list-decoding type algorithm: In every decoding step the decoder selects the L most likely sequences in the error locator tree and calculate its survived successors.

To characterize the algorithm we introduce the characteristic parameter $z = (1-R)/\log_q(q-1)$ and the **effective decoding distance** d_{ef}, which plays the same role as the free distance does for the Viterbi algorithm: The decoder corrects all combinations of $\left\lfloor \frac{d_{ef}-1}{2} \right\rfloor$ or less errors.

Theorem 1: There exists a rate R q-ary time-invariant convolutional code, whose effective distance resulting from algebraic type Viterbi decoding of list size L is lowerbounded by the unequality

$$d_{ef} \geq \frac{z \log_2 L}{h_2(z)} + const,$$

where $h_2(z) = -z\log_2 z - (1-z)\log_2(1-z)$ and the constant does not depend of L.

Comparison with the Costello bound for the free distance of convolutional codes, shows that for large q the algebraic type Viterbi decoding gives essentially better complexity-reliability tradeoff than conventional Viterbi decoding.

Using modified random coding technique we obtain a random coding bound and an expurgation bound for the probability of decoding error for the algorithm.

To formulate the expurgation bound, we introduce the **algebraic** computational cutoff rate:

$$R_0^{(a)} = \max\{1 - z_0 \cdot \log_q(q-1),\ R_0\},$$

where z_0 is the largest root of the equation

$$z \log \frac{z}{\sqrt{\varepsilon}} + (1-z)\log \frac{1-z}{\sqrt{1-\varepsilon}} = 0$$

and R_0 is the "conventional" computational cutoff rate of the DSMC:

$$R_0 = 1 - 2\log_q\{(1-\varepsilon)^{1/2} + [\varepsilon(q-1)]^{1/2}\}.$$

Theorem 2 (expurgation bound): For a q-ary DSMC, there exist a rate R q-ary time-invariant convolutional code and a L-list algebraic type decoder, whose burst error probability is upperbounded by

$$P_e \leq L^{-\gamma_{ex}(R)+o(1)},\ R \leq R_0^{(a)},$$

where

$$\gamma_{ex}(R) = -\frac{z\log\sqrt{\varepsilon} + (1-z)\log\sqrt{1-\varepsilon}}{h_2(z)}.$$

Theorem 3 (random coding bound): For a q-ary DSMC, there exists a rate R q-ary time-invariant convolutional code and a L-list algebraic type decoder, whose burst error probability upperbounded by

$$P_e \leq L^{\gamma_r(R)+o(1)},\ R > R_0^{(a)},$$

where

$$\gamma_r(R) = \frac{z\log_2 \frac{z}{\varepsilon} + (1-z)\log_2 \frac{1-z}{\varepsilon}}{h_2(z)}.$$

REFERENCES

[1] K.Sh. Zigangirov, "Mathematical analysis of algebraic-sequential decoding", *Problems of Information Transmission*, vol. 28, No 1, pp. 3–13, 1992.

[2] K.Sh. Zigangirov, "Algebraic-sequential decoding - ideas and results", in *Communication and Cryptology*, Kluwer Academic Publishers, pp. 451–459, 1994.

On The General Threshold Decoding Rule and Related Codes

Xiao-Hong Peng and **Patrick G. Farrell**

Communication Research Group, Division of Electrical Engineering

University of Manchester, Manchester, M13 9PL, UK

Abstract ---- **A new concept of λ-order orthogonalization, and a general threshold decoding method is introduced. Many codes decodable using general threshold decoding can be constructed and are superior in performance to other majority-logic decodable codes.**

I. INTRODUCTION

Many efforts have been made [1] [2] in trying to decode more codes using majority-logic decoding with nonorthogonal parity-check sums. This approach, however, requires more parity-check sums for each error digit than orthogonal majority-logic decoding, and is only applicable to a small class of codes. Rudolph and Robinson [3] claimed that any decoding function for a linear binary code can be realized as a weighted majority-logic decoding. However, each weight element in this scheme is a function of all the 2^{n-k} parity-check sums, so it will involve a large number of computations in addition to the majority-logic operation. The method presented in this paper can be viewed as an alternative to applying the threshold decoding method to more types of codes, but involving fewer computations.

II. THE GENERAL THRESHOLD DECODING RULE

Let syndrome digits $s_0, s_1, \ldots, s_{n-k-1}$ be Boolean variables and $\hat{e}_l(s_0, s_1, \ldots, s_{n-k-1})$ the decoding function for the error digit in the lth position of the received vector. The general one-step threshold decoding (simply GTD) rule is defined to be

$$\hat{e}_l(s_0, s_1, \ldots, s_{n-k-1}) = \begin{cases} 1 & \text{if } \sum_{j=0}^{J-1} A_j \geq T \\ 0 & \text{if } \sum_{j=0}^{J-1} A_j < T \end{cases} \quad (1)$$

where $A_j = a_0 s_0 + a_1 s_1 + \ldots + a_{n-k-1} s_{n-k-1}$ $(a_i \in GF(2))$, A_j is a parity-check sum, J is the number of parity-check sums on e_l, and T is the threshold value.

Definition: A code is said to be λ-order orthogonal if for any set of parity-check sums, e.g., a set on e_l, e_l appears in each parity-check sum, but no other error digits appear more than λ ($\lambda \geq 1$) times in the set.

This definition covers both orthogonal ($\lambda = 1$) and nonorthogonal ($\lambda > 1$) cases. In the table below, the parameters J and t_c for three decoding methods, majority-logic decoding for orthogonalizable codes (simply orth-M-L), majority-logic decoding for nonorthogonalizable codes (non-orth-M-L) [1] and GTD for λ-order orthogonalizable codes (λ-orth-threshold), are listed for the purpose of comparisons.

It is easy to show that GTD is always applicable where majority-logic decoding can be used, and requires fewer parity-check sums than the second case.

	orth-M-L	non-orth-M-L	λ-orth-threshold
J	$\geq 2t_c$	$\geq 2t_c\lambda$	$\geq (2t_c\text{-}1)\lambda + 1$
$t_c \leq$	$\left\lfloor \frac{n-1}{2(\bar{d}-1)} \right\rfloor$	$\leq \left\lfloor \frac{n-1}{2(\bar{d}-1)} \right\rfloor$	$\leq \left\lfloor \frac{\lambda(n+\bar{d}-1)-\bar{d}}{2\lambda(\bar{d}-1)} \right\rfloor$

III. THRESHOLD DECODABLE CODES

It can be shown [4] that there are many codes which can be decoded in one step using the rule given by (1).

* Any $(2^m - 1, 2^m - m - 1)$ Hamming code is an $(m-1)$-order orthogonal code which can be decoded by one-step threshold decoding with the threshold value $T = m$. Hence, only one threshold gate is required for hardware implementations in this case, instead of a total of $\sum_{i=0}^{m-2} J^i$ majority-logic gates [5] when Hamming codes were originally treated as $(m-1)$-step majority-logic decodable codes [6].

* The $(v_d, b_d, r_d, k'_d, \lambda)$-configurations [7] with parameters, $v_d = \xi$, $b_d = \binom{\xi}{J}$, $k'_d = J$, $\lambda = J - 1$, where $\xi \geq 3$, $3 \leq J \leq \lfloor \xi / 2 \rfloor$, constructs a class of $(n, k, d) = \left(\binom{\xi}{J} + \xi, \binom{\xi}{J}, 4\right)$ SEC-DED codes which can be decoded by one-step threshold decoding with the threshold value $T = J$, where $\xi = n-k$.

A comparison of some threshold decodable codes with existing majority-logic decodable codes, and a list of multiple-error-correcting threshold decodable codes, is given in [4].

REFERENCES

[1] L.D. Rudolph, "A class of majority logic decodable codes," IEEE Trans. Inform. Theory, IT-13, pp.305-307, April 1967.

[2] C.L. Chen, "Majority decoding with nonorthogonal parity checks," IEEE Trans. Inform. Theory, pp.757-759, Nov. 1976.

[3] L.D. Rudolph and W.E. Robbins, "One-step weighted-majority decoding," IEEE Trans. Inform. Theory, pp.446-448, May 1972.

[4] X.-H. Peng, " Low complexity error control for block codes," Ph.D. thesis, University of Manchester, 1994.

[5] S. Lin and D.J. Costello, Jr., *Error Control Coding: Fundamentals and Applications,* Englewood Cliffs, NJ: Prentice-Hall, 1983.

[6] J.L. Massey, *Threshold Decoding*. MIT press, 1963.

[7] M. Hall, *Combinatorial Theory*. Blaisdell Publishing, Waltham Mass., 1967.

Optimum Distance Profile Trellis Encoders for Sequential Decoding [1]

Per Ljungberg

Department of Information Theory, Lund University, P.O. Box 118, S-221 00 Lund, Sweden. email: perl@dit.lth.se

Abstract — **New trellis encoders over various lattice partitions having optimum distance profile (ODP) and large constraint lengths are constructed. They are attractive to use in combination with sequential decoding algorithms since their ODP property ensures good computational performance.**

I. Introduction

Trellis coded modulation (TCM) can achieve significant coding gains over uncoded transmission without any bandwidth expansion. For error rates of the order of 10^{-6}, the gap between the Shannon limit and uncoded high-rate QAM-signaling is approximately 9 dB, being the maximum achievable coding gain for any coded modulation scheme operating in this region. A perhaps more realistic performance limit is the computational cut-off rate, R_0, beyond which the average number of computations for sequential decoding becomes unbounded. The possible coding gain under the R_0-criterion is 7.5 dB. It can be separated in two parts, viz., fundamental coding gain and shaping gain [1]. The maximum values of these gains are approximately 6 dB and 1.5 dB, respectively.

The signal constellation can be viewed as a set of 2^{n+1} points from an infinite N-dimensional lattice Λ. A sublattice Λ' of Λ induces a partition Λ/Λ' of Λ into $|\Lambda/\Lambda'|$ cosets of Λ'. The output of a rate $R_c = \frac{k}{k+1}$ convolutional encoder is used to select one of the 2^{k+1} cosets. Then the $n-k$ uncoded bits select one of the points in the specified coset. The fundamental coding gain is determined by the convolutional encoder and the lattice partition, whereas the shaping gain depends on the bounding region of the constellation.

The aim of this work is to increase the fundamental coding gain compared to current systems by increasing the number of states in the encoder. The decoding is performed with a sequential decoder since its decoding effort is essentially independent of the number of states. It is well-known that the code should have an optimum distance profile (ODP) in order to minimize the average number of computations for the sequential decoder.

II. Search for ODP-encoders

It is convenient to search for $R_c = \frac{k}{k+1}$ encoders on a systematic feedback form. The corresponding generator matrix is

$$G(D) = \left(I_k \mid H^i(D)/H^0(D)\right) \ , \ i = 1, \cdots, k \ ,$$

where

$$H^i(D) = h_0^i + h_1^i D + \cdots + h_\nu^i D^\nu$$

are the parity check polynomials in the delay operator D. The search is performed as follows:

Assume that the set of ODP-encoders of constraint length ν is known. Form the 2^{k+1} possible extensions of every encoder within this set and calculate their distance profiles. Retain the encoders with the best distance profile, these form the set of ODP-encoders of constraint length $\nu+1$.

Later on, we will require the encoder to be on a feedforward form. The transformation from a systematic rational to a non-systematic polynomial encoding matrix is performed as follows:

The encoding matrices $G(D)$ and $G_1(D)$ are equivalent if $G_1(D) = T(D)G(D)$ and $T(D)$ nonsingular. If $T(D)$ is chosen to be $I_k \cdot H^0(D)$, $G_1(D)$ has a feedforward realization.

Let $G_1(D) = A(D)\Gamma(D)B(D)$ be the Smith factorization of $G_1(D)$. Choosing $G_2(D)$ as the k upper rows of $B(D)$ ensures that $G_2(D)$ is basic and equivalent to $G(D)$. Using the algorithm in [2], we are now able to construct a minimal-basic matrix $G_3(D)$ that is equivalent to $G(D)$.

III. Performance evaluation

The distance spectrum of the encoders are computed with the FAST algorithm [3], which is considered to be very efficient. However, it requires knowledge of the smallest number of steps needed to drive the encoder from a certain state to the zero state. This is very difficult to compute for an encoder with feedback, but trivial for a feedforward encoder, which is the reason for the encoding matrix transformation. The complexity of the matrix transformation is small compared to the increase in number of node visits that would occur if another search algorithm would be employed.

In the table we give ODP-encoders over $\mathbb{Z}^2/2R\mathbb{Z}^2$ maximizing the effective coding gain, γ_{eff}, at an error rate of 10^{-6}. Following [1], we compute the three first components of the distance spectrum, $\tilde{N}_i$. The dominant error coefficient is starred, and the parity check polynomials are given in octal notation.

We will continue to search for large constraint length ODP-encoders over 4- and 8-dimensional lattice partitions and perform simulations of their performance.

ν	H^2	H^1	H^0	d^2_{free}	$\tilde{N}_0$	$\tilde{N}_1$	$\tilde{N}_2$	γ_{eff}
5	20	12	41	5	*4	16	72	3.98
6	100	2	45	5	*4	20	116	3.98
7	100	32	261	7	*20	104	520	4.98
8	100	272	601	7	4	*32	232	5.43
9	100	622	1215	8	12	*96	480	5.66
10	100	1062	2275	8	4	*48	256	5.84
11	1500	3052	6601	8	*4	8	88	6.02

[1] This research was supported in part by the Swedish Research Council for Engineering Sciences under the Grants 92-661 and 94-83.

References

[1] G. D. Forney, Jr., "Coset Codes – Part I: Introduction and geometrical classification", *IEEE Trans. Inform. Theory*, **IT-34**, pp. 1123-1151, September 1988.

[2] G. D. Forney, Jr., "Convolutional codes I: Algebraic structure", *IEEE Trans. Inform. Theory*, **IT-16**, pp. 720-738, November 1970.

[3] M. Cedervall and R. Johannesson, "A fast algorithm for computing distance spectrum of convolutional codes", *IEEE Trans. Inform. Theory*, **IT-35**, pp. 1146-1159, November 1989.

Error Burst Detection with High-Rate Convolutional Codes

Amir Said[1]

amir@densis.fee.unicamp.br, Faculty of Electrical Engineering
State University of Campinas (UNICAMP), Campinas, SP 13081, Brazil

SUMMARY

Binary block codes have been extensively used for error detection, and amongst the most popular we have the CRC (cyclic redundancy check) codes [1, 5]. In contrast, convolutional codes have been used almost exclusively for error correction (the exception being some hybrid applications). A reason for this is that the most popular convolutional codes have a rate that is far too low (e.g., 1/2, 2/3) for the cases where only error detection is desired. Furthermore, while there has been a variety of algebraic methods to design good high-rate block codes, the design of good high-rate convolutional codes seems to be more difficult. Nevertheless, progress has been made in that field [2], which opens the practical possibility of using convolutional codes for error detection.

Clearly, one crucial requirement for the use in (only) error detection is low decoding complexity. We show that for this specific application the decoding complexity of convolutional codes is practically equal to the coding complexity, which is very small. Thus, the encoder/decoder can be implemented directly in hardware (as exemplified in Fig. 1), or use efficient software decoding techniques like those used for CRC error detection codes [5]. Different encoder/decoder implementations are considered.

By studing the properties of high-rate convolutional codes for the purpose of error detection, we show their potential advantages over block codes. For instance, one fundamental limitation of block codes for burst error detection comes from the fact that the decoder can only flag an error at the end of each data block. Consequently, there is conflict between the minimization of the probability of undetected errors, and the minimization of the average error detection delay—which is the amount of time taken by the decoder to flag an error after it occurred. To minimize the delay it is necessary to use short blocks, but, for a given code rate, short block codes may not be powerful enough to detect long bursts. The convolutional codes can be powerful enough to detect those error bursts, and still flag the errors with small delays.

In addition, this study gives a deeper view of CRC codes—which happen to be a special case in a class of codes that we call *unit-rate convolutional codes*. Thus, for the extension of CRCs we can employ techniques used for convolutional codes, like the use of unit-memory [3, 2] or cyclic time-varying codes.

Certain general error detection capabilities of the convolutional codes are derived, as shown in the example below.

Proposition 1 *The fraction $F_u(L,n,k,m)$ of error patterns with duration L not detected by a (n,k,m) convolutional code is:*

$$F_u(L,n,k,m) = \begin{cases} 0, & L \le m, \\ \dfrac{2^k-1}{(2^n-1)^2\,2^{n(m-1)}}, & L = m+1, \\ \dfrac{(2^k-1)^2\,2^{k(L-m-2)}}{(2^n-1)^2\,2^{n(L-2)}}, & L > m+1. \end{cases}$$

As explained above, CRC codes can be considered special convolutional codes, and, as expected, $F_u(L,n,k,m)$ gives the performance a (n_c,k_c) q-ary CRC code when $n=k=1$, $2^k=q$, and $m=n_c-k_c$. A more detailed analysis of particular codes, based on worst-case scenarios [4], can also be used to analyze the performance, or define code design objectives.

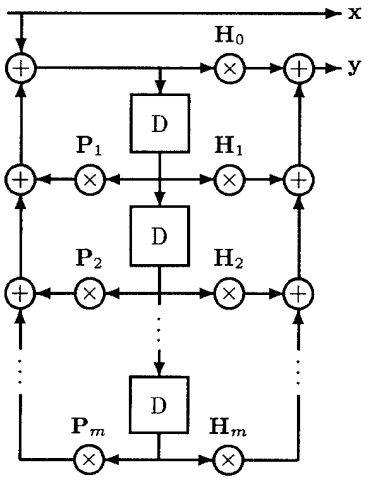


Fig. 1 – Encoder/decoder implementation for systematic codes: $\mathbf{x}$ is a vector with k data bits, and $\mathbf{y}$ contains $n-k$ parity-check bits; the matrices $\mathbf{P}_i$ and $\mathbf{H}_i$ have dimensions $k \times k$ and $k \times (n-k)$, respectively.

[1] This work was supported by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

REFERENCES

[1] W.W. Peterson and E.J. Weldon Jr., *Error Correcting Codes*, 2nd ed., MIT Press, 1972.

[2] A. Said and R. Palazzo Jr., "Using combinatorial optimization to design good unit-memory convolutional codes," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1100–1108, May 1993.

[3] L.N. Lee, "Short unit-memory byte-oriented convolutional codes having maximal free distance," *IEEE Trans. Inform. Theory*, vol. 22, pp. 349–352, May 1976.

[4] J.K. Wolf and D. Chun, "The single burst error detection performance of binary cyclic codes," *IEEE Trans. Commun.*, vol. 42, pp. 11–13, Jan. 1994.

[5] T.V. Ramabadran and S.S. Gaitonde, "A tutorial on CRC computations," *IEEE Micro*, 62–75, Aug. 1988.

A Table-Based Reduced Complexity Sequential Decoding Algorithm

Havish Koorapaty[1], Donald L. Bitzer, Ajay Dholakia, Mladen A. Vouk
Dept. of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA

Abstract — **The table-based soft-decision convolutional decoding method presented here performs a reduced tree search as compared to the M-algorithm. The degree of tree-searching is adapted to the state of the channel by using a syndrome sequence and precomputed information stored in a memory table. This results in a significant reduction in computational complexity while maintaining bit error rate performance comparable to the M-algorithm on a Rayleigh flat-fading channel.**

I. Table-Based Algorithm

We restrict the presentation to rate one-half convolutional coding, although the algorithm presented may be extended to other coding rates. Let the encoded sequence be $\mathbf{v} = (v_0^{(1)}, v_0^{(2)}, v_1^{(1)}, v_1^{(2)}, \ldots) = (\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots)$, where $\mathbf{v}_i = (v_i^{(1)}, v_i^{(2)})$. Let the corresponding received sequence of real-valued (soft) symbols at the receiver be $\mathbf{r} = (r_0^{(1)}, r_0^{(2)}, r_1^{(1)}, r_1^{(2)}, \ldots) = (\mathbf{r}_0, \mathbf{r}_1, \ldots)$. The sequence $\mathbf{r}$ may be hard-quantized (sign detector) to generate a binary received sequence $\mathbf{b} = (b_0^{(1)}, b_0^{(2)}, b_1^{(1)}, b_1^{(2)}, \ldots) = (\mathbf{b}_0, \mathbf{b}_1, \ldots)$. The data-independent syndrome sequence $\mathbf{s} = (s_0, s_1, s_2, \ldots)$ is defined as $\mathbf{s} = \mathbf{b}\mathbf{H}^T$, where $\mathbf{H}$ is the parity check matrix. A section of the syndrome sequence $\mathbf{s}_{[t,\, t+\tau]}$ is generated from $\mathbf{b}_{[t-\nu,\, t+\tau]}$ where ν is the constraint length of the encoder.

In the M-algorithm, at each time-step t, each of the M paths $\mathbf{p}^{(i)}_{[0,\, t-1]}$, $1 \le i \le M$, from time $t-1$ is extended with both *branch extensions* in the code tree to form a total of $2M$ paths from which the best M paths are chosen [1]. The table-based algorithm stores M_c paths at any given time with $M_c \le M$ and differs from the M-algorithm as follows [2]:

For each path $\mathbf{p}^{(i)}$ a syndrome sequence $\mathbf{s}^{(i)}_{[t,\, t+\gamma-1]}$ is calculated. If $\mathbf{s}^{(i)}_{[t,\, t+\gamma-1]} = \mathbf{0}$, $\mathbf{p}^{(i)}_{[0,\, t-1]}$ is extended only with one branch extension $\mathbf{p}^{(i)}_t = \mathbf{b}_t$, i.e., no additional paths are generated. If $\mathbf{s}^{(i)}_{[t,\, t+\gamma-1]} \neq \mathbf{0}$, a finite section of $\mathbf{s}^{(i)}$ is used to retrieve a memory table entry that indicates if a single branch extension must be considered with $\mathbf{p}^{(i)}_t = \mathbf{b}_t$ as above or if both branch extensions of $\mathbf{p}^{(i)}_{[0,\, t-1]}$ must be considered. If $\mathbf{s}_{[t,\, t+\gamma-1]} = \mathbf{0}$ for the path $\mathbf{p}$ with the best metric, the other $M_c - 1$ paths are discarded, and the best path is simply extended with the received symbols $(\mathbf{b}_t, \mathbf{b}_{t+1}, \ldots)$ until the next non-zero syndrome bit occurs. In this stage, $M_c = 1$ and the algorithm is in a depth-first search mode until the next nonzero syndrome bit occurs.

II. Performance

A framed system with interleaving similar to the IS-54 North American digital cellular standard is used, with $F = 84$ information bits and $\nu = 5$ tail bits in each frame. A Rayleigh time correlated flat-fading model is used for the channel. At the receiver, ideal estimation of the fading coefficients is assumed. Simulations were performed for the M-algorithm ($M = 8$), the Viterbi algorithm and the Table-based algorithm (syndrome length $\beta = 15$, $M = 8$, $\gamma = 8$). Figure 1 shows the decoded information bit error rates for the three algorithms and Figure 2 shows the average number of paths per time-step for which both branch extensions were considered which is representative of the reduction in computation.

[1] Havish Koorapaty is with the Dept. of Electrical and Computer Engineering. His work was supported in part by Ericsson Inc.

References

[1] J.B. Anderson and S. Mohan, *Source and Channel Coding: An Algorithmic Approach*, Kluwer Academic Publishers, Boston, 1991.

[2] H. Koorapaty, D.L. Bitzer, A. Dholakia, M.A. Vouk. Table-Based Reduced Complexity Sequential Decoding with Soft Decisions. Tech. Rep. TR-95-06, Dept. Comp. Sci., NCSU, Raleigh, NC, 1995.

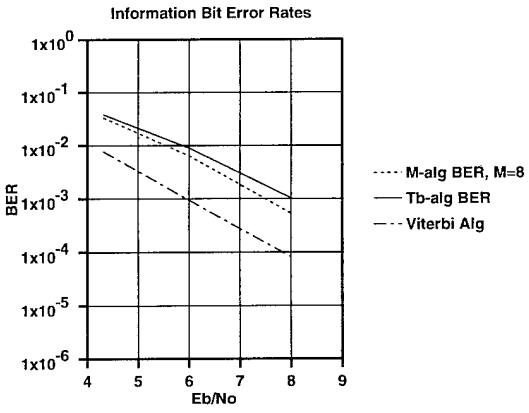


Figure 1: Information Bit Error Rate performance

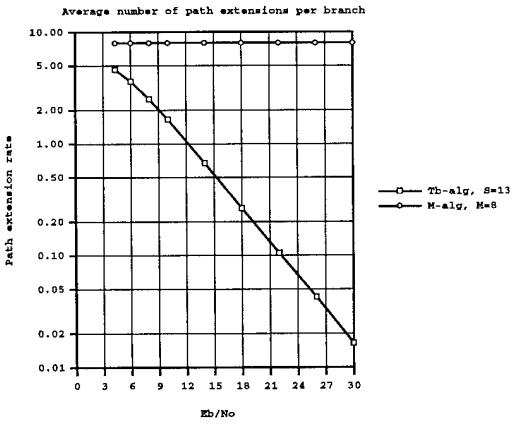


Figure 2: Average number of path extensions performed per branch

Characterization of the Bayes estimator and the MDL estimator for Exponential Families

Jun-ichi Takeuchi

Theory NEC Laboratory[1], RWCP[2]. 4-1-1 Miyazaki, Miyamae-ku, Kawasaki, Kanagawa 216, Japan.

We analyze the relationship between the MDL (Minimum Description Length) estimator (posterior mode) and the P.B.E. (projected Bayes estimator) for exponential families, where the P.B.E. is obtained by projecting the B.E. (Bayes estimator, i.e. posterior mixture) onto the original exponential family and is equal to the B.E. under a certain condition.

An exponential family is defined as $S(\Theta) = \{p(x|\theta) = \exp(\theta^i x_i - \psi(\theta))|\theta \in \Theta\}$ (the range of random variable x is $\mathcal{X} \subseteq \Re^n$), where Θ is a connected subset of $\Re^n$ and $\theta^i x_i$ denotes $\sum_{i=1}^n \theta^i x_i$. θ is called the canonical parameter or the θ-coordinates. We also define the expectation parameter η (η-coordinates) as $\eta_i = E_\theta(x_i)$. θ and η form a dual pair from the point of view of information geometry [1]. Let g^{ij} denote the Fisher information matrix with respect to η. We define $g_{ij} = E_\theta((x_i-\eta_i)(x_j-\eta_j))$ and $T_{ijk} = E_\theta((x_i-\eta_i)(x_j-\eta_j)(x_k-\eta_k))$. Note that g_{ij} equals the inverse of g^{ij}. We refer to a function f which maps $\bigcup_{i=0,1,...} \mathcal{X}^i$ to $\mathcal{H}$ (any set of probability distributions) as an estimator. We let $f[x^N]$ denote the image of x^N by f and $f[x^N](x)$ denote its density at x. Hereafter, we let $\hat{\eta}$ denote the maximum likelihood estimate (MLE) for η, and $\hat{T}$ and $\hat{g}$ denote their values at $\eta = \hat{\eta}$ respectively. Finally, we let 'ln' denote the natural logarithm.

We define the B.E. with the prior $w(\theta)d\theta$ as $f_w[x^N](x) = p_w(x|x^N) = \int p(x|\theta)w(\theta|x^N)d\theta$, where $w(\theta|x^N)$ denotes the posterior density of θ. We let w_J denote the Jeffreys prior ($\propto (\det|g_{ij}|)^{1/2}$). Among the B.E.'s f_{w_J} is particularly important, since it is known ([2]) that $\sup_{\theta\in\Theta} E_\theta(\sum_{i=0}^N D(p(\cdot|\theta)||f_w[x^i]))$ (D denotes Kullback-Leibler divergence) is asymptotically minimized when $w = w_J$, i.e. f_{w_J} has minimax property. We define the projection of f_w to $S(\Theta)$ (let $\tilde{f}_w$ denote it) as $\tilde{f}_w[x^N] \equiv \arg\min_{p\in S(\Theta)} D(f_w[x^N]||p)$. We refer to $\tilde{f}_w$ as the P.B.E. with the prior w. Define $\tilde{\eta} \equiv \int_\Theta \eta(\theta)w(\theta|x^N)d\theta$, then we can show $\eta_i(\tilde{f}_w[x^N]) = \tilde{\eta}_i$ under a certain weak condition. Using this fact, we can show the following for $\tilde{f}_w$.

Theorem 1 *Under a certain weak condition, $\eta_i(\tilde{f}_w[x^N]) = \hat{\eta}_i + N^{-1}\partial \ln w(\hat{\theta})/\partial\theta^i + O(N^{-3/2}\sqrt{\ln N})$ holds. When $w(\theta)$ is uniform over Θ, $\eta_i(\tilde{f}_w[x^N]) = \hat{\eta}_i + O(e^{-CN})$ holds.*

Corollary 1 *Under a certain weak condition, $\eta_i(f_L[x^N]) = \hat{\eta}_i + \hat{T}_{ijk}\hat{g}^{jk}/2N + O(N^{-3/2}\sqrt{\ln N})$ holds.*

The MDL estimator with respect to prior $w(\theta)d\theta$ is defined as $\hat{\theta}_{mdl} = \arg\min_{\theta\in\Theta}(-\ln p(x^N|\theta) - \ln w(\theta) + \ln\det|g^{ab}(\theta)|/2)$ and $f^w_{mdl}[x^N] = p(\cdot|\hat{\theta}_{mdl})$ ([3, 5]). When $w(\theta)d\theta \propto d\eta$, we let f^η_{mdl} denote f^w_{mdl}. We show the following for f^w_{mdl}.

Lemma 1 *$\eta_i(f^w_{mdl}[x^N]) = \hat{\eta}_i + N^{-1}\partial \ln w(\theta)/\partial\theta^i - \hat{T}_{ijk}\hat{g}^{jk}/2N + O(N^{-2})$ holds. In particular, $\eta_i(f^\eta_{mdl}[x^N]) = \hat{\eta}_i + \hat{T}_{ijk}\hat{g}^{jk}/2N + O(N^{-2})$ holds.*

We let f^θ_{bc} denote the bias corrected MLE with respect to θ (see for example [1]). Concerning the expectation parameter of f^θ_{bc}, we can show the following.

Lemma 2 *$\eta_i(f^\theta_{bc}[x^N]) = \hat{\eta}_i + \hat{T}_{ijk}\hat{g}^{jk}/2N + O(N^{-2})$ holds.*

We can show the following theorems using Theorem 1, Corollary 1 and Lemmas 1, 2.

Theorem 2 *Under a certain weak condition, the differences between $\eta(f^\eta_{mdl}[x^N])$, $\eta(f_{w_J}[x^N])$, and $\eta(f^\theta_{bc}[x^N])$ are of order $O((\ln N)^{1/2}N^{-3/2})$.*

Theorem 3 *Under a certain weak condition, the differences between $\eta(f^{w_J}_{mdl}[x^N])$, $\eta(\tilde{f}_{d\theta}[x^N])$, and $\hat{\eta}$ are of order $O(1/N^2)$.*

We summarize the above two theorems in Table 1, where we ignore $o(1/N)$ terms. These results suggest a striking symmetry between the two estimators.

prior $wd\theta$	$d\theta$	$\sqrt{\det\|g_{ij}\|}d\theta$	$d\eta$
f_w	η-unbiased	θ-unbiased	
f^w_{mdl}		η-unbiased	θ-unbiased

Table 1: dependency of estimators on prior

We exhibit an example (Bernoulli sources) below. Define $S = \{p(x|\eta) = \eta^x(1-\eta)^{1-x}|0 < \eta < 1\}$ ($x = 0, 1$), and $\theta = \ln(\eta/(1-\eta))$. In this case, $\eta(f_{w_J}[x^N]) = (k+0.5)/(N+1)$ holds, which is well known as the Laplace estimator. Next, we derive the MDL estimator. The total description length for MDL with respect to η is $-k\ln\eta - (N-k)\ln(1-\eta) - (\ln\eta + \ln(1-\eta))/2 = -(k+0.5)\ln\eta - (N-k+0.5)\ln(1-\eta)$. This is minimized when $\eta = (k+0.5)/(N+1)$, which strictly equals $\eta(f_{w_J}[x^N])$. Finally, we derive the bias-corrected MLE. Since $T_{111} = E_\theta((x-\eta)^3) = \eta(1-\eta)(1-2\eta)$, we have $\hat{T}_{111}\hat{g}^{11} = 1-2\hat{\eta} = 1-2k/N$. Hence, we have $\eta_{bc} = k/N + (1-(2k/N))/2N + O(1/N^2) = \eta(f_{w_J}[x^N]) + O(1/N^2)$.

Theorem 2 implies that we can approximate the P.B.E. with Jeffreys prior (which is hard to derive in general) simply by deriving an appropriate MDL estimator or a bias-corrected MLE. Some important topics of future research are as follows: To analyze the difference between the B.E. and the P.B.E. and to evaluate directly the performance of f_{w_J}, f^η_{mdl}, and f^θ_{bc}. The argument in this abstract is restricted to the case in which the class of sources is an i.i.d. exponential family. The same problem for Markov sources is discussed in [4]. We would also like to analyze the case for curved exponential families.

ACKNOWLEDGEMENTS

The author would like to give his sincere gratitude to Dr. Hiroshi Nagaoka, Dr. Tsutomu Kawabata and Dr. Naoki Abe for their suggestions and advices.

REFERENCES

[1] S. Amari, *Differential-geometrical methods in statistics (2nd pr.)*, Lecture Notes in Statistics, Vol.28, Springer-Verlag, 1990.

[2] B. Clarke. & A. Barron, "Jeffreys prior is asymptotically least favorable under entropy risk," *the JSPI*, 1994.

[3] J. Rissanen "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.

[4] J. Takeuchi & T. Kawabata, "Approximation of Bayes code for Markov sources," *IEEE ISIT*, 1995.

[5] C. Wallace & P. Freeman, "Estimating and inference by compact coding," *J. Roy. Statist. Soc. B*, vol. 49. No.3 pp. 240-265, 1987.

[1]c/o C&C Res. Labs., NEC Corp. e-mail tak@SBL.CL.nec.co.jp

[2]RWCP: Real World Computing Partnership

Concept Learning using Complexity Regularization[1]

Gábor Lugosi[2] and Kenneth Zeger[3]

[2] Dept. of Mathematics, Faculty of Elect. Engineering, Technical University of Budapest, Hungary.
email: lugosi@vma.bme.hu.

[3] Coordinated Science Lab., Dept. of Elect. and Comp. Engineering, University of Illinois, Urbana-Champaign, IL 61801
email: zeger@uiuc.edu.

Abstract — **We apply the method of complexity regularization to learn concepts from large concept classes. The method is shown to automatically find the best balance between the approximation error and the estimation error. In particular, the error probability of the obtained classifier is shown to decrease as $O(\sqrt{\log n/n})$ to the achievable optimum, for large nonparametric classes of distributions, as the sample size n grows.**

In pattern recognition—or concept learning—the value of a $\{0,1\}$-valued random variable Y is to be predicted based upon observing an $\mathcal{R}^d$-valued random variable X. A *prediction rule* (or *decision*) is a function $\phi : \mathcal{R}^d \to \{0,1\}$, whose performance is measured by its error probability $\mathbf{P}\{\phi(X) \neq Y\}$. The error probability $L^* = \mathbf{P}\{g^*(X) \neq Y\}$ of the optimal decision g^* is called the Bayes risk. Assume that a training sequence

$$D_n = ((X_1, Y_1), \ldots, (X_n, Y_n))$$

of independent, identically distributed random variables is available, where the (X_i, Y_i) have the same distribution as (X, Y), and D_n is independent of (X, Y). A *classifier* is a function $\phi_n : \mathcal{R}^d \times (\mathcal{R}^d \times \{0,1\})^n \to \{0,1\}$, whose error probability is the random variable $L(\phi_n) = \mathbf{P}\{\phi_n(X, D_n) \neq Y | D_n\}$.

The method of empirical risk minimization picks a classifier from a class $\mathcal{C}$ of functions $\mathcal{R}^d \to \{0,1\}$ that minimizes the empirical error probability over $\mathcal{C}$. More precisely, define the empirical error probability of a decision ϕ by $\hat{L}_n(\phi) = (1/n)\sum_{i=1}^n I_{\{\phi(X_i) \neq Y_i\}}$, where I denotes the indicator function. Let $\tilde{\phi}_n$ denote a classifier chosen from $\mathcal{C}$ by minimizing $\hat{L}_n(\phi)$, i.e., $\hat{L}_n(\tilde{\phi}_n) \leq \hat{L}_n(\phi)$, $\phi \in \mathcal{C}$. Vapnik and Chervonenkis [4], [5] proved distribution-free exponential inequalities for empirical error minimization. One of the implications is that $\mathbf{E}L(\tilde{\phi}_n) - \inf_{\phi \in \mathcal{C}} L(\phi) \leq c\sqrt{(V \log n)/n}$, where V is the VC dimension of the class $\mathcal{C}$ and c is a universal constant (independent of the distribution). Thus, the error probability of the empirically chosen decision is always within $O(\sqrt{\log n/n})$ of that of the best in $\mathcal{C}$. Unfortunately, if $V < \infty$, then for some distributions, $\inf_{\phi \in \mathcal{C}} L(\phi)$ may be arbitrarily far from L^* On the other hand, if $V = \infty$, then $L(\tilde{\phi}_n) - \inf_{\phi \in \mathcal{C}} L(\phi)$ will be large for some distributions [3], [5].

A possible solution to this problem may be derived from the idea of *structural risk minimization* (Vapnik and Chervonenkis [5]), also known as *complexity regularization* (see Barron [1], Barron and Cover [2]). The basic idea is to minimize the sum of the empirical error and a term corresponding to the "complexity" of the candidate classifier. In our application, this complexity is a simple function of the VC dimension of the class from which the candidate classifier is taken.

[1]The research was supported in part by the National Science Foundation under Grants No. NCR-92-96231 and INT-93-15271.

Theorem 1 *Let $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \ldots$ be a sequence of classes of classifiers whose VC dimensions $V_1, V_2, \ldots$ are finite. Let ϕ_n^* be the classification rule based on structural risk minimization. Then for all n,*

$$\mathbf{E}\{L(\phi_n^*)\} - L^* \leq \inf_{k \geq 1} \left(\sqrt{\frac{16V_k \log n + 8(k+11)}{n}} + \left(\inf_{\phi \in \mathcal{C}^{(k)}} L(\phi) - L^* \right) \right).$$

This result is close on spirit of those obtained by Barron [1], and Barron and Cover [2], who select a classifier from a countable list of candidates by minimizing the sum of the empirical error and a properly chosen penalty. A significant difference is that the method we study here does not restrict the search to a countable set of candidates, allowing thus better approximation ability.

Corollary 1 *Let $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \ldots$ be a sequence of classes of classifiers such that the VC dimensions $V_1, V_2, \ldots$ are all finite. Assume further that the Bayes rule is contained in the union of these classes, i.e., $g^* \in \mathcal{C}^* \stackrel{\text{def}}{=} \cup_{j=1}^{\infty} \mathcal{C}^{(j)}$. Let K be the smallest integer such that $g^* \in C^{(K)}$. Then for every n, the error probability of the classification rule based on structural risk minimization, ϕ_n^*, satisfies*

$$\mathbf{E}L(\phi_n^*) - L^* \leq 4\sqrt{\frac{V_K \log n + K/2 + 6}{n}}.$$

Corollary 1 shows that the rate of convergence is always of the order of $\sqrt{\log n/n}$, and the constant factor V_K depends on the distribution. The number V_K may be viewed as the inherent complexity of the Bayes rule for the distribution. One great advantage of structural risk minimization is that it finds automatically where to look for the optimal classifier.

References

[1] A. R. Barron. Complexity regularization with application to artificial neural networks. In G. Roussas, editor, *Nonparametric Functional Estimation and Related Topics*, pages 561–576, Dordrecht, 1991. NATO ASI Series, Kluwer Academic Publishers.

[2] A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37:1034 – 1054, 1991.

[3] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36:929–965, 1989.

[4] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.

[5] V. N. Vapnik and A. Ya. Chervonenkis. *Theory of Pattern Recognition*. Nauka, Moscow, 1974. (in Russian); German translation: *Theorie der Zeichenerkennung*, Akademie Verlag, Berlin, 1979.

Minimax Redundancy through Accumulated Estimation Error

Bin Yu[1]

Statistics Department, University of California, Berkeley, CA 94720-3860, USA.
Email: binyu@stat.berkeley.edu

Abstract — **In this paper, minimax expected redundancies over memoryless source classes of smooth densities are studied, through their connections with accumulated prediction errors and using available techniques from nonparametric statistics. To derive lower bounds on the minimax expected redundancy rates, two methods are used and compared. One is the Assouad's technique from statistical density estimation and the other is the information-theoretic (generalized) Fano's inequality. Both methods are applied to hypercube sub-classes and a connection between Assouad's and Fano's is established using a packing number result from error-correcting coding theory. Finally, optimal (rate) codes, which achieve the minimax rate lower bounds on expected redundancy, are formed based on optimal density estimators.**

Summary

Minimax expected redundancy was studied as early as 1973 by Davisson [3] for Markov sources. For other regular parametric source classes, minimax lower bounds on expected redundancy follow from [2] (see also [7]). Lower bounds on expected redundancy, minimax or Rissanen's pointwise one ([8]), play an important role in Rissanen's Stochastic Complexity Theory since they justify, together with codes achieving these lower bounds, the complexity measures of the source classes. While Rissanen's pointwise lower bound has difficulty extending to nonparametric (or smooth) classes of densities, the minimax approach has its natural counterpart for those classes. The minimax expected redundancy rates measures the complexities of the nonparametric source classes, just as in the regular parametric case.

For a given memoryless data string $x^n = (x_1, x_2, ..., x_n)$ and without knowing the density f on [0,1] which generated the data, we would like to compress the data in an efficient way; that is, we would like to find a joint density q_n on the n-tuples, which may be regarded as corresponding to a prefix code with code length $-\log q_n(x^n)$ for an n-tuple x^n, such that its expected redundancy is small over, say, source classes of smooth densities. On the other hand, the expected redundancy of q_n can be decomposed into accumulated prediction or estimation error $\sum_t E_{f^{t-1}} D(f, \hat{q}_{t-1})$ because our source is memoryless with density f on [0,1]. The estimation error $E_{f^{t-1}} D(f, \hat{q}_{t-1})$ in terms of information divergence D is very much related to other errors which correspond to real distances such as the Hellinger distance

$$H^2(u, v) = \int (\sqrt{v} - \sqrt{w})^2.$$

Density estimation errors in terms of Hellinger distance have been well studied in the statistical density estimation literature (cf. Birgé [1]) for classes of smooth densities. Therefore, to obtain minimax lower bounds on redundancy over the same type of smooth density classes, we may borrow techniques from density estimation.

One well-known technique is Assouad's method. It bounds the minimax estimation error from below by the average estimation error over a sub-class, i.e., by the Bayes estimation error corresponding to the uniform prior on the sub-class. This sub-class is indexed by a hypercube whose dimension can be optimalized in the end. More importantly, the sub-class is chosen in such a way that the Hellinger distances of densities on the neighboring vertexes of the hypercube can be calculated easily. This Assouad's method can also be understood through another useful and well-known technique called Le Cam's method which deals with two sets of hypotheses. (cf. [9] and [10]).

The other (more powerful) technique is the generalized Fano's inequality (cf. [4] and [6]), which deals with finite number of hypotheses. Using a packing number result from the error-correcting coding theory ([5]), a sub-set of the vertexes of the hypercube class can be selected to apply the Fano's inequality to, giving the same rate lower bounds on redundancy as Assouad's (cf. [9]).

However, since minmax and summation do not exchange, lower bounds on redundancy do not follow directly from the lower bounds in density estimation, but require a separate Assouad's type of arguments.

As to upper bounds, when the density f is bounded away from zero, the minimax rate lower bounds on expected redundancy can be achieved if we take q_t as any optimal rate density estimator based on the first t observations. Hence the minimax rates in the lower bounds are the optimal redundancy rates over classes of smooth densities.

[1]This work was partially supported by ARO Grant DAAH04-94-G-0232 and NSF Grant DMS-9322817.

References

[1] Birgé, L. "On estimating a density Hellinger distance and some other strange facts," *Probab. Th. Rel. Fields* **71** 217-291, 1986

[2] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inform. Theory*, **36** 453-471, May, 1990.

[3] L. Davisson, "Universal Noiseless Coding," *IEEE Trans. Inform. Theory*, **19** 783-795, Nov, 1973.

[4] Devroye, L. *A course in density estimation.* Boston: Birkhauser, 1987.

[5] Gilbert, E. N. "A comparison of signaling alphabets," *Bell System Tech. J.* **31** 504-522, 1952.

[6] T. S. Han and S. Verdú "Generalizing the Fano inequality," *IEEE Trans. Inform. Theory*, **40** 1247-1250, July 1994.

[7] N. Merhav and M. Feder, "The Minimax Redundancy is a Lower Bound for Most Sources," Preprint.

[8] J. Rissanen, "Stochastic complexity and modeling," *Annals of Statistics*, **14** 1080-1100, 1986.

[9] B. Yu, "Assouad, Fano, and Le Cam," To appear in *Festschrift in Honor of L. Le Cam on His 70th Birthday*, 1995.

[10] B. Yu and T. Speed, " Data compression and histograms," *Probab. Th. Rel. Fields*, **92** 195-229, 1992.

An Algorithm for Designing a Pattern Classifier by Using MDL Criterion

Hideaki Tsuchiya Shuichi Itoh Takeshi Hashimoto

Dept. of Electronic Engineering, University of Electro-Communications,
1-5-1, Chofugaoka, Chofu-shi, Tokyo, 182 Japan

Abstract — **The algorithm for designing a pattern classifier, which uses MDL criterion and a binary data structure, is proposed. The algorithm gives a partitioning of the space of the K-dimensional attribute and gives an estimated probability model for this partitioning. The volume of bins in this partitioning is asymptotically upper bounded by $\mathcal{O}((\log N/N)^{K/(K+2)})$ for large N in probability, where N is the length of training sequence. The redundancy of the code length and the divergence of the estimated model are asymptotically upper bounded by $\mathcal{O}(K(\log N/N)^{2/(K+2)})$. The classification error is asymptotically upper bounded by $\mathcal{O}(K^{1/2}(\log N/N)^{1/(K+2)})$.**

I. Introduction

Pattern classification is a problem of assigning each data *attribute* $\boldsymbol{X}$, which is typically obtained from measuring instruments, a *label* Y which indicates the class that data belongs to[1]. Suppose that the label Y assumes a value y in a binary set $\mathcal{Y} = \{0,1\}$ according to a probability distribution $P_Y(y)$ and the observed attribute, denoted by $\boldsymbol{X} = (X_1, \cdots, X_K)$, assumes a value $\boldsymbol{x} = (x_1, x_2, \cdots, x_K)$ in a subset $\mathcal{X} = [0,1)^K$. The optimal decision rule is expressed as $\hat{y} = f(\boldsymbol{x}) = \arg\max_{y\in\mathcal{Y}} P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$. Thus, the problem of designing a pattern classifier turns out to be the problem of estimating $P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ from a given training sequence $\{(\boldsymbol{X}_i, Y_i), i = 1, \cdots, N\}$ of length N. We assume that $(\boldsymbol{X}_i, Y_i)$ are independent and identically distributed.

In the case of discrete-valued attributes, Quinlan and Rivest[2] first showed the possibility of applying Rissanen's Minimum Description Length principle[3] to the construction of decision trees for the pattern classification problem.

We propose an algorithm based on MDL two stage coding to design a pattern classifier using a sequence of independent training examples. A binary tree structure is used to represent the partition and MDL criterion is used to optimize the tree. The asymptotic performance is derived.

II. Algorithm

Our strategy is as follows: For one-dimensional continuous valued X, its range $\mathcal{X} = [0,1)$ is partitioned into finite s subsets called bins $b_i, i = 1, \cdots, s$, and then the probability model $\hat{P}_{Y|X}(y|x)$ is obtained by the histogram-like estimator. This approach is used by Rissanen[4]. However, the complexity of the estimated model soon becomes excessively large unless the model complexity is appropriately controlled. Here, we restrict the partitioning such that $|b_i| = 2^{-d_i}, i = 1, \cdots, s$, where $d_i \in \boldsymbol{N}$. With this restriction, bins are represented as leaf nodes of a complete binary tree. Let $\boldsymbol{t} = t_1 t_2 \cdots t_d$ be a path, string of edges of length d in the tree leading from root node to a leaf. With each leaf node represented by a path $\boldsymbol{t}$, we associate a bin $b(\boldsymbol{t}) = [0.\boldsymbol{t}00\cdots, 0.\boldsymbol{t}11\cdots)$. Let L_t be the cost of the leaf node $\boldsymbol{t}$, that is, the code length associated with bin b_t. The minimization of the sum of the costs can be done easily with the dynamic programing which uses the recursive structure of the binary tree[5].

The binary tree structure is extended to K-dimensional attribute case if we let each node $\boldsymbol{t} = t_{11}t_{21}\cdots t_{K1}t_{12}t_{22}\cdots t_{K2}t_{13}\cdots$ represent a bin $b(\boldsymbol{t}) = \{(x_1, x_2, \cdots x_K) | x_1 \in b(\boldsymbol{t}^1), x_2 \in b(\boldsymbol{t}^2), \cdots, x_K \in b(\boldsymbol{t}^K)\}$, where $\boldsymbol{t}^i = t_i t_{2i}\cdots$.

The computational time complexity of the algorithm is dramatically improved over that of [4].

III. Main Result

Theorem: Assume that $p_{\boldsymbol{X}}(\boldsymbol{x})$ and $P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ are upper and lower bounded and their first differentials are upper bounded, then we have

$$|b(\boldsymbol{t}^{(N)})| \le C\left(\frac{\log N}{N}\right)^{\frac{K}{K+2}} \quad a.s.,$$

$$\frac{L}{N} - \int_{\mathcal{X}} H(P_{Y|\boldsymbol{X}}(0|\boldsymbol{x}))p_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \le \mathcal{O}\left(K\left(\frac{\log N}{N}\right)^{\frac{2}{K+2}}\right), a.s.,$$

and

$$\int_{\mathcal{X}} \sum_{y=0,1} P_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) \log \frac{P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}{\hat{P}_{Y|\boldsymbol{X}}(y|\boldsymbol{x})} p_{\boldsymbol{X}}(\boldsymbol{x})d\boldsymbol{x} \le \mathcal{O}\left(K\left(\frac{\log N}{N}\right)^{\frac{2}{K+2}}\right) \quad a.s.$$

for all sufficiently large N, where $|b(\boldsymbol{t}^{(N)})|$ is the volume of the bin associated with the node $\boldsymbol{t}$ and L is the code length. The resubstituted classification error $R(N)$ of our classifier satisfies

$$|R(N) - R_{\min}| \le \mathcal{O}\left(K^{\frac{1}{2}}\left(\frac{\log N}{N}\right)^{\frac{1}{K+2}}\right)$$

for all sufficiently large N, where $R_{\min}$ is the Bayes risk.

References

[1] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees.* Belmont, California: Wadsworth Inc., 1984.

[2] J. Quinlan and R. L. Rivest, "Inferring Decision Trees Using the Minimum Description Length Principle," *Inf. and Comp.*, vol. 80(3), pp. 227–248, 1989.

[3] J. Rissanen, "Universal Coding, Information, Prediction, and Estimation," *IEEE Trans. Inform. Theory*, vol. **IT-30**, pp. 629–636, July 1984.

[4] Rissanen J. and Yu B.: "MDL Learning", Progress in Automation and Information Systems, Springer-Verlag, New York, 1991.

[5] S. Itoh, "A Piecewise Linear Approximation of Planar Curves by MDL Modeling," in *Proceedings of 1993 IEEE Information Theory Workshop*, pp. 34-35, June 1993.

An Extension on Learning Bayesian Belief Networks Based on MDL Principle

Joe Suzuki
Dept of Mathematics, Osaka University,
Toyonaka, Osaka 560, Japan

Abstract — **Bayesian belief network (BBN) is a framework for representation/inference of some knowledge with uncertainty [1]. Since the process of constructing a BBN manually by experts is time-consuming in general, some method supporting the task is needed. We proposed an algorithm for acquiring some BBN automatically from finite examples based on minimum description length (MDL) principle [2]. This paper addresses an improvement which relaxes a constraint that the original scheme held on the representation.**

In BBNs, attributes and stochastic dependences between them are expressed as nodes and directed links connecting them, respectively, where each attribute may be a predicate, a numerical data, etc., and each dependence is numerically expressed as the conditional probability of one attribute given other attributes if their dependence exists. Therefore, in general, BBNs are represented in terms of the network structure and the conditional probabilities.

Suppose that we have N possible attributes $j = 1, 2, \cdots, N$ ($N \geq 1$), where each attribute value ranges over $A[j] = \{0, 1, \cdots, \alpha[j] - 1\}$ ($2 \leq \alpha[j] < \infty$), and also that we induce the network structure $g \in G$ of a BBN from n examples $x_1^n = x_1 x_2 \cdots x_n$, where $x_i = (x_{i,1}, x_{i,2}, \cdots, x_{i,N})$, $x_{i,j} \in A[j]$, $j = 1, 2, \cdots, N$, $i = 1, 2, \cdots, n$, and a set of the possible network structures, G, is prepared. The problem is to determine the set $\pi[j, g] \subseteq \{1, 2, \cdots, j-1\}$ ($\pi[1, g] = \phi$, $g \in G$) of attributes which each attribute $j = 1, 2, \cdots, N$ depends on, provide that the N attributes have been arranged in such an order that the directed dependence is valid [1]. Then, the number of the possible network structures is $|G| = \prod_{j=1}^{N} 2^{j-1} = 2^{N(N-1)/2}$.

If we applied MDL principle to this problem, a possible description length $L^A(g, x_1^n)$ based on network structure $g \in G$ would be [2]

$$\mathcal{H}^A(x_1^n|g) + \frac{k^A(g)}{2} \log \frac{n}{2\pi} + \sum_{j=1}^{N} |S(j,g)| \log \frac{[\Gamma(1/2)]^{\alpha[j]}}{\Gamma[\alpha[j]/2]} + l_G(g)$$

except some constant terms, where $S[j, g] = \prod_{k \in \pi[j,g]} A[k]$, $\mathcal{H}^A(x_1^n|g)$ and $k^A(g)$ are respectively the empirical entropy and the number of the conditional probabilities to be fixed, and $l_G(g)$ is the description length of model $g \in G$. In the original scheme [2], the network structure $g \in G$ that minimizes $L^A(g, x_1^n)$ is selected from n examples so that the best compromise between the complexity of the network and the fitness of the n examples to the network is achieved in terms of the description length.

The description length $L^A(g, x_1^n)$ is optimal in the sense that the redundancy $E_\theta[L^A(g, x_1^n) + \log p(x_1^n|\theta)]$ is asymptotically upperbounded by the optimal minimax redundancy except the length of $l_G(g)$ for $g \in G$ when the source θ generating the data $x_1^n \in (\prod_{j=1}^{N} A[j])^n$ is expressed as one of those BBNs, where $p(x_1^n|\theta)$ is the probability of $x_1^n \in (\prod_{j=1}^{N} A[j])^n$ given source θ. However, we should note that in some cases where some $\alpha[j]$ is large or $A[j]$ takes continuous values for an attribute, $j = 1, 2, \cdots, N$, the j-th node is not connected to any other nodes even when the dependence is actually significant. So, we propose such an extended scheme that the alphabet $A[j]$ is clustered into another alphabet $B[j, g] = \{0, 1, \cdots, \beta[j, g] - 1\}$ ($1 \leq \beta[j, g] \leq \alpha[j]$), where $y \cap y' = \phi$ for any $y \neq y' \in B[j, g]$ and $\cup_{y \in B[j,g]} y = A[j]$, and we implement a similar procedure for such a new alphabet $B[j, g]$, for $j = 1, 2, \cdots, N$ and $g \in G$. In most cases, such a clustering procedure is manually done as a pre-process for both learning and inferece processes.

Therefore, the structure $g \in G$ refers to the clustering structure $B[j, g]$ as well as the network structure $\pi[j, g]$, for $j = 1, 2, \cdots, N$, in the proposed scheme. The counterpart $L^B(g, x_1^n)$ of $L^A(g, x_1^n)$ is

$$\mathcal{H}^B(x_1^n|g) + \frac{k^B(g)}{2} \log \frac{n}{2\pi} + \sum_{j=1}^{N} |T(j,g)| \log \frac{[\Gamma(1/2)]^{\beta[j,g]}}{\Gamma[\beta[j,g]/2]} + l_G(g), \tag{1}$$

where

$$\mathcal{H}^B(x_1^n|g) = \sum_{j=1}^{N} \sum_{t \in T[j,g]} \sum_{y \in B[j,g]} m[y,t,j] \log\{|y| \frac{m[t,j] + \beta[j,g]/2}{m[y,t,j] + 1/2}\}$$

and $k^B(g) = \sum_{j=1}^{N} (\beta[j, g] - 1) \prod_{k \in \pi[j,g]} \beta[k, g]$, $T[j, g] = \prod_{k \in \pi[j,g]} B[k, g]$, and $m[t, j]$ and $m[y, t, j]$ denote the occurrence, in $y_1^n \in (\prod_{j=1}^{N} B[j, g])^n$, of $t \in \prod_{k \in \pi[j,g]} B[k, g]$ and that of $y \in B[j, g]$ given $t \in \prod_{k \in \pi[j,g]} B[k, g]$, respectively, for $j = 1, 2, \cdots, N$ and $g \in G$. Note that the same length $-\log\{|y|(m[t, j] + \beta[j, g]/2)/(m[y, t, j] + 1/2)\}$ is assigned to the $|y|$ symbols in a group $y \in B[j, g]$, assuming that they occur equiprobably.

Theorem 1: The redundancy $E_\theta[L^B(g, x_1^n) + \log p(x_1^n|\theta)]$ is asymptotically upperbounded by the optimal minimax redundancy except the length of $l_G(g)$ for $g \in G$ when the source θ ranges over the BBNs in which the elements in $A[j]$ can be clustered into any exclusive groups for $j = 1, 2, \cdots, N$, and the elements in the same group occurs equiprobably.

Theorem 2: The number of the possible structures in the proposed scheme is $|G| = 2^{N(N-1)/2} \prod_{j=1}^{N} f(\alpha[j])$, where

$$f(\alpha) = \sum_{i=1}^{\alpha} \sum_{j=1}^{i} \frac{j^\alpha (-1)^{i-j}}{(i-j)!\, j!}.$$

References

[1] J. Peral, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, CA (1988).

[2] J. Suzuki. "Learning Bayeesian Belief Networks Based on the Minimum Description Length Principle", submitted to *IEEE Trans. on Information Theory* (1993).

Optimal Universal Learning and Prediction of Probabilistic Concepts

Meir Feder, Yoav Freund and Yishay Mansour

Department of Electrical Engineering - Systems, Tel Aviv University; AT& T Bell Labs, Room 2B-428, Murray Hill NJ; Department of Computer Science, School of Mathematics, Tel Aviv University.

I Introduction

We consider the following setting of the (supervised) learning problem. A sequence of input data $x_1, \ldots, x_t, \ldots$, is given, one by one, and the goal is to predict the corresponding outputs $y_1, \ldots, y_t, \ldots$. It is assumed that at time t the predictor has an access to the previous input-output pairs $(x_i, y_i)_{i=0}^{t-1}$, and then it has to predict y_t given the new input x_t. The input and output are connected by some unknown functional relation, given by a conditional probability distribution $p_\theta(y|x)$, where it is only known that θ belong to some general index set Θ.

The prediction outcome is an estimated probability distribution $q_t(\cdot) = q(\cdot|x_t; x_1^{t-1}, y_1^{t-1})$ for the unobserved value y_t. (The notation x_i^j means $x_i, \ldots, x_j$). When y_t is revealed, a prediction "log-loss" $-\log q_t(y_t)$ is incurred. The goal is to minimize the expected accumulated log-loss, for the entire sequence of decisions, $E_\theta\left\{\sum_{t=1}^n -\log q_t(y_t)\right\}$, where the expectation is with respect to the "true" distribution $P_\theta(y_1^n|x_1^n)$.

Since θ is unknown, we wish to find a *universal* predictor, independent of θ. We note that the simpler problem of predicting y_t with log-loss, in the absence of the input x_1^n, is completely equivalent to universal coding. As is became evident, from recent results in universal coding, the optimal prediction is based on a Bayesian approach in which a "mixture" probability measure $Q(y_1^n) = \int_{\theta\in\Theta} w(d\theta) P_\theta(y_1^n)$ is assigned to the observation sequence. The "prior" $w(d\theta)$ is chosen to attain

$$\sup_w \int_\theta w(d\theta) \sum_{y_1^n} P_\theta(y_1^n) \log \frac{P_\theta(y_1^n)}{\int_{\theta'} w(d\theta') P_{\theta'}(y_1^n)}, \qquad (1)$$

i.e., to achieve the capacity of the "channel" between Θ and and Y_1^n. The prediction at each time point is given by $q_t(y_t|y_1^{t-1}) = Q(y_1^t)/Q(y_1^{t-1})$. A classical result in universal coding [1] states that this encoder attains the min-max redundancy, implying that the associated predictor minimizes the maximal extra accumulated log-loss. Recently, [2] it was shown that the performance of this predictor, given by the capacity (1), is a lower bound on the performance of any universal coder, in the sense that any other encoder cannot have a smaller redundancy (or a smaller excess log-loss) for "most" $\theta \in \Theta$.

Our proposed solution for the supervised learning problem is likewise Bayesian, and the contribution of this work lies in determining the optimal way to choose the Bayesian "prior" for the superviesd learning problem, and observing the strong sequential, non-anticipating, structure of the resulting universal predictor.

II Optimal Universal Learning

In our problem, where a side information x_1^n is given, one may try to generalize the universal coding results in the following way. Since x_1^n is known (or will be known as we predict the output), an optimal predictor may be chosen for each x_1^n. This predictor will be based on a Bayesian universal probability, in which the weights depend on x_1^n, and each such probability will attain a capacity $C(x_1^n)$ of the channel between Θ and Y_1^n, given that x_1^n. However, this solution turns out to be unacceptable because it leads to too pessimistic, or "too careful" prediction procedures. This is because we try to minimize the extra loss, for the worst θ, and that for each x_1^n. In addition the resulting $w(d\theta)$ depends on the entire x_1^n, and so the Bayesian mixture probability does not factor into a sequential assignment.

We overcome these drawbacks by postulating a probability distribution, $\mu(x_1^n)$, over the input. This is a common assumption in many learning problems, where at least in the training stage, the input is indeed randomly chosen according to some pre-defined distribution. The extra average accumulated log-loss, is now $R(\theta, Q) = -\sum_{x_1^n} \mu(x_1^n) \sum_{y_1^n} P_\theta(y_1^n|x_1^n) \log \frac{P_\theta(y_1^n)}{Q(y_1^n)}$. Here, also, the solution to $\min_Q \max_\theta R(\theta, Q)$, is the max-min solution which (almost by definition) is given by the Bayesian mixture $Q(y_1, \ldots, y_n|x_1, \ldots, x_n) = \int_{\theta\in\Theta} w(d\theta) P_\theta(y_1, \ldots, y_n|x_1, \ldots, x_n)$ where $w(d\theta)$ is a weight over Θ, independent of x_1^n, which maximizes the conditional mutual information,

$$I(\Theta, Y_1^n|X_1^n) = \int_\theta w(d\theta) \sum_{x_1^n} \mu(x_1^n) \sum_{y_1^n} P_\theta(y_1^n|x_1^n) \log \frac{P_\theta(y_1^n|x_1^n)}{Q(y_1^n|x_1^n)}. \qquad (2)$$

The quantity $\sup_w I(\Theta, Y_1^n|X_1^n) \triangleq C_n$ can be interpreted as the capacity of an "auxiliary channel" between Θ and Y_1^n, with side information X_1^n. Similarly to [2] we prove that C_n, which is the loss incurred by our Bayesian predictor, cannot be improved by any other predictor, for "most" $\theta \in \Theta$.

For prediction we need a sequential probablity assignment. First, we observe that the universal probablity above can always be factored as $Q(y_1, \ldots, y_n|x_1, \ldots, x_n) = \prod_{t=1}^n q(y_t|y_1^{t-1}, x_1^n)$. Furthermore, under the common assumption in learning theory $P_\theta(y_1^n|x_1^n) = \prod_{t=1}^n p_\theta(y_t|x_t)$. In this case the universal probability can actually expressed as $Q(y_1, \ldots, y_n|x_1, \ldots, x_n) = \prod_{t=1}^n q(y_t|y_1^{t-1}, x_1^t)$ and so it provides a fully sequential, non-anticipating prediction procedure.

Finally, since C_n is an attainable lower bound on the performance of any learning and prediction algorithm, we make the following claim: *A class of conditional distributions $\{p_\theta(y|x), \theta \in \Theta\}$ is learnable if and only if $C_n/n \to 0$, as the data length $n \to \infty$.* Thus, C_n can replace other measures, such as these of Vapnik and Chervonenkis, to determine the complexity of a class of models.

References

[1] R. G. Gallager, "Source Coding with Side Information and Universal Coding," unpublished manuscript, Sept. 1976.

[2] N. Merhav and M. Feder, "A Strong Version of the Redundancy-Capacity Theorem for Universal coding," *IEEE Trans. Inform. Theory*, May 1995.

Covering Radius 1985-1994

G. D. Cohen, S. Litsyn, A. Lobstein, H. F. Mattson, Jr.

G. D. Cohen and A. Lobstein are with Centre National de la Recherche Scientifique and Télécom Paris, Département INF, 46 rue Barrault 75634 Paris Cedex 13, France

S. Litsyn is with Dept. of Electrical Engineering-Systems Tel-Aviv University, Ramat-Aviv, 69978, Israel

H. F. Mattson, Jr. is with School of Computer and Information Science 4-116 Center for Science & Technology, Syracuse, New York 13244-4100, USA

Abstract — **We survey important developments in the theory of covering radius during the period 1985-1994. We present lower bounds, constructions and upper bounds, the linear and nonlinear cases, density and asymptotic results, normality, specific classes of codes, covering radius and dual distance, tables, and open problems.**

I. Background

Interest in covering radius has grown markedly since about 1980. The topic has applications to problems of data compression, testing, and write-once memories. It is also interesting for its own sake. It is a fundamental geometric parameter of a code, characterizing its maximal error correcting capability in the case of minimum distance decoding. Although some of these applications are recent, others are old. Yet after the 1960 paper of Gorenstein, Peterson, and Zierler [4] showing that the double-error-correcting binary BCH code has covering radius 3, (though there were some papers on the football-pool problem), there was nothing on covering radius until the seminal paper [3] of Delsarte in 1973.

An earlier survey, [1], published in 1985, has seemingly contributed to the increase in the number of papers on this topic in the last decade. Covering radius has evolved into a subject in its own right, and we feel the need to give a summary of many works on covering codes that have appeared since [1].

II. Plan of the paper

We discuss lower bounds in Section 2, mentioning several methods but especially linear programming and the method of excess. These methods usually improve on the sphere-covering bound.

In Section 3 we discuss asymptotic density of coverings when the length goes to infinity while the radius remains fixed.

In Section 4 we treat upper bounds for linear codes, focusing on the deficiency of a code, "worst" codes (useful in designing write-once memories), and Griesmer, optimum, and maximum codes.

Section 5 discusses upper bounds obtained from constructions. There are blockwise direct sums, amalgamated direct sums, variants on the $u|u+v$ construction, and simulated annealing. This section closes with codes over mixed alphabets.

In Section 6 we discuss normality and some of its many offshoots, closing with the conjecture $K(n+2,t+1) \leq K(n,t)$.

Section 7 deals with specific classes of error correcting codes, among which are Reed-Muller, BCH and their duals, cyclic, self-dual, and algebraic-geometric codes.

Section 8 is a brief account of relations between covering radius and dual distance.

Section 9, on generalizations of coverings, treats mixed, weighted, and multiple coverings.

In Section 10 we discuss the open problems of [1], add two new ones, and disprove a conjecture.

We provide extensive tables of bounds for coverings.

In our bibliography of some 270 items we have tried to include all papers bearing on the covering radius of block codes.

References

[1] G. D. Cohen, M. G. Karpovsky, H. F. Mattson, Jr., and J. R. Schatz, "Covering radius—survey and recent results," *IEEE Trans. Inform. Theory*, vol. 31, pp. 328–343, 1985.

[2] G. D. Cohen, S. Litsyn, A. Lobstein, and H. F. Mattson, Jr., "Covering radius 1985–1994," Technical Report 94-D-025, Ecole Nationale Supérieure des Télécommunications, Paris, 76 p., 1994. Submitted to *AAECC Journal.*

[3] P. Delsarte, "Four fundamental parameters of a code and their combinatorial significance," *Information and Control*, vol. 23, pp. 407–438, 1973.

[4] D. Gorenstein, W.W. Peterson, and N. Zierler, "Two-error correcting Bose-Chaudhury codes are quasi-perfect," *Information and Control*, vol. 3, pp. 291–294, 1960.

Greedy Generation of Non-Binary Codes

by Laura Monroe and Vera Pless

Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago
Chicago, IL 60607

We get a B-ordering of all binary n-tuples $\mathbf{V}_n$ by choosing an ordered basis $\{y_1,...,y_n\}$ of V_n and ordering the n-tuples as follows: 0, y_1, y_2, y_2+y_1, y_3, y_3+y_1, y_3+y_2, $y_3+y_2+y_1$, y_4,... Given a minimum distance d, choose a set of vectors $\mathbf{S}$ with the zero vector first, then go through the vectors in their B-ordering and choose the next vector which has distance d or more from all vectors already chosen. The surprising result that $\mathbf{S}$ is linear has been shown in several different ways [1, 2, 3, 4, 6]. Linear codes found in this fashion are called greedy codes.

An ordered basis $\{y_i\}$ of $\mathbf{V}_n$ is called triangular [1] if $y_i = (0,..., 0, 1, *,..., *)$, with the 1 in the ith position. When the y_i are unit vectors, the order is the lexicographic order. The columns $\mathbf{h}_n$, $\mathbf{h}_{n-1}$,...$\mathbf{h}_1$ of the g-parity check matrix $\mathbf{H}_n$ are constructed one by one. We associate numbers with their binary representations. We let $\mathbf{h}_1$ be the number 1. Let $y_{i+1} = (0,..., 0, 1, \varepsilon_i,..., \varepsilon_1)$, where the ε_i are 0 or 1. If $\mathbf{H}_i = [\mathbf{h}_i,...\mathbf{h}_1]$ is known, we let β be the smallest number so that $\mathbf{h}_{i+1} = \beta + (\varepsilon_i\mathbf{h}_i + ...+\varepsilon_1\mathbf{h}_1)$ is not a sum of d-1 or fewer columns of $\mathbf{H}_i$. Then $\mathbf{H}_{i+1} = [\mathbf{h}_{i+1},...\mathbf{h}_1]$. Each $\mathbf{H}_i$ is a parity check matrix of the greedy code chosen using the ordered basis $\{y_1,...,y_i\}$ [1]. Further, the syndrome of any vector with regard to $\mathbf{H}_i$ is the g-value which is assigned to it by generalizing the greedy algorithm for choosing vectors in the code, hence the name g-parity check matrix.

The non-binary case has also aroused quite a bit of interest. One may generalize the concept of B-ordering to the case of an arbitrary base field. For example, in the case $\mathbf{GF}(4) = \{0, 1, \omega, \underline{\omega}\}$, the B-ordering is generated by choosing an ordered basis $\{y_1,...,y_n\}$ of V_n and ordering the n-tuples as follows: 0, y_1, ωy_1, $\underline{\omega} y_1$, y_2, y_2+y_1, $y_2+\omega y_1$, $y_2+\underline{\omega} y_1$, ωy_2, ωy_2+y_1, $\omega y_2+\omega y_1$, $\omega y_2+\underline{\omega} y_1$, $\underline{\omega} y_2$, $\underline{\omega} y_2+y_1$, $\underline{\omega} y_2+\omega y_1$ $\underline{\omega} y y_2 \underline{\omega} y_1$,... The greedy code is then generated from the B-ordering as in the binary case. It has been shown by Conway and Sloane in the case of lexicodes [2], and independently by Fon-Der-Flaass [3], and Van Zanten [6] in the case of general greedy codes that those codes for which the base field is of order 2^{2^r} is linear. When the base field is not of order 2^{2^r}, the situation is a little less clear. In general, the greedy codes generated in this case have been linear only for small n. In every case examined, linearity breaks down at some point early in the generation of the code. It is possible, however, to extend the parity check matrix generating algorithm to this case. Although this algorithm does not produce the greedy code itself, it still produces a very good code which is generated in a greedy-like fashion.

The parity check matrix is generated in the same way as in the binary case. This algorithm also assumes that the ordered basis $\{y_i\}$ of $\mathbf{V}_n$ being used is triangular, and that the first non-zero entry in each basis vector is 1. Then if $\mathbf{H}_i = [\mathbf{h}_i,...\mathbf{h}_1]$ is known, we let β be the smallest number so that $\mathbf{h}_{i+1} = \beta + (\varepsilon_i\mathbf{h}_i + ...+\varepsilon_1\mathbf{h}_1)$ is not a linear combination of d-1 or fewer columns of $\mathbf{H}_i$. Then $\mathbf{H}_{i+1} = [\mathbf{h}_{i+1},...\mathbf{h}_1]$.

Many interesting codes are generated via the parity check algorithm. We have generated many such parity check matrices via the computer for base fields of orders 3, 4, and 5. In all examined cases, the codes generated have had dimension within 1 of the best known codes, for a given n and d, and most of the codes generated had dimension equal to that of the best known codes. Better yet, we have generated more than 100 record breaking codes over the base field of order 4 [5]. Most of these are shortened codes of larger greedy codes. The following table lists the parameters of the codes from which the shortened codes are derived.

Table. Parameters of record breaking codes over **GF**(4) obtained via the parity check matrix algorithm.

n	k	d
52	44	5
128	118	5
35	26	6
71	60	6

This work was supported in part by NSA grant MDA 904-91-H-0003.

REFERENCES

1. R. A. Brualdi and V. Pless, *Greedy Codes*, JCT(A) 64 (1993), 10-30.

2. J. H. Conway and N. J. A. Sloane, *Lexicographic Codes: Error Correcting Codes from Game Theory*, IEEE Trans. Inform. Theory IT-32 (1986), 337-348.

3. D. Fon-Der-Flaass, *A Note on Greedy Codes,* to appear.

4. L. Monroe, *Binary Greedy Codes*, to appear in Congressus Numerantium, vol.100-104.

5. N. J. A. Sloane, *Table of Lower Bounds on dmax (n, k) for Linear Codes over Fields of Order 4*, The Handbook of Coding Theory, edited by R. A. Brualdi, C. Huffman, and V. Pless, published by Elsevier Science Publishers, to appear.

6. A. J. van Zanten, *Lexicodes Over Fields of Characteristic 2,* to appear.

Lee Distance Gray Codes

Bella Bose[1] and Bob Broeg
Department of Computer Science, Oregon State University Corvallis, OR 97331-3902
Phone: 503-737-5573, Fax: 503-737-3014, E-mail: {bose, broegb}@cs.orst.edu

Abstract — **This papers presents 4 methods of generating a Lee distance Gray code. The first two methods presented are for radix k numbers, and the other two methods are for mixed radix numbers.**

I. Introduction

Some recently developed parallel machines have a multi-dimensional torus topology for their processor interconnection structure. Many algorithms can be solved efficiently by embedding a Hamiltonian cycle or a Hamiltonian path within this topology. This correspondence addresses the embedding problem by presenting four methods of constructing a Lee distance Gray code. For each method, Let $\mathbf{R} = (r_{n-1}r_{n-2}\cdots r_0)$ be a number in radix notation, and let $\mathbf{G} = (g_{n-1}g_{n-2}\cdots g_0)$ be the Gray code representation given by f_i, i.e., $\mathbf{G} = f_i(\mathbf{R})$.

II. Single Radix Codes

First, assume there are n dimensions, each having the same number of processors, k, where $k \geq 3$. Each processor node is labeled with a distinct n-digit, radix k vector $(r_{n-1}r_{n-2}\cdots r_0)$, where $r_i \leq k$ for $0 \leq i \leq n-1$. Two nodes, $\mathbf{A} = (a_{n-1}a_{n-2}\cdots a_0)$ and $\mathbf{B} = (b_{n-1}b_{n-2}\cdots b_0)$, are adjacent if the Lee distance between them, $D_L(\mathbf{A},\mathbf{B})$, is one. Lee distance is defined as

$$D_L(\mathbf{A},\mathbf{B}) = \sum_{i=0}^{n-1} \min(a_i - b_i, b_i - a_i).$$

Two methods are given below for constructing a Gray code base on the assumption of the previous paragraph.

Method 1:

$$f_1(r_{n-1}r_{n-2}\cdots r_0) = r_{n-1}(r_{n-2} - r_{n-1})\cdots(r_0 - r_1)$$

Method 2: This method produces a Hamiltonian cycle if k is even, and a Hamiltonian path if k is odd. Let $\overline{r_i} = k-1-r_i$, and let $g_{n-1} = r_{n-1}$. Then, for $i = n-2,\ldots,0$,
if k is even then

$$g_i = \begin{cases} r_i, & \text{if } r_{i+1} \text{ is even} \\ \overline{r_i}, & \text{otherwise} \end{cases}$$

or, if k is odd, let $r' = \sum_{j=i+1}^{n-1} r_j$, and

$$g_i = \begin{cases} r_i, & \text{if } r' \text{ is even} \\ \overline{r_i}, & \text{otherwise} \end{cases}$$

[1]This work is supported in part by the National Science Foundation under Grant MIP-9404924.

III. Mixed Radix Codes

In many cases, however, the number of processors per dimension varies. Let $\mathbf{K} = k_{n-1}k_{n-2}\cdots k_0$ be an n-dimensional vector where k_i is the radix of dimension i and $k_i \geq 3$ for $0 \leq i \leq n-1$. In this case, Method 3 gives a Gray code design resulting in a Hamiltonian cycle if k_i is even for at least one value of i. If each k_i is odd, the resulting Gray code produces a Hamiltonian path. Method 4 produces a Hamiltonian cycle if all k_i are odd.

Let each processor node be labeled with a distinct n-digit vector $\mathbf{R} = (r_{n-1}r_{n-2}\cdots r_0)$, where $0 \leq r_i \leq k_i - 1$ for $i = 0,1,\ldots,n-1$. Vector $\mathbf{R}$ is said to be in *mixed-radix notation*, and the integer value of $\mathbf{R}$ is given by

$$\begin{aligned} I(\mathbf{R}) &= r_0 + r_1k_0 + r_2k_0k_1 + \cdots + r_{n-1}k_0k_1\ldots k_{n-2} \\ &= \sum_{i=1}^{n-1}\left(r_i\prod_{j=0}^{i-1}k_j\right) + r_0 \end{aligned}$$

In mixed-radix notation, the Lee distance, $D_L(\mathbf{A},\mathbf{B})$, between $\mathbf{A} = (a_{n-1}a_{n-2}\cdots a_0)$ and $\mathbf{B} = (b_{n-1}b_{n-2}\cdots b_0)$ is defined as

$$D_L(\mathbf{A},\mathbf{B}) = \sum_{i=0}^{n-1} \min((a_i - b_i) \bmod k_i, (b_i - a_i) \bmod k_i).$$

Method 3: Assume that at least one of the k_i's is even. Without loss of generality, assume that the dimensions are ordered so that if k_i is even and k_j is odd, then $i > j$. Let ℓ be the index of the lowest even dimension. That is, the dimensions are ordered as follows.

$$\overbrace{k_{n-1}\cdots k_\ell}^{even}\ \overbrace{k_{\ell-1}\cdots k_0}^{odd}$$

Now, letting $\overline{r_i} = k_i - 1 - r_i$ and $r'_i = \sum_{j=i+1}^{\ell} r_j$, f_3 is defined as follows.

$$\begin{aligned} & g_{n-1} = r_{n-1}, \text{ and} \\ \text{for } i = n-2 \text{ downto } \ell: \quad & g_i = \begin{cases} r_i, & \text{if } r_{i+1} \text{ is even} \\ \overline{r_i}, & \text{otherwise} \end{cases} \\ \text{for } i = \ell-1 \text{ downto } 0: \quad & g_i = \begin{cases} r_i, & \text{if } r'_i \text{ is even} \\ \overline{r_i}, & \text{otherwise} \end{cases} \end{aligned}$$

Method 4: Assume that k_i is odd for $0 \leq i \leq n-1$, and that the dimensions are ordered such that $k_{n-1} \geq k_{n-2} \geq \ldots \geq k_0$. Also, define

$$\overline{r_i} = \begin{cases} r_i & , \text{if } r_{i+1} \text{ is odd} \\ k_i - 1 - r_i & , \text{otherwise} \end{cases}$$

Now, f_4, which produces a Gray code yielding a Hamiltonian cycle, is defined as follows.

$$\begin{aligned} g_{n-1} &= r_{n-1}, \text{ and for } 0 \leq i \leq n-2 \\ g_i &= \begin{cases} (r_i - r_{i+1}) \bmod k_i & , \text{if } r_{i+1} < k_i \\ \overline{r_i} & , \text{otherwise} \end{cases} \end{aligned}$$

Diffuse Difference Triangle Sets

Torleiv Kløve*, Yuri V. Svirid**

*Department of Informatics, University of Bergen, HIB, N–5020 Bergen, Norway

**Chair for Communications, TU Munich, D-80290 Munich, Germany & Dept. for RTS, BGUIR, 220027 Minsk, Belarus

Abstract — **The properties of introduced diffuse difference triangle sets (DTS) are considered.**

I. Definitions

An (I, J)-set is a set $\Sigma = \{\Sigma_1, \Sigma_2, \ldots, \Sigma_I\}$ where $\Sigma_i = \{\sigma_{ij} \mid 0 \le j \le J\}$ for $1 \le i \le I$ and the elements are integers such that $\sigma_{i0} = 0$ for $1 \le i \le I$, $\sigma_{ij} \le \sigma_{i,j+1}$ for $1 \le i \le I$ and $0 \le j < J$. Let $m(\Sigma) = \max\{\sigma_{iJ} \mid 1 \le i \le I\}$ and $\mu(\Sigma) = \sum_{i=1}^{I} \sigma_{iJ}$. An (I, J)-DTS (in normalized form) is an (I, J)-set Σ such that all the differences $\sigma_{ij} - \sigma_{ij'}$ with $1 \le i \le I$ and $0 \le j' < j \le J$ are distinct.

A *diffuse* DTS satisfies some additional conditions:

$$\sigma_{i,j+1} \ge \sigma_{ij} + \delta \text{ for } 1 \le i \le I \text{ and } 0 \le j < J, \quad (1)$$

$$|\sigma_{ij} - \sigma_{i'j'}| \ge \delta_c \text{ for } 1 \le i \ne i' \le I \text{ and } 0 \le j, j' \le J \quad (2)$$

except when $j = j' = 0$.

An (I, J)-DTS satisfying (1) and (2) is called an (I, J, δ, δ_c)-DTS. The set of (I, J, δ, δ_c)-DTS is denoted by $\mathcal{S}(I, J, \delta, \delta_c)$. We note that for $I = 1$ the condition (2) is empty, we will put $\delta_c = 0$ in this case. For applications, which are mainly found in the constructions of diffuse codes [1], we want (I, J, δ, δ_c)-DTS Σ with $m(\Sigma)$ as small as possible. Let $m(I, J, \delta, \delta_c) = \min\{m(\Sigma) \mid \Sigma \in \mathcal{S}(I, J, \delta, \delta_c)\}$. If $m(\Sigma) = m(I, J, \delta, \delta_c)$, then Σ is called optimal. Similarly, define $\mu(I, J, \delta, \delta_c) = \min\{\mu(\Sigma) \mid \Sigma \in \mathcal{S}(I, J, \delta, \delta_c)\}$.

We will study here the structure of the set $\mathcal{S}(I, J, \delta, \delta_c)$ when one or both of δ, δ_c are increasing and the other parameters are kept fixed.

II. Increasing δ

Let $\mathcal{G}(I, J, \delta_c)$ denote the set of (I, J)-sets $\Gamma = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_I\}$ where the $\Gamma_i = \{\gamma_{ij} \mid 0 \le j \le J\}$ are such that for each fixed l, where $0 < l \le J$, all the differences $\gamma_{i,l+j} - \gamma_{ij}$ with $1 \le i \le I$ and $0 \le j \le J - l$ are distinct, and $|\gamma_{ij} - \gamma_{i'j}| \ge \delta_c$ for $1 \le i \ne i' \le I$ and $1 \le j \le J$.

Let $g(I, J, \delta_c) = \min\{m(\Gamma) \mid \Gamma \in \mathcal{G}(I, J, \delta_c)\}$. For $\Gamma \in \mathcal{G}(I, J, \delta_c)$ and $\delta > 0$, define $\Sigma = f_\delta(\Gamma)$ by $\sigma_{ij} = \gamma_{ij} + j\delta$ for $1 \le i \le I$ and $0 \le j \le J$. We note that $m(f_\delta(\Gamma)) = m(\Gamma) + J\delta$.

Lemma 1 *If $\Sigma \in \mathcal{S}(I, J, \delta, \delta_c)$, then $\Sigma = f_\delta(\Gamma)$ for some $\Gamma \in \mathcal{G}(I, J, \delta_c)$.*

Lemma 2 *For each $\Gamma \in \mathcal{G}(I, J, \delta_c)$ there exists a bound $\delta_0(\Gamma)$ such that $f_\delta(\Gamma) \in \mathcal{S}(I, J, \delta, \delta_c)$ for $\delta \ge \delta_0(\Gamma)$.*

Based on Lemmata 1 and 2, we give the following

Theorem 1 *For given I, J, δ_c, and $\zeta \ge 0$, there exists a bound $\delta_0(I, J, \delta_c, \zeta)$ such that*
a) $m(I, J, \delta, \delta_c) = g(I, J, \delta_c) + J\delta$ for $\delta \ge \delta_0(I, J, \delta_c, 0)$,
b) for $\delta \ge \delta_0(I, J, \delta_c, \zeta)$ we have

$$\{\Sigma \in \mathcal{S}(I, J, \delta, \delta_c) \mid m(\Sigma) = m(I, J, \delta, \delta_c) + \zeta\}$$
$$= \{f_\delta(\Gamma) \mid \Gamma \in \mathcal{G}(I, J, \delta_c) \text{ and } m(\Gamma) = g(I, J, \delta_c) + \zeta\}.$$

Corollary 1 *For $\delta \ge \delta_0(I, J, \delta_c, \zeta)$, the size of $\{\Sigma \in \mathcal{S}(I, J, \delta, \delta_c) \mid m(\Sigma) = m(I, J, \delta, \delta_c) + \zeta\}$ is independent of δ.*

III. Increasing δ_c

We can assume without loss of generality that for $\Sigma \in \mathcal{S}(I, J, \delta, \delta_c)$ we have $\sigma_{iJ} < \sigma_{i+1,J}$ for $1 \le i < I$. With this assumption, we partition the set $\mathcal{S}(I, J, \delta, \delta_c)$ into two sets: $\mathcal{S}_1(I, J, \delta, \delta_c) = \{\Sigma \in \mathcal{S}(I, J, \delta, \delta_c) \mid \sigma_{iJ} \le \sigma_{i+1,1} \text{ for } 1 \le i < I\}$, $\mathcal{S}_2(I, J, \delta, \delta_c) = \mathcal{S}(I, J, \delta, \delta_c) \setminus \mathcal{S}_1(I, J, \delta, \delta_c)$.

For $\Theta \in \mathcal{S}(I, J-1, \delta, 0)$ and non-negative integers $\delta_1, \delta_2, \ldots, \delta_I$, define $\Sigma = h(\Theta, \delta_1, \delta_2, \ldots, \delta_I)$ by $\sigma_{i0} = 0$ for $1 \le i \le I$, $\sigma_{ij} = \sum_{l=1}^{i} \delta_l + \sum_{l=1}^{i-1} \theta_{l,J-1} + \theta_{i,j-1}$ for $1 \le i \le I$ and $1 \le j \le J$. We note that $m(\Sigma) = \sum_{l=1}^{I} \delta_l + \mu(\Theta)$.

Lemma 3 *If $\Sigma \in \mathcal{S}_1(I, J, \delta, \delta_c)$, then there exist a $\Theta \in \mathcal{S}(I, J-1, \delta, 0)$ and non-negative integers $\delta_i \ge \delta_c$ for $1 \le i \le I$ such that $\Sigma = h(\Theta, \delta_1, \delta_2, \ldots, \delta_I)$. In particular, $m(\Sigma) \ge \mu(I, J-1, \delta, 0) + I\delta_c$.*

Lemma 4 *For each $\Theta \in \mathcal{S}(I, J-1, \delta, 0)$ there exists a bound $\delta_{c0}(\Theta)$ such that if $\delta_i \ge \delta_c \ge \delta_{c0}(\Theta)$ for $1 \le i \le I$, then $h(\Theta, \delta_1, \delta_2, \ldots, \delta_I) \in \mathcal{S}_1(I, J, \delta, \delta_c)$.*

Lemma 5 *If $\Sigma \in \mathcal{S}_2(I, J, \delta, \delta_c)$, then $m(\Sigma) \ge (I+1)\delta_c$.*

Based on Lemmata 3, 4, and 5, we obtain

Theorem 2 *For given I, J, and δ there exists a bound $\delta_{c0}(I, J, \delta)$ such that if $\delta_c \ge \delta_{c0}(I, J, \delta)$, then $m(I, J, \delta, \delta_c) = \mu(I, J-1, \delta, 0) + I\delta_c$.*

Corollary 2 *For $\delta_c > \delta_{c0}(I, J, \delta, \zeta)$, the size of $\{\Sigma \in \mathcal{S}(I, J, \delta, \delta_c) \mid m(\Sigma) = m(I, J, \delta, \delta_c) + \zeta\}$ is independent of δ_c.*

IV. Both δ and δ_c increasing

Combining Lemmas 2 and 4, we get the following Lemma.

Lemma 6 *If $\Gamma \in \mathcal{G}(I, J-1, 0)$, $\delta \ge \delta_0(\Gamma)$, and $\delta_c \ge \delta_{c0}(f_\delta(\Gamma))$, then $\Sigma = h(f_\delta(\Gamma), \delta_c, \delta_c, \ldots, \delta_c) \in \mathcal{S}(I, J, \delta, \delta_c)$, and $m(\Sigma) = \mu(\Gamma) + (J-1)I\delta + I\delta_c$.*

Lemma 7 *a) If $\delta \le \delta_c$, then $m(I, J, \delta, \delta_c) \ge I(J-1)\delta + I\delta_c$*
b) If $\delta > \delta_c$, then $m(I, J, \delta, \delta_c) \ge (IJ-1)\delta_c + \delta$.

We note that for $\delta_c \ge \delta$, the lower bound in Lemma 7 a) and the upper bound implied by Lemma 6 differs by a constant independent of δ and δ_c. Based on this observation and support from numerical data, we put forward the following conjectures:

Conjecture 1 *For given I and J there exists a bounds $\delta(I, J)$ and $\Delta(I, J)$ and a constant $\nu(I, J)$ such that $m(I, J, \delta, \delta_c) = \nu(I, J) + (J-1)I\delta + I\delta_c$ for $\delta \ge \delta(I, J)$ and $\delta_c \ge \delta + \Delta(I, J)$.*

Conjecture 2 *For given I, J, and l, there exists a bound $\delta(I, J, l)$ and a constant $\nu(I, J, l)$ such that $m(I, J, \delta, \delta + l) = \nu(I, J, l) + IJ\delta$ for $\delta \ge \delta(I, J, l)$.*

If both conjectures are true, then $\nu(I, J, l) = \nu(I, J) + Il$ for $l \ge \Delta(I, J)$. Hence, three of four conjectures formulated in [1] are proved.

References

[1] Yu. V. Svirid, "Diffuse Codes with Minimal Guard Space," *IEEE International Symposium on Information Theory*, Trondheim, Norway, 26th June – 1st July 1994, p. 24.

Two Classes of Binary Optimum Constant-Weight Codes

Fang-Wei Fu and Shi-Yi Shen
Department of Mathematics, Nankai University
Tianjin 300071, P.R.China

[1] presented a new construction method for binary constant-weight cyclic codes. By slightly modifying this method, we could construct new constant-weight codes (not necessaryly cyclic). Furthermore, two classes of binary optimum constant-weight codes could be constructed by using this modified method. In general, we show that binary optimum constant-weight codes, which achieve Johnson bound, could be constructed from codes over $GF(q)$ which achieve Plotkin bound.

The cyclic order of $a = (a_0, \ldots, a_{N-1}) \in [GF(2)]^N$ is denoted as $t(a)$, i.e. the smallest positive integer t such that $a = S^t(a) = (a_t, \ldots, a_{N-1}, a_0, \ldots, a_{t-1})$. It is clear that $\mathcal{A}(a) = \{a, S(a), \ldots, S^{t(a)-1}\}$ form a binary constant-weight code with length N, weight $w(a)$, and its minimum distance is denoted as $d(a)$. Given a (n, M, d) code C in $GF(q)$, $v \in [GF(2)]^N$ with cyclic order q, and an one to one mapping $f : GF(q) \to \mathcal{A}(v)$, denote

$$C(v, f) = \{(f(c_0), \ldots, f(c_{n-1}) | c = (c_0, \ldots, c_{n-1}) \in C\}$$

Proposition 1 *$C(v, f)$ is a binary constant-weight code with length nN, weight $nw(v)$, minimum distance $d(v)d$, and codeword number M.*

Proposition 2

$$A_2(nN, d(v)d, nw(v)) \geq A_q(n, d)$$

Construction 1 (ref. [1]) $\alpha = (1, 0, \ldots, 0) \in [GF(2)]^q$, $t(\alpha) = q$, $w(\alpha) = 1$, $d(\alpha) = 2$
Construction 2 (ref. [1]) $q = p$, prime, and $\frac{p-1}{2}$ is odd, $\beta \stackrel{def}{=}$ Legendre sequence of length p, $t(\beta) = p$, $w(\beta) = \frac{p+1}{2}$, $d(\beta) = \frac{p+1}{2}$

Proposition 3 *(1)* $A_2(nq, 2d, n) \geq A_q(n, d)$
(2) if p is prime, and $\frac{p-1}{2}$ is odd, then

$$A_2(np, d\frac{p+1}{2}, n\frac{p+1}{2}) \geq A_p(n, d)$$

Lower bounds for $A_2(n', d', w)$ could be obtained from lower bounds for $A_q(n, d)$, e.g. Gilbert-Varshamov bound, and optimum codes in $GF(q)$, e.g. Hamming codes, Golay codes, R-S codes, MDS codes, simplex codes.

Proposition 4 *If C is a optimum (n, M, d) code in $GF(q)$, which achieves Plotkin bound, i.e. $M = \frac{d}{d-\lfloor n(q-1)/q \rfloor}$, $d > n(q-1)/q$. then $C(\alpha, f)$ and $C(\beta, f)$ are binary optimum constant-weight codes, which achieve Johnson bound.*

Generalized Hadamard matrix in $GF(q)$ could be used to construct codes in $GF(q)$, which achieve Plotkin bound, e.g. ref.[2]. If we take C be the simplex code, i.e. dual code of Hamming code in $GF(q)$, we obtain two classes of binary optimum constant-weight codes.

Proposition 5 *(1)* $A_2(q\frac{q^m-1}{q-1}, 2q^{m-1}, \frac{q^m-1}{q-1}) = q^m$
(2) if p is prime, and $\frac{p-1}{2}$ is odd, then

$$A_2(p\frac{p^m-1}{p-1}, p^{m-1}\frac{p+1}{2}, \frac{p^m-1}{p-1}\frac{p+1}{2}) = p^m$$

If C is a binary optimum code which achieves Plotkin bound, then $C(\alpha, f)$ is an optimum blanced error-correcting code, therefore we could use Hadamard matrix to construct optimum blanced error-correcting codes.

References

[1] Nguyen Q.A., L. Györfi and J.L. Massey, " Constructions of binary constant weight cyclic codes and cyclically permutable codes," *IEEE Trans. Inform. Theory,* Vol.38, No.3, pp.940-949, 1992

[2] G. Mackenzie, and J. Seberry, " Maximal ternary codes and Plotkin's bound," *ARS Combinatoria,* Vol.17A, pp.251-270,1984

Tensor Codes for the Rank Metric

Ron M. Roth
Computer Science Department
Technion, Haifa 32000, Israel.
e-mail: ronny@cs.technion.ac.il

Abstract — **Linear spaces of $n \times n \times n$ tensors over finite fields are investigated where the rank of any nonzero tensor in the space is at least a prescribed number μ. Such spaces can recover any $n \times n \times n$ tensor of rank $\leq (\mu-1)/2$, and, as such, they can be used to correct three-way crisscross errors. Bounds on the dimensions of such spaces are given for $\mu \leq 2n+1$, and constructions are provided for $\mu \leq 2n-1$ with redundancy which is linear in n. These constructions can be generalized to spaces of $n \times n \times \cdots \times n$ hyper-arrays.**

I. Introduction

An $n \times n \times n$ tensor over a field F is an $n \times n \times n$ array $\Gamma = [\Gamma_{i,j,\ell}]_{i,j,\ell=1}^{n}$ whose entries $\Gamma_{i,j,\ell}$ are in F. A tensor $\Gamma = [\Gamma_{i,j,\ell}]_{i,j,\ell=1}^{n}$ over F is called a *rank-one* tensor if there exist three nonzero vectors $[a_i]_{i=1}^{n}$, $[b_j]_{j=1}^{n}$, and $[c_\ell]_{\ell=1}^{n}$ over F such that $\Gamma_{i,j,\ell} = a_i b_j c_\ell$ for $i,j,\ell = 1,2,\ldots,n$. The *rank* of an $n \times n \times n$ tensor Γ is the smallest number ρ of rank-one tensors Γ_m such that $\Gamma = \sum_{m=1}^{\rho} \Gamma_m$. The definition of tensor rank is a generalization of that of matrix rank and can be extended to $n^{\times\Delta}$ hyper-arrays over F.

A μ-$[n^{\times\Delta}, k]$ hyper-array code $\mathcal{C}$ over a field F is a k-dimensional linear subspace of the vector space of all $n^{\times\Delta}$ hyper-arrays over F where μ is the smallest rank of any nonzero hyper-array in $\mathcal{C}$. We call $n^\Delta - k$ the *redundancy* of $\mathcal{C}$ and μ the *minimum rank* of $\mathcal{C}$. We will use the terms *array codes* and *tensor codes* for the cases $\Delta = 2$ and $\Delta = 3$, respectively.

The minimum-rank Singleton bound for μ-$[n^{\times\Delta}, k]$ hyper-array codes over a field F takes the form

$$n^\Delta - k \;\geq\; (\mu - 1)\, n \;.$$

This bound was stated by Delsarte in [1] for the case $\Delta = 2$. Furthermore, Delsarte obtained a construction of μ-$[n \times n, k]$ array codes over $GF(q)$ that attains this bound for every $\mu \leq n$ (see also [2] and [4]).

In [3] and [4], it was shown how a certain model of errors — so-called crisscross errors — can be handled optimally by using such array codes. A discussion was given in [4] also for larger Δ. There are various applications of the crisscross error model. In particular, the three-way crisscross model of errors in tensors (i.e., the case $\Delta = 3$) can be found in practice in certain memory chips. Tensor rank is closely related also to the multiplicative complexity of sets of bilinear forms, such as polynomial multiplication or matrix multiplication.

The purpose of this work is to continue the work of [1], [2], and [4] and present constructions of linear spaces of $n^{\times\Delta}$ hyper-arrays for $\Delta \geq 3$ while obtaining bounds on the dimensions of such spaces. We mainly concentrate on bounds and constructions of μ-$[n^{\times\Delta}, k]$ hyper-array codes over finite fields with $\mu = O(n)$.

II. Bounds

Theorem 1. *For any 3-$[n^{\times\Delta}, k]$ hyper-array code,*

$$n^\Delta - k \;\geq\; \Delta n \;-\; (\Delta - 1)\log_q(q-1) \;-\; O(\Delta/(q^n \log q)) \;.$$

Theorem 2. *Let $\mu \leq 2n+1$. Then, for every μ-$[n^{\times\Delta}, k]$ hyper-array code,*

$$n^\Delta \;-\; k \;\geq\; \Delta\lfloor(\mu-1)/2\rfloor\, n\,(1 - \epsilon_\Delta(n)) \;,$$

where $\lim_{n\to\infty} \epsilon_\Delta(n) = 0$.

III. Construction of tensor codes

Let $\{\alpha_i\}_{i=1}^{n}$, $\{\beta_j\}_{j=1}^{n}$, and $\{\omega_\ell\}_{\ell=1}^{n}$ be three bases of $GF(q^n)$ over $GF(q)$. Define the tensor code $\mathcal{C}(n,\mu,3;q)$ as the set of all tensors $\Gamma = [c_{i,j,\ell}]_{i,j,\ell=1}^{n}$ over $GF(q)$ such that

$$\textstyle\sum_{i,j,\ell=1}^{n} c_{i,j,\ell}\, \alpha_i^{q^r} \beta_j^{q^s} \omega_\ell \;=\; 0 \;,$$

where r and s range over all nonnegative integers such that (a) $0 \leq r,s < n$, and (b) there exists a (conventional) linear $[\mu-1, r+1]$ code over $GF(q)$ with minimum Hamming distance $s+1$. In particular, by the Singleton bound on the minimum Hamming distance we have $r + s \leq \mu - 2$. Hence, we obtain the following upper bound on the redundancy of $\mathcal{C}(n,\mu,3;q)$:

$$n^3 - k \;\leq\; \begin{cases} \binom{\mu}{2}\, n & \text{for } \mu = 1,2,\ldots,n \\ n^3 - \binom{2n-\mu+1}{2}\, n & \text{for } \mu = n+1,\ldots,2n-1 \end{cases} .$$

Theorem 3. *The minimum rank of $\mathcal{C}(n,\mu,3;q)$ is at least μ.*

Generalizing the construction for any Δ, we can obtain μ-$[n^{\times\Delta}, k]$ hyper-array codes $\mathcal{C}(n,\mu,\Delta;q)$ whose redundancy is bounded from above by $\binom{\mu+\Delta-3}{\Delta-1}\, n$. For $\Delta = 2$, the codes $\mathcal{C}(n,\mu,2;q)$ coincide with those of Delsarte [1].

The construction $\mathcal{C}(n,\mu,\Delta;q)$ attains the Singleton bound when $\mu = 2$. For $\mu = 3$ we get redundancy $n^\Delta - k = \Delta n$, which, in view of Theorem 1, is optimal over $GF(2)$ for any fixed Δ and sufficiently large n. In general, for any fixed μ, the redundancy of $\mathcal{C}(n,\mu,\Delta;q)$ is *linear* in n, which is smaller than a redundancy proportional to $n \log_q n$ that would be needed in the simpler skewing crisscross coding method.

References

[1] Ph. Delsarte, *Bilinear forms over a finite field, with applications to coding theory, J. Comb. Th. A,* 25 (1978), 226–241.

[2] E.M. Gabidulin, *Theory of codes with maximum rank distance, Probl. Peredach. Inform.,* 21 (1985), 3–16 (in Russian; pp. 1–12 in the English translation).

[3] E.M. Gabidulin, *Optimal array error-correcting codes, Probl. Peredach. Inform.,* 21 (1985), 102–106 (in Russian).

[4] R.M. Roth, *Maximum-rank array codes and their application to crisscross error correction, IEEE Trans. Inform. Theory,* IT-37 (1991), 328–336.

On the Asymptotic Properties of a Class of Linearly Expanded Maximum Distance Separable Codes

Siddhartha Ray-Chaudhuri
Memotec Communications Corporation, North Andover, MA 01845, USA

Abstract— **Based on the average weight distribution of linearly expanded codes, we study their asymptotic characteristics.**

I. Introduction

In this paper, we study the asymptotic properties of linearly expanded (LE) maximum distance separable (MDS) codes. We show that there is a class of LE MDS codes in which most members are asymptotically good. A *time-varying* code is also discussed, based on the asymptotic goodness of LE MDS codes.

II. Average Weight Distribution

The *average weight distribution* (AWD) of LE codes is defined as follows. Pick any (N, K, D) block code $\mathcal{C}$ over GF(q^m), and list all nonzero N-tuples over the multiplicative group of GF(q^m). With each N-tuple, multiply the columns of the code, which yields a total of $(q^m-1)^N$ block codes over GF(q^m). Finally, expand each code with a fixed basis, to obtain a class $\mathbf{C}_x$ of (n,k) q-ary LE codes, where $n = mN$ and $k = mK$. Consider now the q-ary weight $i, 0 \le i \le mN$. Let G_i denote the average number of weight-i codewords in a code in $\mathbf{C}_x$. We refer to the set $\{G_i\}_{i=0}^{mN}$ as the q-ary AWD of $\mathbf{C}_x$. The sum $G_{h\to j} = \sum_{i=h}^{j} G_i$ is called the *cumulative* AWD (CAWD) of $\mathbf{C}_x$ between the weights h and $j, h \le j$. G_i has been derived for a class of generalised Reed-Solomon (GRS) codes [2], where $N = 2^m - 1$ and $q = 2$. More general expressions for G_i and the CAWD, applicable to any q^m-ary MDS code, $q \ge 2$, have also been derived [1].

Most well-known MDS codes, *e.g.,* GRS codes, satisfy $2 \le K \le N-2$. For such codes, the CAWD is upper-bounded by [1],

$$\begin{aligned} G_{0\to mN\delta} &< q^{N+3+m(K-N)} \sum_{i=0}^{mN\delta} (q-1)^i \binom{mN}{i} \\ &\le q^{N+3+m(K-N)+mNH_q(\delta)} \end{aligned} \quad (1)$$

where $0 < \delta < (q-1)/q$, and $H_u(x)$ is the u-ary *entropy function*, $u \ge 2$.

III. Asymptotic Properties

Let $\mathbf{C}_x$ be the class of q-ary LE codes obtained from a q^m-ary (N, K, D) MDS code, where $2 \le K \le N-2$. Let d denote the minimum distance of any member of $\mathbf{C}_x$.

Theorem 1 *For any $\epsilon > 0$, there is an integer $N_0 > 0$ such that a majority of codes in $\mathbf{C}_x$ satisfy $H_q(\frac{d}{mN}) > 1 - \frac{K}{N} - \epsilon, \forall N > N_0$.*

In other words, most codes in $\mathbf{C}_x$ are asymptotically good. The CAWD can be used to study the minimum distances of the LE codes, as stated below.

Proposition 1 *The smallest q-ary weight d_{MDS}, such that $G_{0\to d_{\text{MDS}}} \ge 1$, is a lower-bound on the minimum distance of the best codes in the class. The largest q-ary weight d_{most}, such that $G_{0\to d_{\text{most}}} \le 0.5$, is a lower-bound on the minimum distance of most codes in the class.*

Fig. 1 shows the asymptotic behaviour of d_{most}. Here, $\mathcal{C}$ is a primitive Reed-Solomon (RS) code over GF(2^m). We have computed d_{most} for $5 \le m \le 9$. We also show the BCH bound for comparable primitive binary BCH codes. Evidently, d_{most} is asymptotically good.

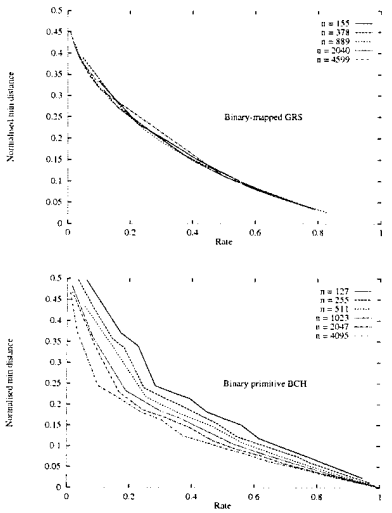


Figure 1: Asymptotic Behaviour

IV. Time-Varying Code

Time-varying code is a pseudo-random code based on all the members of the class $\mathbf{C}_x$. To explain how the code works, let us index the member codes with consecutive integers: $\mathbf{C}_x = \{\mathcal{C}_{x,l}\}_{l=0}^{l=(q^m-1)^N-1}$. To encode the r^{th} block of information, we shall use the code $\mathcal{C}_{x,\ell}$, where $\ell = r \pmod{(q^m-1)^N}$. Since most members of $\mathbf{C}_x$ are good, we are likely to pick a good code most of the time. With $\mathcal{C}$ being the same Reed-Solomon code as before, it has been found that the best codes in $\mathbf{C}_x$ dominate the overall performance of the time-varying code, provided every member is decoded to its true minimum distance [1].

Acknowledgements

This work was supported by NSF Grant No. NCR90031004. The author wishes to thank Dr. Charles Retter for many valuable comments, and acknowledges the support of Memotec Communications Corporation.

References

[1] Ray-Chaudhuri, S.: "On Coordinate Codes of Nonbinary Codes," Ph.D. Thesis, ECE Dept., Northeastern University, Boston, MA 02115, USA, August 1994

[2] Retter, C.T.: "The Average Binary Weight-Enumerator for a Class of Generalized Reed-Solomon Codes," *IEEE Trans. Info. Theory,* **IT-37,** pp.346-349, March 1991

Generalized Partial Spreads, Geometric Forms of Bent Functions

Claude Carlet
INRIA Bat 10, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France

I. Introduction

Let $n = 2p$ be a positive even integer. Let $GF(2)$ denote the Galois field of order 2 and V_n the $GF(2)$-vectorspace $(GF(2))^n$.
δ_0 denotes Dirac symbol ($\delta_0(x) = 1$ if $x = 0$ and 0 otherwise).
For any subset E of V_n or of V_p, the symbol ϕ_E denotes the characteristic function of E in V_n or V_p.
We distinguish between the addition in $\mathbf{Z}$, denoted by $+$, and the addition in $GF(2)$, denoted by $\oplus$.
For any Boolean function f, the complement of f is the function $f \oplus 1$.
We denote by $\chi(x)$ the character $(-1)^x$ on $GF(2)$. The Walsh transform of any real-valued function φ on V_n is defined on V_n by : $\widehat{\varphi}(s) = \sum_{x \in V_n} \varphi(x)\chi(x \cdot s)$, where "$\cdot$" denotes the usual dot product on V_n.
Let f be a Boolean function on V_n. We denote by $f_\chi(x)$ the function $\chi(f(x))$. The Walsh transform of $f_\chi(x)$ is the function :

$$\widehat{f_\chi}(s) = \sum_{x \in V_n} \chi(f(x) \oplus x \cdot s).$$

The Boolean function f is called bent if for any element s of V_n, $\widehat{f_\chi}(s)$ has absolute value 2^p . That is equivalent to the fact that f is at maximum Hamming distance from the set of all affine functions $g(x) = a \cdot x \oplus \epsilon$ ($a \in V_n$, $\epsilon \in GF(2)$).
A class of bent functions is called complete if it is globally invariant under any affine nonsingular transformation of the variable and under the addition of any affine function.
If a Boolean function f on V_n is bent, then the Boolean function $\tilde{f}$ defined by : $\widehat{f_\chi}(s) = 2^p \chi(\tilde{f}(s))$ is bent. $\tilde{f}$ is called the "Fourier" transform of f (cf. [2]).

The known bent functions belong to the completed versions of four classes:
1) Maiorana-Mc Farland's class [2], denoted by $\mathcal{M}$. It is the set of all the Boolean functions on V_n of the form : $f(x, y) = x \cdot \pi(y) \oplus g(y)$ where x and y belong to V_p, π is a permutation on V_p and g is a Boolean function on V_p.
2) Partial Spreads class [2], denoted by $\mathcal{PS}$, whose elements are the sums (modulo 2) of the characteristic functions of 2^{p-1} or $2^{p-1}+1$ "disjoint" p-dimensional subspaces of V_n ("disjoint" meaning that any two of these spaces intersect in 0 only, and therefore that their sum is direct and equal to V_n).
3) Class $\mathcal{D}$ [1] which is the set of all the functions of the form : $f(x, y) = x \cdot \pi(y) \oplus \phi_{E_1}(x)\phi_{E_2}(y)$ where π is any permutation on V_p and E_1, E_2 are any linear subspaces of V_p such that $\pi(E_2) = E_1^\perp$.
4) Class $\mathcal{C}$ [1] which is the set of all the functions of the form : $f(x, y) = x \cdot \pi(y) \oplus \phi_L(x)$ where π is any permutation on V_p, L is any linear subspace of V_p such that, for any element λ of V_p, the set $\pi^{-1}(\lambda + L^\perp)$ is a flat.

II. Generalized Partial Spreads, geometric forms of bent functions

Our main result is the following:

Theorem 1 *Let $\{E_1, \cdots, E_k\}$ be a family of p-dimensional subspaces of V_n and $m_1, \cdots, m_k$ (positive or negative) integers. Let $f(x)$ be a Boolean function on V_n. Assume that :*

$$\sum_{i=1}^{k} m_i \phi_{E_i}(x) = 2^{p-1}\delta_0(x) + f(x) \quad (i)$$

then f is bent and

$$\sum_{i=1}^{k} m_i \phi_{E_i^\perp}(x) = 2^{p-1}\delta_0(x) + \tilde{f}(x). \quad (ii)$$

We denote by $\mathcal{GPS}$ the class of all functions which satisfy (i). We call (i) a geometric form of f.
Any element of class $\mathcal{PS}$ belongs to class $\mathcal{GPS}$. Any element of class $\mathcal{M}$ or of class $\mathcal{D}$ is equivalent, up to a translation on the variable, to one of the elements of class $\mathcal{GPS}$, or to its complement. Thus, it belongs to the completed version of class $\mathcal{GPS}$.
For any element f of class $\mathcal{GPS}$ and any linear isomorphism ψ of V_n, the functions $f \circ \psi$ and $f \oplus 1$ belong to $\mathcal{GPS}$. However, class $\mathcal{GPS}$ is not complete.

III. New bent functions deduced from the theorem

Proposition 1 *Let $n = 2p$ be any even integer. Let π, π' and π'' be three permutations on V_p such that $\pi + \pi'$ and $\pi + \pi''$ are permutations. Assume that their inverses are $\pi^{-1} + \pi'^{-1}$ and $\pi^{-1} + \pi''^{-1}$ (respectively). Let ϵ and η be two elements of $GF(2)$, and $f(x, y)$ the Boolean function defined on V_n by :*

$$f(x, y) = (x \cdot \pi(y) \oplus 1)(x \cdot \pi'(y) \oplus \epsilon) \oplus (x \cdot \pi(y))(x \cdot \pi''(y) \oplus \eta)).$$

Then f is bent.

We have checked that these functions do not belong in general to the completed class of $\mathcal{M}$.

IV. A new characterization of bent functions

The theorem extends straightfully to a more general framework: let f be a Boolean functions on V_n ; let $E_1, \cdots, E_k$ be p-dimensional subspaces of V_n and $m_1, \cdots, m_k$ integers; assume that

$$\sum_{i=1}^{k} m_i \phi_{E_i}(x) = 2^{p-1}\delta_0(x) + f(x) \ [mod\, 2^p] \ (i')$$

then f is bent.
We have proved, with Philippe Guillot, that the class of those Boolean functions that satisfy (i') is that of all bent functions.

References

[1] C. Carlet, Two New Classes of Bent Functions , Proceedings of EUROCRYPT'93, Advances in Cryptology, Lecture Notes in Computer Science 765, p. 77-101 (1994)

[2] J. F. Dillon, Elementary Hadamard Difference Sets, Ph. D. Thesis, Univ. of Maryland (1974).

Constructing Covering Codes via Noising

I. Charon, O. Hudry, A. Lobstein
Centre National de la Recherche Scientifique
Télécom Paris, Département INF
46 rue Barrault 75634 Paris Cedex 13, France

Abstract — **We show how a combinatorial optimization method, the noising method, can be used for constructing covering codes.**

I. Introduction

The *noising method* and its applications to some graph problems were described in [1] and [2] (see also [5] and [4]). It is a heuristic for combinatorial optimization problems of the form $min\{f(s) : s \in S\}$. The elements in S are called *solutions* and f is the *evaluation function*. A *transformation* is any operation transforming a solution $s \in S$ into a solution $s' \in S$. An *elementary transformation* is a transformation changing one feature of s without changing its global structure; it defines the *neighbourhood* $N(s)$ of a solution s as the set of all solutions s' obtained from s by an elementary transformation.

This makes possible the definition of an *iterative-improvement* method, the *descent* method : from a current solution s, take a solution $s' \in N(s)$. If $f(s') < f(s)$, take s' as the current solution, otherwise keep s. Iterate this process. When no $s' \in N(s)$ is better than the current solution s, a local minimum is reached (with respect to this neighbourhood, i.e., to this elementary transformation).

The noising method is based on descent. Starting with an initial solution, repeat the following steps :

- add noise to the data (in order to change the values of f).
- apply the descent method to the current solution for the noised data.

For each iteration, the amount of noise is decreased until it reaches 0 in last iteration. The final solution is the best solution computed during the process.

II. Noising for Covering Codes

Let $C \subseteq F_q^n$ be a q-ary code of length n. Its *covering radius* $t(C)$ is $t(C) = max\{d(z,C) : z \in F_q^n\}$. Let $K_q(n,t)$ be the smallest size of a q-ary code with length n and covering radius t. Function K has been extensively studied, in particular for q=2 or 3 (see [3] for a recent survey on covering radius). Upper bounds on K are obtained by constructions; some of them use heuristics based on descent, for exemple *simulated annealing*.

In the following we restrict ourselves to q=2, but there is no difficulty in extending it to any q.

The set of solutions S is the set of all binary codes of given length n and given size. The evaluation function f is the number of vectors in F_2^n at distance greater than t from the current solution $C \subset F_2^n$: $f(C) = |\{z \in F_2^n : d(z,C) > t\}|$. The goal is to have $f(C) = 0$, proving that $K_2(n,t) \leq |C|$. From a random initial solution C, a new solution C' is obtained by complementing one bit of one codeword (this defines our neighbourhood).

To add noise, we give to each vector $z \in F_2^n$ a value $v(z) \in [1-r, 1+r]$, where v is uniformely distributed and r is the rate of the additional noise. The noised function, $f_N(C)$, is given by : $f_N(C) = \sum_{z \in F_2^n, d(z,C)>t} v(z)$. When rate r is zero, then $v(z) = 1$ for all $z \in F_2^n$, and $f = f_N$.

If we find a code C such that $f(C) = 0$, we start again the whole process with a size decreased by one.

References

[1] I. Charon and O. Hudry, "The noising method: a new method for combinatorial optimization," *Operations Research Letters*, No. 14, pp. 133–137, 1993.

[2] I. Charon and O. Hudry, "La méthode du bruitage: application au problème du voyageur de commerce," *Rapport Interne* 93–D–003, Ecole Nationale Supérieure des Télécommunications, Paris, 12 p., 1993.

[3] G. D. Cohen, S. Litsyn, A. Lobstein, and H. F. Mattson, Jr., "Covering radius 1985–1994," *Rapport Interne* 94–D–025, Ecole Nationale Supérieure des Télécommunications, Paris, 76 p., 1994, submitted to *AAECC Journal*.

[4] B. Hajek, "Locating the maximum of a simple random sequence by sequential search," *IEEE Trans. Inform. Th.*, No. 6, pp. 877–881, 1987.

[5] R. Khasminskii, "Application of random noise to optimization and recognition problems," *Problems of Information Transmission*, No. 3, pp. 113–117, 1965.

On the Diamond Code Construction

C.P.M.J. Baggen and L.M.G.M. Tolhuizen

Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Abstract — **We introduce a new error correcting code, which we call *Diamond code*. Diamond codes combine the error correcting capabilities of product codes and the reduced memory requirements from CIRC, the code applied in the CD system.**

I. Diamond Code Construction

The Diamond Code C calls for two codes, C_1 and C_2, of equal length n and defined over the same alphabet. C consists of the bi-infinite strips of height n, with each column in C_1 and each diagonal in C_2.

A convenient way of constructing Diamond codes is by using linear weakly cyclic codes for C_1 and C_2.

Definition: A linear code B is called *weakly cyclic* if $(b_0, b_1, \ldots, b_{n-2}, 0) \in B \Leftrightarrow (0, b_0, b_1, \ldots, b_{n-2}) \in B$.

Suppose both C_1 and C_2 are weakly cyclic codes, with p and q parity symbols, respectively. The minimal span codewords in C look like $(p+1) \times (q+1)$ diamonds as indicated in Figure 1. By the weakly cyclic property of C_1 and C_2, these elementary

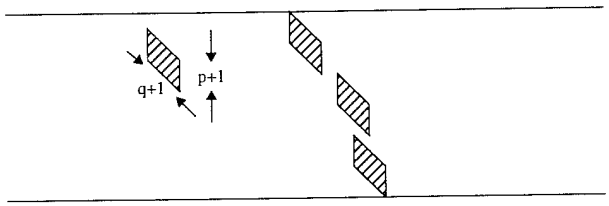


Figure 1: Elementary Diamond codewords

diamonds can be positioned anywhere within the code array. By taking suitable linear combinations of shifted elementary diamonds, we can produce codewords that are systematic in the $s = n - (p + q)$ top rows. Moreover, this construction shows that each information symbol in the upper s rows can affect the parities in at most $n - p$ columns.

II. Encoding

A Diamond code word whose columns contain only zeros for negative time, can efficiently be encoded by alternating C_1 and C_2 encodings, starting with the leftmost nonzero C_1 word and ending with the rightmost nonzero C_2 word. Such an encoding can be realized by the structure from Figure 2. The memory contents should be set to zero before the first data is fed into the encoder. The symbols immediately after the C_1 encoder correspond to columns of the Diamond code C, that are written to the channel. The feedback link in Figure 2 makes it an infinite impulse response structure. The remark at the end of Section I, however, implies that the structure from Figure 2 has a finite impulse response if C_1 and C_2 both are weakly cyclic and the encoder is initialized at the all zero state.

III. Decoding

A decoder for C is obtained by combining decoders for C_1 and C_2. In Figure 3, we show how a Diamond decoder is related to the decoder configuration of CIRC (Compact Disc) [1]. The

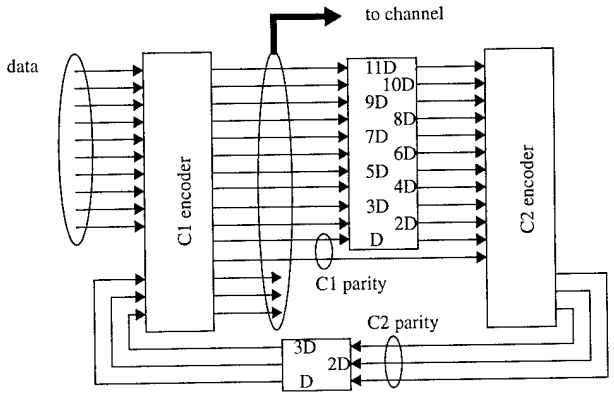


Figure 2: Systematic Diamond encoder

Diamond decoder applies iterative decoding which is known to be very powerful, especially for correcting random errors and short bursts. For an optimal performance, all symbols should be checked by both C_1 and C_2. A Diamond code does so (like a product code), but CIRC does not. Like CIRC, a Diamond code allows for a Forney interleaver between consecutive decoding stages (the "delay" triangles in Fig. 3), thus reducing the memory requirements, while retaining the distance and decoding potential of the product code of C_1 and C_2.

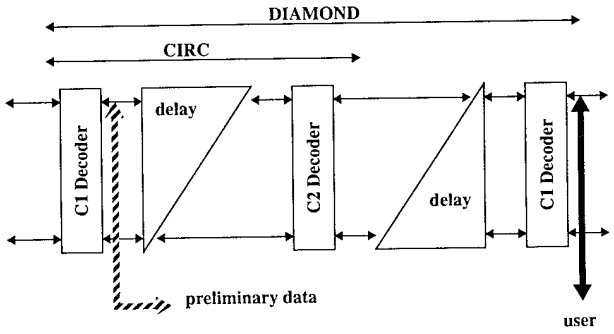


Figure 3: Decoding formats

IV. Block variations

For data recording applications, there is a need for independent, randomly rewritable data blocks. Three types of block codes will be discussed that share many features with C, which allows us to share much of their encoding and decoding hardware with the hardware for C. Each of them offers a different trade-off between rate, performance and similarity with the parent code C as a function of the blocksize.

References

[1] J.B.H. Peek. Communications aspects of the compact disc digital audio system. *IEEE Communications Magazine*, 23(2):7–15, February 1985.

A Five-head, Three-track, Magnetic Recording Channel

Emina Soljanin
AT&T Bell Labs, Rm. 2C-169, 600 Mountain Avenue, Murray Hill, NJ 07974
Costas N. Georghiades
Electrical Engineering Department, Texas A&M University, College Station, Texas 77843-3128

Abstract — **Performance loss caused by the inter-track interference (ITI) in recording channels may be alleviated through the use of multiple-head systems simultaneously writing and reading a number of adjacent tracks. We consider a five-head, three-track system, and show that there is no loss in performance of the system due to ITI, under some broad assumptions. We also show that, under these assumptions, the codes designed to provide certain coding gain in single-track, single-head systems, provide the same coding gain in five-head, three-track systems.**

I. Summary

We consider disk recording systems where inter-track interference can be described as follows: only adjacent tracks interfere, and when a reading head is positioned over one of the tracks, it responds to the magnetization of an adjacent track as if it were positioned over that track but with an amplitude modified by a weighting parameter α. Information is written in N_t adjacent tracks and simultaneously detected by N_h reading heads. We analyse a discrete-time model for the magnetic recording channel with input $\{a_n^k\}$, $1 \le k \le N_t$, impulse response $\{h_n\}$, and output $\{y_n^k\}$ $1 \le k \le N_h$, given by

$$y_n^k = \sqrt{E}\sum_m (\alpha a_m^{k-1} + a_m^k + \alpha a_m^{k+1})h_{n-m} + \eta_n^k,$$

where h_n are integer, η_n^k are independent Gaussian random variables with zero mean and variance σ^2, and E is a constant related to the output voltage amplitude. We refer to E/σ^2 as the signal-to-noise ratio (SNR) per track, and to $H(D) = \sum_n h_n D^n$ as the channel transfer function. Special cases of these systems with $N_t = N_h = 2$ have been studied by Barbosa [1], Siala and Kaleh [2], and Soljanin and Georghiades [3].

We compare the performance of various detection systems on the basis of minimum Euclidean distance, $d_{\min}$. This distance determines the performance for high values of SNR, when the probability of an error event in the system is closely approximated by $Q(d_{\min}\sqrt{\mathrm{SNR}})$.

Proposition 1 Let d_0 be the minimum distance of the composing single-track channels. (For example, for the $H(D) = (1-D)$ channel, $d_0^2 = 2$.) Then

$$d_{\min}^2 = \begin{cases} (1+2\alpha^2)d_0^2 & \text{if } 0 \le \alpha \le 1-\sqrt{2}/2, \\ 2(1+2\alpha^2-2\alpha)d_0^2 & \text{if } 1-\sqrt{2}/2 \le \alpha \le 1/2, \end{cases}$$

as long as $d_0^2 \le 6$.

Note that there is no performance loss due to ITI as long as $d_{\min}^2 \ge d_0^2$, *i.e.*, $0 \le \alpha \le 1/2$, which is the entire interval under consideration. Note also that the above condition holds for $H(D) = (1-D)(1+D)^N$, $N \in \{0,1,2,3\}$, *i.e.*, for the most common magnetic recording channel transfer functions.

Corollary 1 Under the assumptions of the preceding proposition, a single-track code that provides an increase in the single-track minimum distance to $d_0^c = \sqrt{g}d_0$ when applied to each track, results in an increase in the two-track minimum distance to $d_{\min}^c = \sqrt{g}d_{\min}$, as long as $d_{\min}^c \le \sqrt{6}$.

Note that the above holds for a dc–free coded $1-D$ channel as well as for a Nyquist–free coded $(1-D)(1+D)^2$ channel.

Performance of five different detection systems are compared Fig. 1, which plots $d_{\min}^2$ for each of the five cases as a function of the interference parameter α.

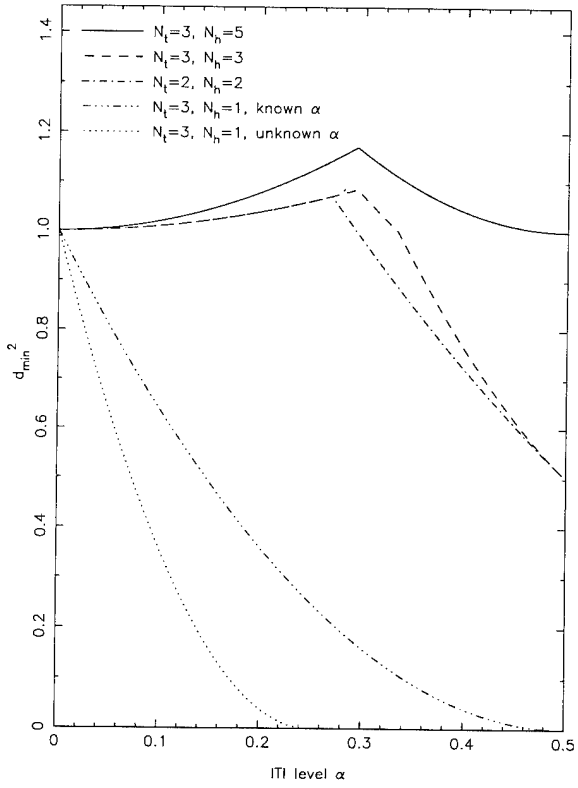


Figure 1: Performance of five different detection systems for channel of three interfering tracks.

References

[1] L. C. Barbosa, "Simultaneous detection of readback signals from interfering magnetic recording tracks using array heads," *IEEE Trans. Magn.*, vol. 26, no. 5, pp. 2163–2165, Sept. 1990.

[2] M. Siala and G. K. Kaleh, "A Two-track matched spectral-null code with rate 3/4 for the binary dicode channel," *IEEE Trans. Magn.*, vol. 30, no. 5, pp. 2778–2787, Sept. 1994.

[3] E. Soljanin, C. N. Georghiades "Coding for two-head recording systems," *IEEE Trans. Inform. Theory.* vol. 41, no. 3, pp. 747–755, May 1995.

Multi-track MSN Codes for Magnetic Recording Channels

Erozan Kurtas John G. Proakis Masoud Salehi
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115

Abstract — **In this paper we consider designing multi-track codes for parallel networks of partial-response channels. The codes we design have the capability of combating inter-track interference and also providing additional gains.**

In single-track magnetic recording systems data is coded, modulated and written on tracks independently from neighboring or any other tracks, with the same modulation code used on each track. The main idea of a multi-track system is to encode, write and read in parallel. Decreasing the k constraint is always desirable in recording systems, because this brings better synchronization, consequently higher access speeds. However, the price of a low k constraint is the lower code-rate, in other words the capacity loss. The advantage of using multi-track systems is the altered form of the k constraint. In a multi-track recording system, the k constraint on each channel is removed and a joint(vector) k constraint is imposed on the channel output sequences. As pointed out in [1], this provides higher rates and the information for timing and gain control can be obtained from any of the tracks which are coded jointly.

The multi-track system under investigation is modeled as a set of parallel partial-response channels. Each channel is of the form $(1-D)(1+D)^n$. In addition to this, the effect of the interference from each track to others is formulated by an inter-track interference (ITI) matrix, $\mathbf{T}$. Matched Spectral-Null (MSN) modulation codes which are intended for use on noisy partial-response channels with a finite input alphabet size were described in [2]. These codes provide significant increase in the minimum Euclidean distance, limit the maximum run length of identical samples and are designed to eliminate quasi-catastrophic sequences. The simplest approach to designing MSN codes for multi-track systems is to code each track independently. However, independent coding of each track for a multi-track system is undesirable for two reasons. First, it does not have the effect of decreasing the k constraint. Second, such a scheme ignores the existence of ITI which is inherent in all multi-track systems under investigation and is not expected to provide adequate coding gains except in very special forms of inter-track interference.

In this paper we consider designing multi-track codes for parallel networks of $(1-D)$ channels. The codes we design have the capability of combating ITI. Formally, let $\mathbf{F}$ be the graph representing the cross product of two canonical diagrams of single-track MSN codes with edge labels from $\{0,1\}^2$. Let $\mathbf{a}=\vec{a}_0,\ldots,\vec{a}_n$ and $\mathbf{b}=\vec{b}_0,\ldots,\vec{b}_n$ be sequences generated by paths in $\mathbf{F}$, $\Gamma_a=\{\sigma_0,\sigma_1^a,\sigma_2^a,\ldots,\sigma_n^a,\sigma_{n+1}^a\}$ and $\Gamma_b=\{\sigma_0,\sigma_1^b,\sigma_2^b,\ldots,\sigma_n^b,\sigma_{n+1}^b\}$. The sequence $\epsilon=\vec{\epsilon}_0,\ldots,\vec{\epsilon}_n$ with $\vec{\epsilon}_i=(a_{i1}-b_{i1},a_{i2}-b_{i2})$ is referred as the difference sequence corresponding to $\mathbf{a}$ and $\mathbf{b}$. If $\sigma_{n+1}^a=\sigma_{n+1}^b$ then ϵ is called a difference event, and if $\sigma_{n+1}^a=\sigma_{n+1}^b=\sigma_0$, then ϵ is called a difference cycle. The difference sequences corresponding to $\mathbf{a}$ and $\mathbf{b}$ on track 1 and track 2 are denoted as $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively. Let $\mathbf{A}_{(n+1)\times(n+1)}$ be the autocorrelation matrix corresponding to $(1-D)$ channel and define an inner product in R^{n+1} as $<\mathbf{u},\mathbf{v}>_{\mathbf{A}}=\mathbf{u}\mathbf{A}\mathbf{v}^t$ with $\mathbf{u},\mathbf{v}\in R^{n+1}$ For the network of two parallel $(1-D)$ channels with symmetric ITI coefficient α the distance between two allowable recorded sequences is given by [3],

$$d^2(\epsilon)=(1+\alpha^2)(\|\mathbf{e}_1\|_{\mathbf{A}}^2+\|\mathbf{e}_2\|_{\mathbf{A}}^2)+2.2\alpha<\mathbf{e}_1,\mathbf{e}_2>_{\mathbf{A}}.$$

To design two-track codes which are able to combat the performance loss caused by ITI, we consider subgraphs of $\mathbf{F}$ with the property that for any difference cycle contained in this subgraph the difference cycles corresponding to track 1 and track 2 satisfies, $\mathbf{e}_1\neq-\mathbf{e}_2$. We denote such a subgraph of $\mathbf{F}$ with $\mathbf{H}$. Then we have the following proposition which provides improvements over the results of [3],

Proposition : Let $\mathbf{a}=\vec{a}_0,\ldots,\vec{a}_n$ and $\mathbf{b}=\vec{b}_0,\ldots,\vec{b}_n$ be two distinct sequences generated by paths in $\mathbf{H}$, $\Gamma_a=\{\sigma_0,\sigma_1^a,\sigma_2^a,\ldots,\sigma_n^a,\sigma_0\}$ and $\Gamma_b=\{\sigma_0,\sigma_1^b,\sigma_2^b,\ldots,\sigma_n^b,\sigma_0\}$, with $\vec{a}_0\neq\vec{b}_0$.

$$\min_{\epsilon\neq\mathbf{0}} d^2(\epsilon)=\begin{cases}2(1+\alpha^2) & \text{if } 0\leq\alpha\leq\frac{3-\sqrt{5}}{2}\\ 4(1-\alpha)^2+2\alpha & \text{if } \frac{3-\sqrt{5}}{2}\leq\alpha\leq\frac{1}{2}\end{cases}$$

The capacity of $\mathbf{H}$ we considered was 0.7997. We designed codes with rates up to capacity and confirmed the gains predicted by the proposition with simulations.

References

[1] Michael W. Marcellin and Harold J. Weber, " Two-dimensional modulation codes ", IEEE J. Select. Areas Commun. Jan. 92 pp 254-265.

[2] Razmik Karabed and Paul H. Siegel, "Matched Spectral-Null Codes for Partial-Response Channels", IEEE Trans. on Inform. Theory May 91 pp 818-855.

[3] Emina Soljanin and Costas N. Georghiades, "Coding for two-head recording Systems", IEEE Trans. on Inform. Theory, May 95 pp 747-755.

MDS Array Codes with Independent Parity Symbols

Mario Blaum
IBM Almaden Research Center
650 Harry Rd., San Jose, CA 95120
blaum@almaden.ibm.com

Jehoshua Bruck*
California Institute of Technology
Pasadena, CA 91125
bruck@systems.caltech.edu

Alexander Vardy**
Coordinated Science Laboratory
University of Illinois, Urbana, IL 61801
vardy@golay.csl.uiuc.edu

Abstract — **A new family of MDS array codes is presented. The code arrays contain p information columns and r independent parity columns, where p is a prime. We give necessary and sufficient conditions for our codes to be MDS, and then prove that if p belongs to a certain class of primes these conditions are satisfied up to $r \leq 8$. We also develop efficient decoding procedures for the case of two and three column errors, and any number of column erasures. Finally, we present upper and lower bounds on the average number of parity bits which have to be updated in an MDS code over GF(2^m), following an update in a single information bit. We show that the upper bound obtained from our codes is close to the lower bound and does not depend on the size of the code symbols.**

I. Introduction

This work is concerned with maximum distance separable (MDS) codes. The Reed-Solomon (RS) codes are a well-known example of MDS codes. However, with Reed-Solomon codes, (a) the encoding and decoding procedures are performed as operations over a finite field, and (b) an update in a single information bit requires an update in all the parity symbols and affects a number of bits in each symbol. These two properties of RS codes are quite undesirable for certain channels. Firstly, the fact that encoding/decoding is performed in a finite field makes it unfeasible to use large symbols, since the size of the field grows exponentially with the symbol size. Secondly, the fact that an update in a single information bit requires to re-compute most of the parity bits is particularly undesirable in storage applications where the stored data has to be frequently updated in real-time. In this work, we present a new family of MDS codes having the following two properties: encoding and decoding may be accomplished with simple cyclic shifts and XOR operations on the code symbols, without finite field operations; and an update in an information bit affects a minimal number of parity bits.

II. The New MDS Array Codes

Our new codes are based on recent work in array codes [1, 3]. We assume that the information is presented as a two-dimensional array of bits. Henceforth we will identify the symbols of an MDS code with the columns of such an array. Thus the errors that can occur are column errors.

A trivial example of an MDS array code of this type is a simple parity code. This code is defined by requiring that the last column in the array is a parity column, given by the exclusive-OR of the other columns. The first nontrivial generalization of the parity code is the EVENODD code introduced in [1]. The EVENODD code has columns of size $p-1$ for some prime p, and requires two parity symbols. It can correct one error or two erasures.

In this paper, we generalize the construction of the EVENODD code to a family of codes with p information columns and r parity columns, for $r \geq 1$. We assume that p is prime number, and let $M_p(x) = 1 + x + \cdots + x^{p-1}$ with $M_p(x) \in \mathbb{F}_2[x]$. Consider the code C whose entries are in the ring of polynomials modulo $M_p(x)$, defined by the parity-check matrix:

$$H = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & \alpha & \dots & \alpha^{p-1} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{r-1} & \dots & \alpha^{(r-1)(p-1)} & 0 & 0 & \dots & 1 \end{pmatrix}$$

It is not difficult to show that this code is MDS for all p when $r = 2$ or $r = 3$. However, this is no longer true when $r \geq 4$. We give necessary and sufficient conditions for the code to be MDS when $r \geq 4$. Although we determined completely the primes $p \leq 100$ for which this code is MDS when $r \leq 8$, checking the necessary and sufficient conditions in the general case may be very complex. Our solution to this problem is related to certain generalizations of Vandermonde determinants called alternants. Using alternants, we have been able to show that if 2 is primitive in $\mathbb{F}_p$, then our codes are MDS up to $r = 5$ for all $p \neq 3$, and up to $r = 8$ for all $p \notin \{3, 5, 11, 13, 19, 29\}$.

III. Decoding and Information Updates

We present a decoding algorithm for the case of two symbol errors, that is for $r = 4$. Notably, this algorithm does not require finite field operations. This extends the algorithms of [3], applicable only for the case of a single symbol error.

Finally, we present lower and upper bounds on the average number $\eta(C)$ of parity bits affected by an update in a single information bit. In particular, we investigate the behavior of $\eta(C)$ for MDS codes over GF(2^m). It is shown that for our codes $\eta(C)$ *does not depend* on the size of the code symbols. In contrast, we also show that for Reed-Solomon codes, as well as for the MDS codes of Blaum and Roth [3], $\eta(C)$ increases linearly with the symbol size.

All these properties of the new MDS array codes make them very well suited for applications where the size of the code symbols is required to be large. We refer the reader to [2] for further details.

*Supported by the NSF Young Investigator Award CCR-9457811, by the Sloan Research Fellowship, and by grants from the IBM Almaden Research Center and the AT&T Foundation.

**Research supported in part by a grant from the Joint Services Electronics Program.

References

[1] M. Blaum, J. Brady, J. Bruck and J. Menon, "EVENODD: an optimal scheme for tolerating double disk failures in RAID architectures," *IEEE Trans. Computers*, vol. 44, pp. 192–202, 1995.

[2] M. Blaum, J. Bruck and A. Vardy, "MDS array codes with independent parity symbols," IBM Research Report, RJ 9887 (87267), September 1994.

[3] M. Blaum and R.M. Roth, "New array codes for multiple phased burst correction," *IEEE Trans. on Information Theory*, vol. 39, pp. 66–77, 1993.

On The Capacity Rates Of Two-Dimensional Runlength Limited Codes

Zhongxing Ye[1]
Math.Science and Technology Institute
Department of Applied Mathematics
Jiao Tong University
Shanghai 200030
P.R.China

Zhen Zhang[2]
Communication Sciences Institute
Electrical Engineering-Systems
University of Southern California
Los Angeles,CA 90089-2565
U.S.A.

Abstract – We give a rigorous proof of Etzion and Wei's conjecture about the capacity rate of two-dimensional Runlength Limited Codes.We also provide an alternative approach to compute the capacity rate.

SUMMARY

Runlength-limited(RLL) codes are binary codes with the maximum and minimum runlength constraints of its codewords. Etzion and Wei[1] have studied the extension of runlength- limited codes to two-dimensions. One of the most fundamental problem is to determine the capacity rate of 2-D RLL codes, i.e., the highest code rate possible under a given set of runlength constraints.

A 2-D $n_1\times n_2$ $(d_1,k_1,d_2,k_2;d_3,k_3,d_4,k_4)$ array is an $n_1\times n_2$ binary array with the following parameters:

(1) $d_1(d_2)$ is the shortest run of ZEROs (ONEs) horizontally,

(2) $k_1(k_2)$ is the longest run of ZEROs (ONEs) horizontally,

(3) $d_3(d_4)$ is the shortest run of ZEROs (ONEs) vertically,

(4) $k_3(k_4)$ is the longest run of ZEROs (ONEs) vertically,

If the horizontal constraints are the same as the vertial constraint, then it is a (d_1,k_1,d_2,k_2) array.

The capacity rate of 2-D (d_1,k_1,d_2,k_2) arrays is defined as:

$$C = \lim_{n\to\infty} \frac{\log F(n_1,n_2)}{n_1\times n_2}$$

where $F(n_1,n_2)$ is the number of valid $n_1\times n_2$ -configurations with runlength constraints.

To determine the capacity rate of 2- D RLL code, we assign a Gibbs measure associated with the given 2-D runlength constraints as follows. Let

$$\Lambda^{(n)}=\{n=(n_1,n_2);0\leqslant |n_1|, |n_2|\leqslant n\}\triangleq \Lambda$$

be the finite box of Z^2. We define an energy function for configurations on the finite set Λ by

$$U(x^{\wedge}) = \sum_{C\,\Lambda} V_C(x^{\subset})$$

where

$$V_C(x^{\subset})=\begin{cases} J & \text{if C is a minimal violating subconfiguration} \\ 0 & \text{otherwise}\end{cases}$$

Then define an Gibbs field by:

$$P_\Lambda(x^{\wedge})=Z_\Lambda^{-1}\exp\{-U_\Lambda(x^{\wedge})\}. \quad (1)$$

where Z_Λ is the normalization factor called partition function.

Theorem 1. $C = \lim_{J\to\infty} h_j$,where

$$h_j = \lim_{n\to\infty} \frac{1}{|\Lambda^{(n)}|} H_{J,\Lambda}(X^{\wedge(n)})$$

is the entropy rate of the Gibbs field (1) associated with the 2-D RLL constraints for fixed J, and

$$H_{J,\Lambda}(X^{\wedge(n)}) = -\sum_{x^{\wedge}} P_\Lambda(x^{\wedge})\log P_\Lambda(x^{\wedge})$$

This result was first conjectured by Etzion and Wei [1]. We give a rigorous proof.

The determination of the entropy rate of a Gibbs measure is a notoriously difficult problem. We find an alternative formula for the capacity rate of 2-D RLL codes which can be used to approximate the capacity rate. Let us describe the idea by an simple example. For 2-D $(1,\infty,1,1)$ RLL code we define a sequence of matrices A_n as follows:

First find all $n\times 1$ column vectors which satisfy the column constraints and label them by $x_1,x_2,\ldots,x_{F(n)}$, where F(n) is the number of n-vectors which satisfy the constraints. We assign an $F(n)\times F(n)$ merging indicator matrix $A_n=(A_{i,j})$, where

$$A_{i,j}=\begin{cases} 1 & \text{if merging } x_i \text{ and } x_j \text{ results} \\ & \text{a valid } n\times 2 \text{ array} \\ 0 & \text{otherwise}\end{cases}$$

Then $$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

We define another sequence of matrices B_n's by

$$B_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Then we have the recursive form:

$$A_n = \begin{pmatrix} A_{n-1} & B_{n-1} \\ B_{n-1}^t & 0 \end{pmatrix} \text{ and } B_n = \begin{pmatrix} A_{n-1} \\ B_{n-1}^t \end{pmatrix}$$

where B_n^t is the transpose of B_n. Denote by u_n the largest eigenvalue of the matrix A_n, then we have the following theorem.

Theorem 2. For 2-D $(1,\infty,1,1)$ code, the capacity rate

$$C = \lim_{n\to\infty} \frac{1}{n} \log u_n.$$

This method can be extented to other 2-D RLL codes.

[1]Supported in part by Chinese NNSF and U.S. NSF Grant NCR-9205265

[2]Supported in part by U.S.NSF Grant NCR-9205265

REFERENCES

[1] Etzion & Wei, On two- dimensional run- length -limited codes, preprint, 1993

Physical Limits on the Storage Capacity of Magnetic Recording Media

Donald G. Porter and Joseph A. O'Sullivan[1]
Department of Electrical Engineering, Washington University, St. Louis, MO USA 63130-4899

Abstract — **A model representing the physical laws which govern magnetic recording media is presented. Previous results in the analysis of Hopfield neural networks may be extended and applied to this model to determine its storage capacity limits.**

I. Introduction

The central component of any magnetic recording system is the medium on which information is stored in magnetic patterns. The storage capacity of a recording medium cannot exceed the logarithm of the number of distinct magnetic patterns which can be sustained over time. These limits are fundamental to the physical nature of a given recording medium, irrespective of the devices or methods used for recording.

II. Medium Model

Let a medium be represented by a planar array of N square tiles, indexed by $i \in I$. Let θ_i be the orientation of the ith tile's easy axis of anisotropy. The θ_i are random variables whose values are fixed in the manufacture of the medium, independently drawn from a uniform distribution on the interval $[-\pi, \pi]$. The magnetization of tile i is $\mathbf{m}_i = s_i[\cos\theta_i, \sin\theta_i]^T$, where $s_i \in \{\pm 1\}$.

The normalized, effective magnetic field at tile i is

$$\mathbf{h}_i = K_e \sum_{j \in N(i)} \mathbf{m}_j + \sum_{j \in I, j \neq i} \frac{K_m}{(x_{ij}^2 + z_{ij}^2)^{5/2}} \begin{bmatrix} 2x_{ij}^2 - z_{ij}^2 & 3x_{ij}z_{ij} \\ 3x_{ij}z_{ij} & 2z_{ij}^2 - x_{ij}^2 \end{bmatrix} \mathbf{m}_j. \quad (1)$$

The constants K_e and K_m scale the relative strength of the exchange interaction, arising from neighbor tiles $[N(i)]$, and the magnetostatic interaction, arising from all tiles. $\mathbf{h}_i$ is resolved into components parallel and perpendicular to the easy axis of tile i, $h_{\parallel,i}$ and $h_{\perp,i}$, and the magnetic state evolves according to a modified Stoner-Wohlfarth model[1] update rule,

$$s_i^{\text{new}} = \begin{cases} s_i & h_{\parallel,i}^{2/3} + h_{\perp,i}^{2/3} < 1, \\ \text{sgn}(h_{\parallel,i}) & h_{\parallel,i}^{2/3} + h_{\perp,i}^{2/3} \geq 1 \end{cases} \quad (2)$$

The update is repeated at randomly selected tiles until a state is reached which undergoes no further changes.

III. Capacity Analysis

Only fixed points of the update rule are suitable for information storage. When there are F_N fixed points, the storage density is limited by $C = \frac{1}{N}\log_2 F_N$, expressed in units of bits per tile. C is a function of K_e, K_m, and the θ_i. Figure 1 displays the capacity limit C computed via simulation for a medium of 16 tiles arranged in a 4×4 array, using one realization of orientation angles, θ_i [2].

For small values of K_e and K_m, all magnetic states are stable, and $C = 1$. As K_e and K_m increase, the second line of the update rule has an effect. Let

$$\mathcal{A}_i = \{\mathbf{h}_i : h_{\parallel,i}^{2/3} + h_{\perp,i}^{2/3} < 1 \text{ or } h_{\parallel,i} > 0, 1 \leq i \leq N\}. \quad (3)$$

[1]This work was supported in part by NSF Grant NCR-94-06197.

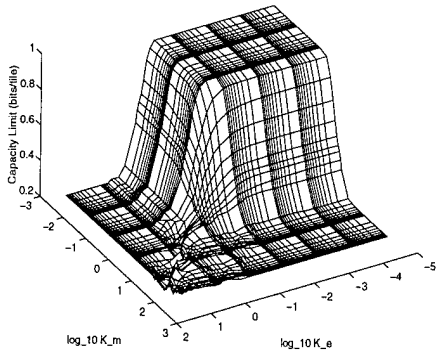


Figure 1: Capacity limit as a function of K_e and K_m

Application of DeMorgan's Law, the Union Bound, and symmetry properties of the model yields

$$C \geq 1 + \frac{1}{N}\log_2(1 - N \max_{1 \leq i \leq N} \Pr\{\mathbf{h}_i \notin \mathcal{A}_i\}). \quad (4)$$

For large N, edge effects may be ignored and the functional dependence of $\Pr\{\mathbf{h}_i \notin \mathcal{A}_i\}$ on K_e and K_m has been estimated numerically. The results determine that the analytical boundary of the region of the K_e-K_m plane for which all states are stable is consistent with simulation results.

For large values of K_e and K_m, the role of $h_{\perp,i}$ becomes insignificant, and the model may be simplified to

$$h_{\parallel,i} = \sum_{j \neq i} w_{ij} s_j, \quad s_i^{\text{new}} = \text{sgn}(h_{\parallel,i}), \quad (5)$$

with appropriate definitions of the w_{ij}. This case corresponds to the Hopfield model with random weights.

When the w_{ij} are independent standard normal random variables, for large N, the number of fixed points is [3]

$$F_N \approx 1.0505 \cdot 2^{0.2874N}. \quad (6)$$

Thus, a storage capacity limit of about 0.29 bits per tile would prevail if independent, identically distributed, zero-mean Gaussian random weights accurately reflected the medium model. For a typical tile size, the corresponding areal storage density limit is about 116 Gbits per square inch. Such analysis must be extended to determine the capacity limits when w_{ij} are given by the medium model.

References

[1] E. C. Stoner and E. P. Wohlfarth, "Mechanisms of magnetic hysteresis in heterogeneous alloys," *Proc. R. Soc.*, vol. A240, p. 599, 1948.

[2] J. A. O'Sullivan, D. G. Porter, R. S. Indeck, and M. W. Muller, "Recording medium properties and capacity bounds," *J. Appl. Phys.*, vol. 75, pp. 5753–5755, May 1994.

[3] P. Baldi and S. S. Venkatesh, "Number of stable points for spin-glasses and neural networks of higher orders," *Phys. Rev. Lett.*, vol. 58, pp. 913–917, March 1987.

Entropy Estimates for Simple Random Fields

Søren Forchhammer and Jørn Justesen
Institute of Telecommunication,
Technical University of Denmark

I. Introduction

We consider the problem of determining the maximum entropy of a discrete random field on a lattice subject to certain local constraints on symbol configurations. The results are expected to be of interest in the analysis of digitized images and two dimensional codes. We shall present some examples of binary and ternary fields with simple constraints. Exact results on the entropies are known only in a few cases, but we shall present close bounds and estimates that are computationally efficient.

II. Fields with Simple Constraints

We consider random variables on a rectangular grid, $x(i,j)$. The lattice is defined by the set of neighbors associated with a given point. We shall not assume that the probability distribution is given, but the structure of the field will be specified in terms of a set of constraints on the values assumed by a particular variable and its neighbors. Constraints could be of one of the following (not necessarily distinct) types:

- the runs of pixels of a given color should satisfy a set of inequalities [1]
- the field is a random tiling of the plane with certain pieces [2]
- certain configurations of values are excluded

Since we are interested in estimates of the entropy which may be related to coding and data compression, we consider fields which are obviously stationary. The existence of solutions to the constraints should not be a problem, and boundary conditions should not be important.

Example 1: As a simple example we shall consider the following problem which is quite well-known: Consider a binary field on a rectangular lattice with the restriction that two neighbors, i.e. $x(i,j)$ and $x(i,j+1)$ or $x(i,j)$ and $x(i+1,j)$, cannot both have the value 1. What is the largest possible entropy, or what is the number of solutions for an N by N segment of the lattice as a function of N? We estimate the entropy to be $H \approx 0.587891161775339$.

III. Markov Chains

As suggested in [1], the maximal entropy may be bounded by the entropy of a band of finite width, i.e. the variable j is restricted to $0 < j \leq m$. This entropy can be calculated as the maximal entropy of a finite state Markov chain, and from this approach we obtain an upper bound (with a suitable relaxation on the restrictions at the boundaries). This estimate converges slowly. For the problem of Example 1 we get $H < 0.5928$ for m=20 imposing no restrictions at the boundaries. Constraining the probability of a 1 at the boundaries a tighter bound may be obtained. In some cases it is possible to derive a very accurate estimate from this sequence of values, H_m. The estimate given in Example 1 was obtained as $H_{m+1} - H_m$ with m=16.

Another type of estimate may be obtained from finite state causal models of the field. If the outcome of the process is generated one pixel at a time, and the probability distribution of $x(i',j')$ is assumed to depend on a finite past context $i < i'$ or $i = i'$ and $j < j'$, then the entropy can be approximated by that of a finite Markov source. This approach gives some information about the properties of the field, but the model is only exact in a very simple case, which is discussed in the following section.

IV. Construction of Stationary Fields

An actual construction of a random field with known entropy is interesting both for simulation purposes and as a method for establishing lower bounds. It would be very desirable to have random fields where rows and columns were described by simple Markov chains. Unfortunately this appears to be possible only in the case of the Pickard lattices [3]. This is also the only case where the causal model of the field becomes a simple finite state source.

Example 2: A Pickard field consistent with the constraint considered in Example 1 may be constructed such that each row or column is a Markov chain with $P(1)$=1/5. In this case the entropy may be found explicitly as $H = 1/10 + 3/10 \log 3 = 0.575...$ Actually a slightly larger value may be obtained by varying the transition probabilities.

Clearly the solution to the maximum entropy problem is always a Markov random field. However, in general such fields are hard to analyze. We shall consider a construction which has much greater flexibility than the Pickard field, but still allows detailed analysis:

Let rows i and $i+1$ be generated by a Markov chain (or another unifilar finite state source), such that there is complete symmetry between the two rows. Their joint entropy can be easily calculated. The probability distribution may be extended to a stationary distribution on the entire plane by assuming that all pairs of rows have the same distribution, and that the probability of each row given the past depends on only the previous row. The entropy of a row given the previous row may, in some cases, be calculated from a hidden Markov source.

Example 3: For the constraint in Example 1, two successive rows may be generated by a symmetric 3-state Markov chain with entropy H_2. From this source it is possible to calculate the entropy of a single row (i), H_1, exactly. Whenever $x(i,j) = 1$, the state of the source is known, and the distribution of zero runs can be calculated. We find the entropy of the process as $H = H_2 - H_1$. The largest lower bound obtained in this way is 0.58783.

References

[1] V. K. Wei and T. Etzion, "On two-dimensional run-length limited codes", *IEEE Information Theory Workshop*, Brazil, 1992.

[2] R. Burton and R. Pemantle, "Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances", *Ann. Prob.*, vol. 21, pp. 1329–71, 1993.

[3] D.K. Pickard, "A curious binary lattice process", *J. Appl. Prob.*, vol. 14, pp. 717–731, 1977.

Wavelets and Lattice Spaces

Todd K. Moon[1]
Elect. & Comp. Eng. Dept., Utah State University,
Logan, UT. tmoon@moon.ece.usu.edu

Abstract — **The shift- and scale-orthogonality properties of wavelets and scaling functions provide a means of producing signal spaces of arbitrary dimensionality. The signal spaces are presented with an example trellis code on the E_6 lattice.**

I. Introduction

Let $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$ and $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k)$ represent scaled and shifted wavelet and scaling functions, respectively, where we take $j, k \in Z$. Normalized orthogonal wavelets and scaling functions have the following orthogonality properties:

$$\begin{aligned} < \psi_{jk}, \psi_{lm} > &= \delta_{j,l}\delta_{k,m} \\ < \psi_{jk}, \phi_{lm} > &= 0 \quad \forall j,k,l,m \\ < \phi_{jk}, \phi_{jm} > &= \delta_{k,m}. \end{aligned}$$

In addition, these functions have attractive frequency localization: $\phi(t)$ is a low-pass function and $\psi(t)$ is a band-pass function. A family of wavelets of interest is the compactly supported orthogonal wavelets introduced by I. Daubechies. Members of these families are denoted by D_N, N even, where N is the number of coefficients in the two-scale implicit description of the scaling function $\phi(t) = \sum_{n=0}^{N-1} c_n \phi(2t - n)$, which has support over $[0, N-1)$. The regularity (and hence the frequency localization) of member of D_N increases with N.

Considerable attention has been focused on wavelet and scaling functions over the last few years, due to their time/frequency localization ability and the existence of fast transform algorithms. In this paper, we introduce the application of wavelets and scaling functions as baseband waveforms for the transmission of digital signals.

Using a basic bit time normalized to unity, a baseband signal may be written as

$$s(t) = \sum_n I_n \phi(t - n)$$

where $\{I_n\}$ represents an alphabet of symbols drawn from a (possibly complex) signal constellation $\mathcal{C}$. For $\phi(t) \in D_N$ for $N > 4$, this signalling scheme has better spectral localization than MSK.

Signalling with several scales of wavelet functions may be written as

$$s(t) = \sum_n \sum_{s \in \mathcal{S}} I_{s,n}\psi_{s,n}(t) + \sum_n J_n \phi_{\sigma,n}(t)$$

where $\mathcal{S}$ is a set of scale indices, σ is a single fixed scale, and J_n is drawn from a constellation which may be empty (if the scaling function is not used for transmission). Table I indicates the dimensions for transmission with various sets of scales, where the scale notation is as follows: If a scaling function is used on a given scale, it is listed before the wavelet function for that scale. Absence of a wavelet on a particular scale is indicated by $-$.

Table 1: Dimensionalities obtainable using multiple scales.

Num. of Levels	Function Listing	Num. Dim.
2	$\psi\psi$	3
2	$\phi - \psi$	3
2	$\phi\psi\psi$	5
2	$\psi\phi-$	3
2	$\psi\phi\psi$	4
3	$\psi\psi\psi$	7
3	$\phi\psi\psi\psi$	11
3	$\phi - \psi\psi$	7
3	$\psi\phi\psi\psi$	9
3	$\psi\phi - \psi$	7
3	$\psi\psi\phi\psi$	8
3	$\psi\psi\phi-$	7

II. Trellis coding in wavelet signal spaces

Multiple-scale signalling provides a spectrally efficient way of producing arbitrary dimensionalities. The dimensions for coding available using multiple-scale signal provide rationale for exploring trellis codes in dimensions other than those usually explored. Due to limited space, we present only one example.

The code is based upon the construction of Calderbank and Sloane over the E_6 lattice. This is a lattice Λ with generator M over the Eisenstein integers and endomorphism Θ

$$M = \begin{bmatrix} \theta & 0 & 0 \\ 0 & \theta & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad \Theta = \begin{bmatrix} (1,2) & 0 & 0 \\ 0 & (1,2) & 0 \\ 0 & 0 & (1,2) \end{bmatrix}$$

where (a,b), $a, b \in Z$ represents the Eisenstein integer $a + b\omega$, $\omega = (1 + i\sqrt{2})/2$. The sublattice Λ' generated by $M' = M\Theta$ produces a quotient group Λ/Λ' with 64 cosets.

A generator for a four-state convolutional coder is

$$g = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This gives coding gains (for example) of 2.43 dB when $k_2 = 0$, 2.4 dB when $k_2 = 1$, and 3.4 dB when $k_2 = 2$. A computer search for better codes, both in E_6 and its dual E_6^*, is underway.

III. Extension

There is fertile ground for trellis code developments in other dimensions and for application of multi-scale signalling to fading channels.

Nonparametric Regression Estimation For Arbitrary Random Processes

S.E. Posner and S.R. Kulkarni[1]
Department of Electrical Engineering
Princeton University, Princeton, NJ 08544

Abstract — **We study nonparametric estimates of $E[Y_n|X_n]$ of the form $\sum_{i=1}^{n-1} W_{ni}(X_1,\ldots,X_n)Y_i$ based on X_n and data $\{(X_i,Y_i)\}_{i=1}^{n-1}$. Our work analyzes the case where $\{X_i\}$ is a completely arbitrary random process. Conditions on the weights are established so that the time-average of the estimation errors converges to zero. One consequence of our work is a recovery and extension of some classical results to stationary processes in separable metric spaces.**

I. Preliminaries

Let $X_1, X_2, \ldots, X_n$ be an arbitrary random process taking values in a compact subset of a separable metric space $(\mathcal{X}, \rho)$. Special cases include nonstationary or nonergodic processes and deterministic sequences. Each $X_i = x_i$ has an associated label Y_i which is a random variable, taking values in $\mathbb{R}$, drawn from an unknown conditional distribution $F(y|X_i = x_i)$. A classical problem is to nonparametrically estimate the regression function $m(X_n) = E[Y_n|X_n]$ given X_n and additional data pairs $\{(X_i,Y_i)\}_{i=1}^{n-1}$. Following the seminal work in [6], we consider estimators of the form $m_n(X_n) = \sum_{i=1}^{n-1} W_{ni}(X_1,\ldots,X_n)Y_i$ where $\overline{W}_n = \{W_{ni}(X_1,\ldots,X_n)\}_{i=1}^{n-1}$ is a sequence of probability weights that satisfy certain conditions.

Most previous work considered the case in which $\{(X_i,Y_i)\}$ are i.i.d. Stone [6] proved that certain conditions on the weights guarantee consistency without making any assumptions on the distributions other than that $\{(X_i,Y_i)\}$ are i.i.d. and that $EY^2 < \infty$. Universal consistency of specific estimators was discussed in, e.g., [2, 4]. In this paper, we impose no restrictions on the random process except that $\{X_i\}$ take values in a compact subset of a separable metric space and that the following assumption holds:

(A0) For each i and for every measureable set S,

$$Pr(Y_i \in S|X_1,\ldots,X_n,Y_1,\ldots,Y_{i-1}) = Pr(Y_i \in S|X_i)$$

This is a very general setting and as a result, in general, one cannot expect to design a consistent estimator. However, we show that when restricted to continuous regression functions, we can prove that conditions on the weights, analogous to those in [6], ensure time-average consistency, i.e., the time-averaged estimation errors go to zero for *every* random process. Our proof techniques are elementary sample path analyses and are in line with current trends in information theory and statistics to evaluate performance in terms of individual data sequences. See, for example, [1, 3, 5].

We impose the following assumptions on $F(y|x)$:

(A1) $\sup_{x\in A} E[Y^2|X = x] < \infty$,

(A2) $m(x) = E[Y|X = x]$ is a continuous function.

[1]This work was supported in part by the National Science Foundation under grants IRI-9209577 and IRI-9457645 and by the U.S. Army Research Office under grant DAAL03-92-G-0320.

II. Consistency

Our result is an extension of [6] to separable metric spaces and to arbitrary random processes. To achieve the latter, we can only conclude that the *time-averaged* squared loss goes to zero. Accordingly, it is natural that our statements on the weights are time-averaged versions of those in [6].

Theorem 1 *Let $\Omega_n = \{X_i\}_{i=1}^n$ be an arbitrary random process in a compact subset $A \subset (\mathcal{X}, \rho)$ with $\{(X_i,Y_i)\}_{i=1}^n$ satisfying (A0). Let $\{\overline{W}_n\}$ be probability weights. If*

(C1) $\frac{1}{N}\sum_{n=2}^{N}\sum_{i=1}^{n-1} W_{ni}(\Omega_n)I(\rho(X_n,X_i) > \epsilon) \to 0 \ \ \forall \epsilon > 0$,

(C2) $\frac{1}{N}\sum_{n=2}^{N} \max_i W_{ni}(\Omega_n) \to 0$,

hold a.s., then for $F(y|x)$ that satisfies (A1) and (A2) we have

$$\frac{1}{N}\sum_{n=2}^{N} E[|m_n(X_n) - m(X_n)|^2|\Omega_n] \to 0$$

Applying this theorem to the k_n-nearest neighbor and kernel algorithms, we can show that (C1)-(C2) are satisfied by the standard conditions $k_n \to \infty, k_n/n \to 0$ and $\epsilon_n \to 0, n\epsilon_n^r \to 0$ (in $\mathbb{R}^r$) resp. This means that these universally consistent estimates are, in fact, also universally time-averaged consistent for *every* random process $\{X_i\}$. As a corollary, we have the following partial extension of the pointwise consistency result of [6] to stationary processes in general metric spaces for continuous regression functions.

Theorem 2 *Let $\Omega_n = \{X_i\}_{i=1}^n$ be a stationary process in a compact subset $A \subset (\mathcal{X}, \rho)$ with $\{(X_i,Y_i)\}_{i=1}^n$ satisfying (A0). Let $\{\overline{W}_n\}$ be a sequence of weights. If*

(S1) $E\sum_{i=1}^{n-1} W_{ni}(\Omega_n)I(\rho(X_n,X_i) > \epsilon) \to 0$, *for all* $\epsilon > 0$,

(S2) $E\max_i W_{ni}(\Omega_n) \to 0$,

hold, then for $F(y|x)$ that satisfies (A1) and (A2) we have

$$E[|m_n(X_n) - m(X_n)|^2] \to 0$$

We have also shown Theorem 2 to hold when (A2) is omitted and (A1) is relaxed to $EY^2 < \infty$.

References

[1] T.M. Cover, "Universal portfolios," *Mathematical Finance* vol. 1, pp. 1-29, 1991.

[2] L. Devroye, "On the almost everywhere convergence of nonparametric regression function estimation," *Ann. Stat.* vol. 9, pp. 1310-1319, 1981.

[3] M. Feder, N. Merhav, and M. Gutman, "Universal prediction of individual sequences," *IEEE Trans. Information Theory*, vol. 38, pp. 1258-1270, 1992.

[4] L. Györfi, "Recent results on nonparametric regression estimate and multiple classification," *Prob. Contr. Inform. Theory*, vol. 10, pp. 43-52, 1981.

[5] S.R. Kulkarni and S.E. Posner, "Rates of convergence of nearest neighbor estimation under arbitrary sampling," *IEEE Trans. Information Theory*, July 1995.

[6] C.J. Stone, "Consistent nonparametric regression," *Ann. Stat.*, vol. 5, pp. 595-645, 1977.

Model Selection Criteria and the Orthogonal Series Method for Function Estimation

Pierre Moulin (moulin@bellcore.com)
Bell Communications Research, 445 South St., Morristown, New Jersey 07960, USA

Abstract — **Closed–form solutions are presented for function estimation using the orthogonal series method and various model selection criteria. While Akaike's AIC criterion does not lead to consistent estimates, a family of criteria that includes Minimum Description Length operates within a logarithmic factor of the minimax rate in a range of Sobolev smoothness classes.**

I. Model for Function Estimation

Consider the following model: $y_i = f(i/N)+\epsilon_i$, $0 \le i < N$, where ϵ_i are iid $N(0,1)$ and the sampled, unknown function f admits the representation $f = \sum_{n=0}^{N-1} \theta_n \phi_n$ in the orthonormal basis $\{\phi_n\}$ of $\mathcal{R}^N$. We investigate estimators of the form $\hat{f} = \sum_{n=0}^{N-1} \hat{\theta}_n \phi_n$ where $\hat{\theta}$ has $k \le N$ nonzero components and $\hat{\theta}$ and k are chosen so as to maximize the criterion

$$-\frac{1}{2}\sum_{i=0}^{N-1} |y_i - \hat{f}(i/N)|^2 - C_N k \qquad (1)$$

where the constant $C_N = 1$ (AIC [1]), or $C_N = \frac{3}{2}\log_2 N$ (MDL* [2][1]), or $C_N = \ln N$ (DJ, see below). Other model selection criteria (choices of C_N) may be considered. Related work may be found in [3], where the largest model order is restricted to be $o(N)$, and [4].

II. Basic Results

Proposition 1. The maximizer of (1) is $\hat{\theta}_n = T_\lambda(\tau_n)$, where $\tau_n = \sum_{i=0}^{N-1} y_i \phi_n(i)$, and

$$T_\lambda(\tau) \triangleq \begin{cases} 0: & |\tau| < \lambda \\ \tau: & \text{else} \end{cases} \qquad (2)$$

is the "hard threshold" function, with threshold $\lambda = \sqrt{2C_N}$.

Proof. By orthonormality of $\{\phi_n\}$ and Parseval's theorem, the criterion (1) takes the form $-\frac{1}{2}\sum_{n=0}^{N-1} |\tau_n - \hat{\theta}_n|^2 - C_N k = \sum_{n=0}^{N-1} L_n(\hat{\theta}_n)$ where

$$L_n(\hat{\theta}_n) = \begin{cases} -\frac{1}{2}|\tau_n - \hat{\theta}_n|^2 - C_N & : \text{if } \hat{\theta}_n \neq 0 \\ -\frac{1}{2}|\tau_n|^2 & : \text{else.} \end{cases}$$

The criterion may thus be maximized over each coordinate independently. The maximizer of $L_n(\hat{\theta}_n)$ is $\hat{\theta}_n = \tau_n$ if $-C_N > -\frac{1}{2}|\tau_n|^2$ and $\hat{\theta}_n = 0$ otherwise. The statement of the proposition follows directly. □

Prop. 1 establishes the fundamental role of thresholding in solving (1) and admits a hypothesis–testing interpretation. It also provides a closed–form expression for $\hat{\theta}$ which may be used to evaluate various properties of $\hat{f}$ for different choices of $\{\phi_n\}$ and C_N. For instance, when $\{\phi_n\}$ is a wavelet basis and $C_N \sim \ln N$, the estimator is almost minimax over a wide class of functions [5].

A basic problem is to evaluate the performance of the AIC and MDL* criteria in (1). We use the squared l^2 risk

$$R_N(\hat{f}) \triangleq E\left[N^{-1}\sum_{n=0}^{N-1} |f(n/N) - \hat{f}(n/N)|^2\right] = N^{-1}\sum_{n=0}^{N-1} \rho(\lambda, \theta_n) \qquad (3)$$

where $\rho(\lambda,\theta) \triangleq E|T_\lambda(\tau) - \theta|^2 = (\theta^2-1)[\Phi(\lambda-\theta) - \Phi(-\lambda-\theta)] + 1 + (\lambda-\theta)\phi(\lambda-\theta) + (\lambda+\theta)\phi(\lambda+\theta) \ge \lambda^2\Phi(-3\lambda)$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the normal pdf and cdf.

Proposition 2. For every choice of basis $\{\phi_n\}$,
(i) a necessary condition for consistency of $\hat{f}$ in the R_N sense is $C_N \to \infty$ as $N \to \infty$.
(ii) the AIC estimator is not consistent.

Proof. Using (3) and the lower bound on $\rho(\lambda,\theta)$ we obtain a necessary condition for $R_N(\hat{f}) \to 0$: $\lambda^2\Phi(-3\lambda) \to 0$, hence (i). (ii) follows immediately. □

III. Estimation Using Fourier Series

Convergence rates are computed for Fourier series and functions in the L_2 Sobolev ball $W_2^s(R) = \left\{f : \int_0^1 |f^{(s)}(t)|^2 dt \le R^2 < \infty,\ f^{(i)}(0) = f^{(i)}(1),\ 0 \le i \le s\right\}$. The estimator does not know s or R.

Proposition 3. For the choice

$$C_N = \beta \ln N, \quad 2s/(2s+1) \le \beta < \infty, \qquad (4)$$

the risk is $R_N(\hat{f}) \le C_{R,s}(\beta N^{-1}\ln N)^{2s/(2s+1)}$, where $C_{R,s}$ does not depend on N or β. The rate above is attained by the MDL* ($\beta = \frac{3}{2\ln 2} \approx 2.16$) and DJ ($\beta = 1$) estimators and is within a logarithmic factor of the minimax rate [6].

[1] This particular formulation of the MDL criterion accounts for the familiar $\frac{1}{2}k\log_2 N$ bits for encoding the k real parameters $\hat{\theta}_n$ and $k\log_2 N$ bits for encoding their index (assuming a uniform prior on indices).

References

[1] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. AC*, Vol. 19, No. 6, pp. 716—723, 1974.

[2] N. Saito, "Simultaneous Noise Suppression and Signal Compression Using a Library of Orthonormal Bases and the MDL Criterion," in *Wavelets in Geophysics*, Eds. E. Foufoula-Georgiou and P. Kumar, pp. 299—324, Academic Press, 1994.

[3] R. Shibata, "An optimal selection of regression variables," *Biometrika*, Vol. 68, No. 1, pp. 45—54, 1981.

[4] K.-C. Li, "Asymptotic Optimality for C_p, C_L, Cross–Validation and Generalized Cross–Validation: Discrete Index Set," *Ann. Stat.*, Vol. 15, No. 3, pp. 958—975, 1987.

[5] D. L. Donoho and I. M. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," Dept. of Statistics preprint, Stanford U., 1992, to appear in *Biometrika*.

[6] A. P. Barron and T. M. Cover, "Minimum Complexity Density Estimation," *IEEE Trans. IT*, Vol. 37, No. 4, pp. 1034—1054, 1991.

ON NONPARAMETRIC CURVE ESTIMATION WITH COMPRESSED DATA

M.Pawlak[1] and U.Stadtmüller

Dept. Elect. & Comp. Eng., University of Manitoba, Winnipeg, Canada
Abteilung für Mathematik III, University of Ulm, Ulm, Germany

Abstract — **Modified kernel estimators calculated from compressed data for density estimation and signal recovering problems are proposed. An asymptotically optimal compression technique utilizing the quantile process and data binning is employed. The statistical accuracy of the introduced kernel estimators is studied, i.e., we derive mean squared error results for the closeness of the these estimators to both the true functions and the kernel estimators determined from non compressed data.**

I. INTRODUCTION

The problem of estimating a nonparametric function $f(t)$ from a finite data record has received a great deal of attention in recent years both in Information Theory as well as Statistics. Two important models for $f(t)$ include density estimation and function recovering. In the former case $f(t)$ is a density function of a random variable X, whereas in the latter situation $f(t)$ is a signal observed in the presence of noise. The estimation problem in both settings can be formulated as follows. Given a sequence of independent random variables $\{X_1, \ldots, X_n\}$ distributed as X we wish to estimate the density $f(t)$. In the second case one observes $y_j = f(t_j) + \varepsilon_j$, at points $\{t_j\}$ and wishes to recover $f(t)$. A popular nonparametric technique for recovering $f(t)$ is the kernel estimator, defined as $\hat{f}(t) = n^{-1} \sum_{i=1}^{n} K_h(t - X_i)$, where, $K_h(t) = K(t/h)/h$, K is a kernel function and h is a smoothing parameter. As for the curve recovering problem we can use $\hat{f}(t) = \sum_{i=1}^{n} y_i \Delta_i K_h(t - t_i)$, where $\Delta_i = t_i - t_{i-1}$ and $t_0 < t_1 < \cdots < t_n$ is an ordered sequence.

Under suitable conditions for the kernel function and the sequence $\{t_j\}$ it is known that if $f \in C^s(R)$ then $E(\hat{f}(t) - f(t))^2 = O(n^{-2s/(2s+1)})$ for h being selected optimally as $cn^{-1/(2s+1)}$. Nevertheless, the aforementioned estimators need $O(n)$ evaluations at each point t and this is often a prohibitive complexity. It is our aim in this paper to propose modified versions of the kernel estimators with a substantially reduced computational complexity and yet with the asymptotically optimal rate $O(n^{-2s/(2s+1)})$.

II. KERNEL ESTIMATORS FROM COMPRESSED DATA

The reduced complexity kernel estimators can be designed first by utilizing some compression techniques to the original data set followed by a binning process applied to the classicial estimators. We utilize a compression technique employing the quantile process generated by a certain density function. A question of considerable practical importance concerns the accuracy of our new kernel estimators based on such compressed data. Hence, if $\tilde{f}(t)$ denotes a kernel estimate from the compressed data then we examine how close $\tilde{f}(t)$ is to $f(t)$ and also to $\hat{f}(t)$. Hence, $\tilde{f}(t)$ can be treated as a new estimate of $f(t)$ or we could think of $\tilde{f}(t)$ as being a compressed approximation to the estimator $\hat{f}(t)$. The general form of such estimators for the density estimation problem is the following $\tilde{f}(t) = \sum_{j=1}^{N} n_j K_h(t - a_j) / \sum_{j=1}^{N} n_j$, where $a_1, \ldots, a_N$ are center points representing the partition of the data set into N clusters, $n_j = \sum_{i=1}^{n} w_j(X_i)$, where $w_j(x)$ is the weight of assigning x to the jth cluster.

For given center points $\{a_1, \ldots, a_N\}$ one would like to select N yielding the largest possible compression ratio. On the other hand, we wish to preserve the rate of convergence attained by the classical kernel estimator. These two conflicting factors allow us to select N as a function of n yielding the desire convergence rate. Recommendations concerning the choice of h and N are presented. In particular, it is shown that N can be selected as $N = cn^{s/2(2s+1)}$, $f \in C^s(R)$ for a certain rule of generation of the center points $\{a_1, \ldots, a_N\}$.

[1] Research supported by NSERC Grant A8131 and Humboldt Foundation

REFERENCES

[1] W.Greblicki and M.Pawlak, "Dynamic system identification with order statistics", IEEE Trans. Inform. Theory, 40, 1474-1489, 1994.

[2] S.Cambanis and N.L.Gerr, "A simple class of asymptotically optimal quantizers", IEEE Trans. Inform. Theory, 29, 664-676, 1983.

[3] A.Gersho and R.M.Gray, Vector Quantization and Signal Compression, Kluwer, Boston, 1992.

[4] M.Pawlak and U.Stadtmüller, "Kernel estimators with compressed data", Technical Report, 1995.

Complexity Regularization Using Data-dependent Penalties[1]

Gábor Lugosi and Andrew Nobel

G. Lugosi is with the Dept. of Mathematics, Faculty of Elect. Engineering, Technical University of Budapest, Hungary.
Email: lugosi@vma.bme.hu.

A. Nobel is with the Department of Statistics at the University of North Carolina, Chapel Hill.
Email: nobel@assistant.beckman.uiuc.edu

Abstract — **We define a regression function estimate based on complexity regularization, where the list of candidate functions and the corresponding penalties are determined from the training data, leading to improved performance.**

Let (X,Y) be an $\mathcal{R}^d \times \mathcal{R}$-valued pair of random variables with regression function $m(x) = \mathbf{E}\{Y|X = x\}$. We assume that Y (and therefore also $m(X)$) is bounded with probability one, i.e., $\mathbf{P}\{|Y| \le B\} = 1$ for some $B < \infty$.

The regression function $m(x)$ is to be estimated based upon the data $T_n = ((X_1,Y_1),\ldots,(X_n,Y_n))$, where the (X_i,Y_i) pairs are independent copies of (X,Y). A regression function estimate is thus a function $m_n : \mathcal{R}^d \times \left(\mathcal{R}^d \times \mathcal{R}\right)^n \to \mathcal{R}$, whose performance is measured by the squared error

$$\begin{aligned} J(m_n) &= \mathbf{E}\left\{(m_n(X,T_n) - Y)^2 \,|T_n\right\} - \mathbf{E}\left\{(m(X) - Y)^2\right\} \\ &= \int (m_n(x,T_n) - m(x))^2 \mu(dx), \end{aligned}$$

where μ denotes the distribution of X. We will use the shorthand notation $m_n(x)$ instead of $m_n(x,T_n)$.

Complexity regularization (see Barron and Cover [2], Barron [1]) selects an estimate m_n from a countable list of candidates Γ_n by minimizing the sum of the empirical error

$$\hat{J}_n(f) = \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - Y_i)^2$$

and an appropriately defined complexity penalty $\Delta_n(f)$ over $f \in \Gamma_n$. Intuitively, more "complex" candidates are penalized by more in order to avoid overfitting. Barron [1] proves that if the best candidate in Γ_n is close to m, then the method indeed performs extremely well. In particular,

$$J(m_n) = O\left(\inf_{f\in\Gamma_n}\left(\frac{\Delta_n(f)}{n} + J(f)\right)\right),$$

where the penalties $\Delta_n(f)$ are required to satisfy the summability constraint $\sum_{f\in\Gamma_n} 2^{-\Delta_n(f)} < \infty$ for each $n \ge 1$.

Our goal is to assure that the list of candidates contains elements—with not too large complexity penalties—that closely approximate the regression function. We let the data determine the list of functions which adds a tremendous amount of flexibility, leading to an improved performance.

The basic idea is splitting the data in two such that the first half

$$T_n^1 = ((X_1,Y_1),\ldots,(X_m,Y_m))$$

is used to determine the (random) list of functions Γ_n and the corresponding penalties, and the second half

$$T_n^2 = ((X_{m+1},Y_{m+1}),\ldots,(X_n,Y_n))$$

is used to carry out minimization of the empirical error. For convenience, we take $m = \lfloor n/2 \rfloor$, but other choices as $m \approx \sqrt{n}$ may be better in certain cases.

As the first step towards defining our estimate, consider a sequence $\Lambda_1, \Lambda_2, \ldots$ of classes of bounded functions $\mathcal{R}^d \to [-B,B]$. These classes may be uncountable, but to avoid certain problems of measurability, we assume that every model class Λ_i contains a countable subclass Λ_i^* with the property that every $f \in \Lambda_i$ is a pointwise limit of a sequence of functions from Λ_i^*.

Following ideas of Buescher and Kumar [3], we construct a proper minimal empirical cover of each class Λ_i based upon the data $X_1^m = X_1,\ldots,X_m$, i.e., for each i, we take a set $\mathcal{G}_i \in \Lambda_i$ with the following properties. For every $f \in \Lambda_i$ there exists $g \in \mathcal{G}_i$ such that

$$\frac{1}{m}\sum_{j=1}^{m}|f(X_i) - g(X_j)| \le \frac{1}{\sqrt{n}},$$

and $\mathcal{G}_i$ has minimal cardinality. Denote it by $|\mathcal{G}_i| = N(X_1^m, \Lambda_i)$.

To each $f \in G_i$, we assign the complexity penalty

$$\Delta_i(f, X_1^m) = B^2\frac{\log N(X_1^m,\Lambda_i) + c_i}{n},$$

where the c_i's are required to satisfy the Kraft-type inequality $\sum_{i=1}^{\infty} e^{-c_i} \le 1$. Define the estimate m_n as a function that minimizes the penalized empirical error

$$\frac{1}{n-m}\sum_{j=m+1}^{n}(f(X_j) - Y_j)^2 + \Delta_i(f, X_i^m)$$

over all $f \in \mathcal{G} = \cup_{i=1}^{\infty}\mathcal{G}_i$.

For the performance of the estimate we have the following result.

Theorem 1 *For a universal constant C,*

$$\mathbf{E}\{J(m_n)\} \le C\inf_{i\ge 1}\left(\frac{\log \mathbf{E}N(X_1^m,\Lambda_i) + c_j}{n} + \inf_{f\in\Lambda_i} J(f)\right).$$

The main improvement in our result is that in the second term, the infimum is taken over an uncountable collection.

References

[1] A. R. Barron. Complexity regularization with application to artificial neural networks. In G. Roussas, editor, *Nonparametric Functional Estimation and Related Topics*, pages 561–576, Dordrecht, 1991. NATO ASI Series, Kluwer Academic Publishers.

[2] A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Transactions on Information Theory*, 37:1034 – 1054, 1991.

[3] K. L. Buescher and P. R. Kumar. Learning by canonical smooth estimation, Part I: Simultanious estimation. *submitted to IEEE Transactions on Automatic Control*, 1994.

[1]The research was supported in part by the National Science Foundation under Grants No. NCR-92-96231 and INT-93-15271.

On the Existence of Strongly Consistent Rules for Estimation and Classification*

S.R. Kulkarni

Department of Electrical Engineering
Princeton University, Princeton, NJ 08544
email: kulkarni@ee.princeton.edu

O. Zeitouni

Department of Electrical Engineering
Technion, Haifa 32000, ISRAEL
email: zeitouni@ee.technion.ac.il

Abstract — **Suppose we observe $x_1, x_2, \ldots \in \mathcal{X}$ drawn i.i.d. according to some unknown distribution P selected from a family of distributions $\mathcal{P}$. Let $f : \mathcal{P} \to \Lambda$ be a parameterization of $P \in \mathcal{P}$. In this paper, we study necessary and sufficient conditions for the existence of strongly consistent estimators of $f(P)$. A number of previous results along these lines are special cases of our main result.**

I. Introduction and Formulation

This paper is concerned with characterizing when strongly consistent estimators exist for hypothesis testing and estimation problems such as those posed in the abstract. Our interest in this problem stems from the work of Cover [1] and subsequent work [5, 8, 6, 2, 4] that significantly generalized the set-up in [1]. This paper presents an extension and alternative approach to recent results of Dembo and Peres [2]. The flavor of the results is also along the lines of the classical work of Hoeffding and Wolfowitz [3] and LeCam and Schwartz [7].

Let $\mathcal{X}$ (the *sample space*) be a complete, separable metric space (e.g., think of $\mathbb{R}^m$), and let $\mathcal{M}_1(\mathcal{X})$ denote the space of all Borel probability measures on $\mathcal{X}$. Let (Λ, d) denote another metric space (the *parameter space*, with d denoting the metric), which is σ-compact. Again, one can think of $\Lambda = \mathbb{R}^m$ as a characteristic example. Let $f : \mathcal{P} \subset \mathcal{M}_1(\mathcal{X}) \to \Lambda$ denote a Borel measurable map. Thus, f denotes a parameterization of the probability measures in $\mathcal{P}$, with $f(P)$ being the parameter associated with P. We are interested in the estimation of $f(P)$, based on a sequence of i.i.d. observations $\{x_i\}_{i=1}^{\infty}$ with marginal distribution $P \in \mathcal{P}$.

Definition 1 (Discernibility and Estimation)
a) $A_0, A_1, \ldots, A_k \subset \mathcal{P}$ are discernible *if there exists a strongly consistent decision rule for deciding to which A_i an unknown $P \in \cup_{i=0}^{k} A_i$ belongs — that is, if there exists a sequence of Borel functions $g^n : \mathcal{X}^n \to \{0, 1, \ldots k\}$ such that, for any $P \in A_i$, almost surely $g^n(x_1, \ldots, x_n) \to_{n\to\infty} i$. b) Λ is* f-estimatable *if there exists a strongly consistent estimator for $f(P)$. That is, if there exists a sequence of Borel functions $g^n : \mathcal{X}^n \to \Lambda$ such that, for all $P \in \mathcal{P}$, almost surely $g^n(x_1, \ldots, x_n) \to_{n\to\infty} f(P)$.*

We use the term discernibility following [2]. We will also use notions of *uniform* discernibility and estimation in which there is a uniformly consistent estimator over all $P \in \mathcal{P}$.

II. Main Results

We first mention a result showing that it is enough to characterize discernibility to obtain results on the more general estimation problem. Namely, we have shown that Λ is f-estimatable iff for all $\Lambda_1, \ldots, \Lambda_k \subset \Lambda$ that are positively separated, the sets $f^{-1}(\Lambda_1), \ldots, f^{-1}(\Lambda_k)$ are discernible. We thus concentrate below only on the notion of discernibility.

We have also shown that $A_0, A_1, \ldots, A_k \subset \mathcal{P}$ are discernible iff there exist sequences $A_i^n \nearrow A_i$ for $i = 0, \ldots, k$ such that $(A_0^n, \ldots, A_k^n)$ are uniformly discernible for each n. Using this result, our approach is to make some quite general assumptions regarding separation properties of the A_i and *uniform* discernibility. The idea is that conditions regarding uniform discernibility are easier to check, but still lead to statements on (non-uniform) consistency.

We now assume that $\mathcal{M}_1(\mathcal{X})$ is endowed with some metric ρ. For $A, B \subset \mathcal{P}$, define the separation between A and B as $\rho(A, B) = \inf_{P_1 \in A, P_2 \in B} \rho(P_1, P_2)$. A, B are called *positively separated* if $\rho(A, B) > 0$. Sets $A_0, \ldots, A_k$ are said to be *separable by closed sets* if there exist closed sets $B_0, \ldots, B_k$ with $A_i \subset B_i$ and $B_i \cap B_{i'} = \emptyset$ for $i \neq i'$. We say that $A_0, \ldots, A_k$ are positively separated by finite covers if each A_i can be covered by a finite number of closed balls, and the covers are pairwise positively separated. We recall that in a topological space a set A is said to be F_σ if it is a countable union of closed sets, that is, if $A = \cup_{i=1}^{\infty} F_i$ where the F_i are closed. $A_0, A_1, \ldots, A_k$ are called *F_σ-separated* if $A_i \subset B_i$, where the B_i are F_σ sets with $B_i \cap B_{i'} = \emptyset$ for all $i \neq i'$.

Definition 2 (Properness Conditions on ρ)
(P1) ρ is said to be (P1)-proper w.r.t. $\mathcal{P}$ if every Borel $A_i \subset \mathcal{P}$ which are uniformly discernible are separable by closed sets.
(P2) ρ is said to be (P2)-proper w.r.t. $\mathcal{P}$ if ρ makes $\mathcal{M}_1(\mathcal{X}) \cap \mathcal{P}$ into a separable space, and any Borel sets $A_i \subset \mathcal{P}$ which are positively separated by finite covers are uniformly discernible.

In a sense, these conditions impose "non-degeneracy" requirements on the metric ρ restricted to $\mathcal{P}$. The following result extends and corrects Theorem 3.3 of [6]. Results from [1], [5], and Theorem 2 of [2] can be recovered using Theorem 1 below. We denote $\mathcal{A} = \cup_{i=0}^{k} A_i$.

Theorem 1
a) (sufficient condition) Assume ρ is (P2)-proper w.r.t. $\mathcal{A}$. If $A_0, \ldots, A_k$ are F_σ separated then they are discernible.
b) (necessary condition) Assume ρ is (P1)-proper w.r.t. $\mathcal{A}$. If $A_0, \ldots, A_k$ are discernible then they are F_σ-separated.

*This work was supported in part by the National Science Foundation under grants IRI-9209577 and IRI-9457645 and by the U.S. Army Research Office under grants DAAL03-92-G-0320 and DAAL03-92-G-0115.

References

[1] T.M. Cover, "On determining the irrationality of the mean of a random variable," *Ann. Stat.*, Vol. 1, pp. 862-871, 1973.

[2] A. Dembo and Y. Peres, "A topological criterion for hypothesis testing," *Ann. Stat.*, Vol. 22, pp. 106-117, 1994.

[3] W. Hoeffding and J. Wolfowitz, "Distinguishability of sets of distributions," *Ann. Math. Stat.*, Vol. 29, pp. 700-718, 1958.

[4] S.R. Kulkarni and D.N.C. Tse, "A paradigm for class identification problems," *IEEE Trans. on Information Theory*, Vol. 40, pp. 696-705, 1994.

[5] S.R. Kulkarni and O. Zeitouni, "Can one decide the type of the mean from the empirical measure?" *Statistics & Probability Letters*, Vol. 12, pp. 323-327, 1991.

[6] S.R. Kulkarni and O. Zeitouni, "On probably correct classification of concepts," *Proc. Sixth Annual ACM Workshop on Computational Learning Theory*, pp. 111-116, 1993.

[7] L. LeCam and L. Schwartz, "A necessary and sufficient condition for the existence of consistent estimates," *Ann. Math. Stat.*, Vol. 31, pp. 130-140, 1960.

[8] O. Zeitouni and S.R. Kulkarni, "A general classification rule for probability measures," to appear *Ann. Stat.*, 1995.

k Nearest Neighbors in Search of a Metric

Robert R. Snapp[1]
Computer Science and Electrical Engineering Department
University of Vermont
Burlington, VT 05405 USA
snapp@emba.uvm.edu

Santosh S. Venkatesh
Department of Electrical Engineering
University of Pennsylvania
Philadelphia, PA 19104 USA
venkates@ee.upenn.edu

Abstract — **The finite-sample risk of the k-nearest neighbor classifier that uses a weighted L_p metric as a measure of class similarity is examined. For a family of multiclass, classification problems with smooth distributions in $\mathbb{R}^n$, the risk is represented as an asymptotic expansion in decreasing fractional powers of the reference-sample size. An analysis of the leading coefficients reveals that the optimal metric (i.e., the metric that minimizes the risk) tends to a weighted Euclidean (i.e., L_2) metric as the sample size is increased. Numerical calculations corroborate this finding.**

I. The k-Nearest-Neighbor Classifier

Let the elements of $\mathbb{L} = \{1,\ldots,C\}$ denote C states of nature, or pattern classes, and let $P_1,\ldots,P_C$ denote their corresponding stationary prior probabilities. Each pattern is represented by a feature vector $\mathbf{x}$, drawn at random from $\mathbb{R}^n$. Specifically, patterns originating from class $\ell \in \mathbb{L}$ are generated by the stationary conditional distribution F_ℓ.

Labeled feature vectors are generated by a two-step process. First, a class $\ell \in \mathbb{L}$ is chosen at random so that $\Pr[\ell = j] = P_j$; then a random feature vector is drawn according to F_ℓ. After m independent repetitions of this process, we obtain the labeled reference sample,

$$\mathcal{X}_m = \{(\mathbf{x}^1,\ell^1),\ldots,(\mathbf{x}^m,\ell^m)\}.$$

Given a weighted L_p metric, $d(\mathbf{x},\mathbf{y}) = \|A(\mathbf{x}-\mathbf{y})\|_p$, where A denotes an n-by-n, positive-definite, symmetric matrix with $\det A = 1$, and an arbitrary point $\mathbf{x} \in \mathbb{R}^n$, the indices of the labeled feature vectors in $\mathcal{X}_m$ can be permuted so that

$$d(\mathbf{x},\mathbf{x}^1) \le d(\mathbf{x},\mathbf{x}^2) \le \cdots \le d(\mathbf{x},\mathbf{x}^m). \qquad (1)$$

Here $\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$ for $1 \le p < \infty$, and $\|\mathbf{x}\|_\infty = \max_{1\le i\le n}|x_i|$, denote the L_p norm. The k nearest neighbors of $\mathbf{x}$ then form the subset $\{(\mathbf{x}^1,\ell^1),\ldots,(\mathbf{x}^k,\ell^k)\}$; and the k-nearest-neighbor classifier assigns $\mathbf{x}$ to class $L'(\mathbf{x}) = \mathrm{maj}(\ell^1,\ldots,\ell^k)$, viz., the most frequently appearing class label in the subset. (Ties, and degeneracies in (1), can be resolved by an arbitrary procedure.) Using this algorithm every point in $\mathbb{R}^n$ can be assigned to a class in $\mathbb{L}$.

II. The Finite-Sample Risk

Given a positive integer k, an L_p metric, and a finite random reference sample $\mathcal{X}_m$, a single test vector $(\mathbf{X},L)$, drawn independently by the same random process, is assigned to class $L' = L'(\mathbf{X})$ by the k-nearest-neighbor classifier. We now consider the m-sample risk,

$$R_m = \sum_{i=1}^{C}\sum_{j=1}^{C} \Lambda_{ij}\Pr[L' = i, L = j],$$

with the zero-one cost matrix $\Lambda_{ij} = 1 - \delta_{ij}$.

For a family of classification problems, $\mathcal{F}_N$, described by class-conditional probability densities f_ℓ with uniformly bounded partial derivative up through order $N+1$, and a mixture density $f = \sum_{\ell=1}^{C} P_\ell f_\ell$ that is bounded away from zero a.e. on its probability-one support $S \subset \mathbb{R}^n$, we obtain the following:

Theorem 1 *There exist constants c_j, for $j = 2,3,\ldots,N$, such that*

$$R_m = R_\infty + \sum_{j=2}^{N} c_j m^{-j/n} + O(m^{-(N+1)/n})$$

where R_∞ is the infinite-sample risk derived by Cover and Hart [1].

(A version of this theorem, restricted to the case $k = 1$, $p = 2$, $A = I$, and $C = 2$, appears in a recent paper [2].) The coefficient c_2 evaluates to

$$c_2 = D_n(p)\,\frac{\Gamma(k+1+\frac{2}{n})}{24\left[\Gamma\left(\frac{k+1}{2}\right)\right]^2}\,\mathrm{tr}\left\{(A^{-1})^T H A^{-1}\right\},$$

where,

$$D_n(p) = \frac{\Gamma\left(\frac{3}{p}+1\right)\Gamma\left(\frac{n}{p}+1\right)^{1+(2/n)}}{\Gamma\left(\frac{n+2}{p}+1\right)\Gamma\left(\frac{1}{p}+1\right)^3},$$

A^{-1} denotes the inverse of the metric weight matrix A, and H is an n-by-n matrix, independent of p. For the two-class problem ($C = 2$),

$$H_{ij} = \int_S d\mathbf{x}\, f^{1-\frac{2}{n}}(\hat{P}_1\hat{P}_2)^{\frac{k+1}{2}}(\hat{P}_2 - \hat{P}_1)\left(\frac{1}{f_1}\frac{\partial^2 f_1}{\partial x_i \partial x_j} - \frac{1}{f_2}\frac{\partial^2 f_2}{\partial x_i \partial x_j}\right).$$

Here, $\hat{P}_\ell = P_\ell f_\ell(\mathbf{x})/f(\mathbf{x})$ denotes the posterior probability that a feature vector with value $\mathbf{x}$ originates from class ℓ.

III. A Desirable Metric

Since R_∞ does not depend upon the chosen metric, Theorem 1 suggests that the finite-sample risk of the k-nearest-neighbor may be reduced, for large values of m, by selecting a metric that minimizes c_2. It can be shown that $D_n(p)$ has a global minimum at $p = 2$ for fixed $n > 1$. Using the Euler-Lagrange multiplier theorem, the trace in c_2 is minimized if the weight matrix A satisfies $A^T A = H/(\det H)^{1/n}$. Although it may be difficult to determine H, and consequently the optimal matrix A, in practice, this analysis and corroborating numerical simulations motivate the use of a weighted Euclidean metric for large reference samples.

[1] This work was supported in part by Rome Laboratory, Air Force Material Command, USAF, under grant number F30602-94-1-0010.

References

[1] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 21-27, 1967.

[2] D. Psaltis, R. R. Snapp, and S. S. Venkatesh, "On the finite sample performance of the nearest neighbor classifier," *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 820-837, 1994.

An Information-theoretic Framework for Optimization with Application to Supervised Learning [1]

David Miller, Ajit Rao, Kenneth Rose, and Allen Gersho
Center for Information Processing Research
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

This work develops a unified approach for hard optimization problems involving data association, i.e. the assignment of elements viewed as "data", $\{x_i\}$, to one of a set of classes, $\{C_j\}$, so as to minimize the resulting cost. The diverse problems which fit this description include data clustering, statistical classifier design to minimize probability of error, piecewise regression, structured vector quantization, as well as optimization problems in graph theory, e.g. graph partitioning. Whereas standard descent-based methods are susceptible to finding poor local optima of the cost, the suggested approach provides some potential for avoiding local optima, yet without the computational complexity of stochastic annealing.

The approach we develop is based on ideas from information theory and statistical physics, and builds on the work of Rose, Gurewitz, and Fox for clustering and related problems [1]. The optimization problem is embedded within a framework in which data are assigned to classes *in probability*, with Shannon's entropy measure used to control the level of uncertainty or randomness in the assignments. We first address "unconstrained" assignment problems such as data clustering and graph partitioning, in which the data elements are freely assigned to any class, specifiable by binary 0-1 assignment variables. We consider the joint distribution over all possible assignments, $P[x_1 \in C_{j(1)}, \ldots, x_N \in C_{j(N)}]$, and choose it to minimize the expected assignment cost $< E >$, given a constraint on Shannon's entropy, H. Thus, we seek the *best* random assignments in the sense of $< E >$ for a given H. This formulation is equivalently stated by invoking the maximum entropy principle, but the former description is more appealing for optimization. The constrained minimization is equivalent to the unconstrained minimization of the Lagrangian: $L \equiv \beta < E > -H$, where β is the Lagrange multiplier controlling $< E >$ and H. Physical inspiration for minimizing L is obtained by recognizing that it is the Helmholtz free energy of a simulated system, with $< E >$ the "energy" and $\frac{1}{\beta}$ the "temperature". Thus, a deterministic annealing approach is naturally suggested, wherein, starting from high temperature ($\beta = 0$), the cost and randomness are reduced with the temperature. At low temperature ($\beta \to \infty$) the hard cost is minimized. Our formulation unifies the deterministic annealing method for clustering with mean-field annealing methods proposed for combinatorial optimization [2]. Moreover, the derivation provides an intuitive, yet precise description of what constitutes annealing in these optimization methods. In particular, the annealing process is characterized as a reduction in the system's entropy and expected cost through the increase of a Lagrange multiplier interpreted as the inverse temperature.

While this description may provide insights into existing methods, a more significant benefit lies in its generality, and hence its potential for stimulating development of novel optimization methods tackling heretofore unaddressed assignment problems. Of prime interest are what we will call *structurally-constrained* problems, wherein the assignments are restricted to be consistent with a (parametrized) classification rule. These problems abound in pattern recognition and source coding, and include statistical classifier design, piecewise regression, and structured vector quantization. The restricted assignments may be produced by a nearest prototype rule, a decision tree, or neural network structures such as radial basis functions or multilayer perceptrons. Thus, the previous optimization framework requires substantial extension in order to enforce the structural constraint on the assignments. To do so, we introduce an additional cost C_s, which *quantifies* achievement of the structural constraint. This cost is incorporated within a generalization of the basic formulation we have described, so that the annealing process controls $< C_s >$, as well as $< E >$ and H. A second Lagrange multiplier is identified which controls $< C_s >$. This parameter is chosen to provide the optimal "level" of structural constraint consistent with $< E >$ and H at each temperature in the annealing process. At the limit $\beta \to \infty$, a "hard" classifier with the requisite structure is achieved, and the assignment cost is minimized directly. This general optimization paradigm has significant potential for outperforming descent-based approaches for structurally-constrained assignment problems. In several coming papers, these ideas are applied to the two fundamental problems of supervised learning – statistical classification and regression – as well as to the design of novel source coding structures (generalized vector quantizers), with promising results achieved in all of these domains [3], [4].

[1] This work was supported in part by the National Science Foundation under grant no. NCR-9314335, the University of California MICRO program, Rockwell International Corporation, Hughes Aircraft Company, Echo Speech Corporation, Signal Technology Inc., Lockheed Missile and Space Company and Qualcomm, Inc.

References

[1] K. Rose, E. Gurewitz, and G. C. Fox, "Vector quantization by deterministic annealing," *IEEE Trans. on Inform. Theory*, vol. 38, pp. 1249–1258, 1992.

[2] G. L. Bilbro, W. E. Snyder, and R. C. Mann, "Mean-field approximation minimizes relative entropy," *Journal of the Opt. Soc. of Amer.*, vol. 8, pp. 290–294, 1991.

[3] D. Miller, A. Rao, K. Rose, and A. Gersho, "An information-theoretic framework for optimal statistical classification." (Submitted for publication.), 1995.

[4] A. Rao, D. Miller, K. Rose, and A. Gersho, "Generalized vector quantization." (To be submitted for publication.), 1995.

Nonparametric Classification using Radial Basis Function Nets and Empirical Risk Minimization

Adam Krzyżak[1], Tamás Linder[2], and Gábor Lugosi[2]
1. Dept. Comp. Sc., Concordia University, Montreal, Canada
2. Dept. Math. & Comp. Sc., Technical University of Budapest, Hungary

Abstract — **In the paper we study convergence properties of radial basis function (RBF) networks in nonparametric classification for a large class of basis functions with parameters of RBF nets learned through empirical risk minimization.**

I. Introduction

In the classification (pattern recognition) problem, based upon the observation of a random vector $X \in \mathbb{R}^d$, one has to guess the value of a corresponding label Y, where Y is a random variable taking its values from $\{-1, 1\}$. The decision is a function $g : \mathbb{R}^d \to \{-1, 1\}$, whose goodness is measured by the *error probability* $L(g) = \mathbf{P}\{g(X) \neq Y\}$. Let g^* be a *Bayes decision* and L^* be the *Bayes risk*. An empirical decision rule g_n is a function $g_n : \mathbb{R}^d \times (\mathbb{R}^d \times \{-1, 1\})^n \to \{-1, 1\}$, whose error probability is given by

$$L(g_n) = \mathbf{P}\{g_n(X, D_n) \neq Y | D_n\}$$

where $D_n = ((X_1, Y_1), \ldots, (X_n, Y_n))$ is a training sequence of i.i.d. random variables independent of (X, Y). A sequence of classifiers $\{g_n\}$ is called *strongly consistent* if $\lim_{n\to\infty}(L(G_n) - L^*) = 0$ almost surely, and $\{g_n\}$ is *strongly universally consistent* if it is consistent for *any* distribution of (X, Y).

Let $K : \mathbb{R}^d \to \mathbb{R}$ be a kernel function. Consider RBF networks given by

$$f_\theta(x) = \sum_{i=1}^{k} w_i K(A_i[x - c_i]) + w_0 \qquad (1)$$

where $\theta = (w_0, \ldots, w_k, c_1, \ldots, c_k, A_1, \ldots, A_k)$ is the vector of parameters, $w_0, \ldots, w_k \in \mathbb{R}$, $c_1, \ldots, c_k \in \mathbb{R}^d$, and $A_1, \ldots, A_k$ are nonsingular $d \times d$ matrices. Let $\{k_n\}$ be a sequence of positive integers. Define $\mathcal{F}_n$ as the set of RBF networks in the form of (1) with $k = k_n$. Given an f_θ as above, we define the classifier g_θ to be 1 if $f_\theta(x) \geq 0$, and 0 otherwise.

Let $\mathcal{G}_n$ be the class of classifiers based on the class of functions $\mathcal{F}_n$. To every classifier $g \in \mathcal{G}_n$, assign the *empirical error probability*

$$\hat{L}_n(g) = \frac{1}{n} \sum_{i=1}^{n} I_{\{g(X_i) \neq Y_i\}}.$$

We pick a classifier g_n from $\mathcal{G}_n$ by minimizing this empirical error probability. The distance $L(g_n) - L^*$ between the error probability of the selected rule and the Bayes risk is decomposed into the *estimation error* and the *approximation error*:

$$L(g_n) - L^* = \left(L(g_n) - \inf_{g\in\mathcal{G}_n} L(g)\right) + \left(\inf_{g\in\mathcal{G}_n} L(g) - L^*\right).$$

[0]This work was supported by Canadian National Networks of Centers of Excellence grant 293

II. Approximation

We consider the approximation error when $\mathcal{F}_k$ is the family of RBF networks of the form of (1).

Theorem 1 . *Suppose $K : \mathbb{R}^d \to \mathbb{R}$ is bounded and*

$$K \in L_1(\lambda) \cap L_p(\lambda)$$

for some $p \in [1, \infty)$, and assume that $\int K(x)dx \neq 0$. Let μ be an arbitrary probability measure on $\mathbb{R}^d$ and let $q \in (0, \infty)$. Then the RBF nets in the form (1) are dense in both $L_q(\mu)$ and $L_p(\lambda)$. In particular, if $m \in L_q(\mu) \cap L_p(\lambda)$, then for any ϵ there exists a $\theta = (w_0, \ldots, w_k, b_1, \ldots, b_k, c_1, \ldots, c_k)$ such that

$$\int_{\mathbb{R}^d} |f_\theta(x) - m(x)|^q \mu(dx) < \epsilon \quad \text{and} \quad \int_{\mathbb{R}^d} |f_\theta(x) - m(x)|^p dx < \epsilon.$$

III. CLASSIFICATION

Define the class of sets

$$\mathcal{C}_1 = \{\{x \in \mathbb{R}^d : K(A[x - c]) > 0\} : c \in \mathbb{R}^d, A \text{ invertible}\}.$$

Theorem 2 *Let K be an indicator such that $V_{\mathcal{C}_1} < \infty$. Then for every n, k_n and $\epsilon > 0$,*

$$\mathbf{P}\{L(g_n) - \inf_{g\in\mathcal{G}_n} L(g) > \epsilon\} \leq 4 \exp\left(-n\left[\frac{\epsilon^2}{32} - \frac{C_2 k_n \log(C_1 n)}{n}\right]\right)$$

for some constants C_1 and C_2 depending only on $V_{\mathcal{C}_1}$.

Suppose that the set of RBF networks given by (1), k being arbitrary, is dense in $L_1(\mu)$ on balls $\{x \in \mathbb{R}^d : \|x\| \leq B\}$ for any probability measure μ on $\mathbb{R}^d$. If $k_n \to \infty$ and $n^{-1}(k_n \log n) \to 0$ as $n \to \infty$, then the sequence of classifiers g_n minimizing the empirical error probability is strongly universally consistent.

Using a result by Macintyre and Sontag we can also prove that empirical error probability minimization can fail to provide a distribution free upper bound on $L(g_n) - \inf_{g\in\mathcal{G}_n} L(g)$ for other kernels however nice these kernels may seem. We have the following counter example:

Theorem 3 *There exists a $K : \mathbb{R} \to \mathbb{R}$ which is symmetric around 0, monotone decreasing on $\mathbb{R}^+$ and infinitely differentiable, such that for $k \geq 2$ there is no distribution free upper bound on the estimation error which converges to zero as $n \to \infty$.*

References

[1] A.Krzyżak, T. Linder and G. Lugosi. Nonparametric estimation and classification using radial basis function nets and empirical risk minimization. *IEEE Trans. on Neural Networks*, 1995 (to appear).

Universal, Nonlinear, Mean-Square Prediction of Markov Processes[1]

Dharmendra S. Modha and Elias Masry

Department of Electrical & Computer Engineering, University of California at San Diego
9500 Gilman Drive, La Jolla, CA 92093-0407, USA

Abstract — **We estimate the best, nonlinear, mean-square predictor for a Markov process from an observed, finite realization of the process—when the true Markov order is unknown. In particular, we propose an universal minimum complexity estimator, which does not know the true Markov order, and yet delivers the same statistical performance as that delivered by a minimum complexity estimator, which knows the true Markov order.**

I. Introduction

Given a sequence of observations $\{X_i\}_{i=1}^N$ drawn from a stationary Markov process $\{X_i\}_{i=-\infty}^{\infty}$ of order q, we are interested in estimating the conditional mean of X_0 given the past $X_{-1}, X_{-2}, \ldots, X_{-q}$, namely

$$m_q(X_{-1}, X_{-2}, \ldots, X_{-q}) = E[X_0 | X_{-1}, X_{-2}, \ldots, X_{-q}].$$

The conditional mean m_q is the best, nonlinear, mean-square predictor for the Markov process $\{X_i\}_{i=-\infty}^{\infty}$, and is thus an important object of knowledge.

In addition to the Markov assumption, we assume that $\{X_i\}_{i=-\infty}^{\infty}$ is strongly mixing [5] with exponential decay [2]. This additional assumption is required, technically, to construct consistent minimum complexity estimators for m_q.

If the true Markov order q is **known**, then we can estimate the predictor m_q by proceeding essentially as in Modha and Masry [2]. In particular, if we assume that the predictor m_q possesses a certain bounded spectral norm [1], then it is possible to construct a minimum complexity estimator, say $\hat{m}_{q,N}$, based on neural networks such that [2]

$$\text{MISE}(\hat{m}_{q,N}, m_q) = O\left((\log N)^{\frac{1}{2}} N^{-\frac{1}{4}}\right), \qquad (1)$$

where MISE denotes a certain mean integrated squared error.

In this paper, we consider the practically important case when the true Markov order q is **unknown**. In particular, assuming, as before, that the predictor m_q possesses a bounded spectral norm, we propose (see Section II below) a universal minimum complexity estimator, say $\hat{m}_N$, based on neural networks such that [3]

$$\text{MISE}(\hat{m}_N, m_q) = O\left((\log N)^{\frac{1}{2}} N^{-\frac{1}{4}}\right). \qquad (2)$$

Precise results and proofs can be found in the full paper [3].

Comparing (1) and (2), we find that our estimator $\hat{m}_N$, which does not know q, achieves the **same** rate of convergence as that of $\hat{m}_{q,N}$, which knows q. Asymptotically, the estimator $\hat{m}_N$ not only learns the true predictor m_q, but also (implicitly) discovers the true Markov order q. In other words, the estimator $\hat{m}_N$ is **universal**. This notion of universality parallels the notions of universality arising in the context of coding of finite alphabet processes and in the context of mean-square prediction of Gaussian ARMA processes [4].

[1]This work was supported by the Office of Naval Research under Grant N00014-90-J-1175

II. Universal Estimation Scheme

We now outline a two-stage estimation scheme to construct the universal minimum complexity estimator $\hat{m}_N$, which was advertized and discussed in the previous section. Our estimation scheme, which can be found in the full paper [3], builds on the results in Modha and Masry [2], and is inspired by the results in Barron [1] and in Rissanen [4].

Stage 1: For each fixed memory $1 \le p \le \log N$, let m_p denote the conditional mean of X_0 given the past $X_{-1}, X_{-2}, \ldots, X_{-p}$. Given N observations $\{X_i\}_{i=1}^N$, we first estimate m_p (for each $1 \le p \le \log N$) using neural networks as follows.

A neural network parametrized by a very small number of parameters has a small variance (estimation error), but also has a large bias (approximation error) in estimating m_p; on the other hand, a neural network parametrized by a very large number of parameters has a small bias, but also has a large variance in estimating m_p. The minimum complexity estimator $\hat{m}_{p,N}$, which minimizes a certain penalized empirical loss, selects the neural network (parametrized by an appropriate number of parameters) that achieves the best trade-off between the bias and variance. Thus, $\hat{m}_{p,N}$ achieves the smallest statistical risk (bias + variance) in estimating m_p.

Stage 2: Having constructed the sequence of minimum complexity estimators $\{\hat{m}_{p,N}\}_{p=1}^{\log N}$, we now select the estimator $\hat{m}_N$ as the element of the sequence that achieves the smallest statistical risk in estimating m_q.

In particular, for a very small p, $\hat{m}_{p,N}$ may be close to m_p, but m_p may be far from m_q; on the other hand, for a very large p, m_p may be close to m_q, but $\hat{m}_{p,N}$ may be far from m_p. The universal minimum complexity scheme selects a data-driven memory $\hat{p}$, which minimizes a certain penalized empirical loss and hence achieves the best trade-off between the competing terms. Finally, we use the element corresponding to $\hat{p}$, namely $\hat{m}_{\hat{p},N}$, as our universal estimator $\hat{m}_N$.

References

[1] A. R. Barron, "Approximation and estimation bounds for artificial neural networks," *Machine Learning*, vol. 14, pp. 115-133, 1994.

[2] D. S. Modha and E. Masry, "Minimum complexity regression estimation with weakly dependent observations," submitted for publication, 1994.

[3] D. S. Modha and E. Masry, "Universal, nonlinear, mean-square estimation for stationary time series," submitted for publication, 1995.

[4] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, Teaneck, NJ: World Scientific Publishers, 1989.

[5] M. Rosenblatt, "A central limit theorem and strong mixing conditions," *Proc. Nat. Acad. Sci.*, vol. 4, pp. 43-47, 1956.

The Quadratic Gaussian CEO Problem

Harish Viswanathan and Toby Berger
School of Electrical Engineering, Cornell University, Ithaca, NY 14850 , USA

I. Introduction

The following problem in multiterminal source coding was introduced in [1]. A firm's CEO is interested in a data sequence $\{X(t)\}_{t=1}^{\infty}$ which cannot be observed directly. The CEO employs a team of L agents who observe independently corrupted versions of $\{X(t)\}_{t=1}^{\infty}$. Let R be the total data rate at which the agents may communicate information about their observations to the CEO. The agents are not allowed to convene. In [1] Berger and Zhang determine the asymptotic behavior of the minimal error frequency in the limit as L and R tend to infinity. Their result is for discrete memoryless source and observations. In this paper we consider a special case of the continuous source and observations problem. We assume that the source is an i.i.d sequence of zero mean Gaussian random variables ($\mathcal{N}(0,\sigma_X^2)$) and the observations are corrupted by identical independent memoryless Gaussian noise ($\mathcal{N}(0,\sigma_N^2)$). The CEO is interested in reconstructing the source with minimum mean squared error. We study the asymptotic behavior of the minimum achievable distortion in the limit as first L and then R tends to infinity. That is we study the behavior of

$$\beta(\sigma_X^2,\sigma_N^2) \quad = \quad \lim_{R\to\infty}\lim_{L\to\infty} R\frac{D(R,L)}{\sigma_X^2}.$$

The solution to this problem differs sufficiently from that for the discrete source problem studied in [1]. Our main result is that asymptotically the distortion decays at best as $1/R$. We also derive the upper bound $\beta(\sigma_X^2,\sigma_N^2) \le \frac{\sigma_N^2}{2\sigma_X^2}$. These results should be contrasted with the fact that, if the agents were allowed to convene before communicating to the CEO, they could smooth out their noisy observations and achieve a rate distortion performance corresponding to that of the source X, i.e., the distortion would decay as 2^{-2R}. Thus, there is a significant performance degradation in the isolated agents case. This problem also serves as an interesting example for connections between information theory and statistics. We use the Cramer-Rao bound for random parameter estimation for lower bounding the achievable distortion. The problem is described in detail in Section 2 and the main result is presented in Section 3.

II. Problem Statement

The formal description of the Gaussian CEO problem is as follows. The CEO is interested in a i.i.d Gaussian data sequence $\{X(t)\}_{t=1}^{\infty}$ with variance σ_X^2. This data sequence cannot be directly observed by the CEO. Versions $\{Y_i(t)\}$ of $\{X(t)\}$ corrupted by independent additive white Gaussian noise with variance σ_N^2 are observed by a team of L agents. The agents are not allowed to convene; Agent i has to send data based solely on his own noisy observations $\{Y_i(t)\}_{t=1}^{\infty}$. The agents are required to send encoded versions of the data observed through noiseless communication channels with a total rate R. Symbolically,

$$Y_i(t) = X(t) + N_i(t)$$

where $X(t)$ is $\mathcal{N}(0,\sigma_X^2)$ distributed and $N_i(t)$ is independent and identical over i,t and is distributed $\mathcal{N}(0,\sigma_N^2)$ for $i = 1,\ldots,L$ and $t = 1,\ldots n,\ldots$.

For $i = 1,\ldots L$, Agent i encodes a block of length n from his observed data $\{y_i(t)\}_{i=1}^{\infty}$ using a source code $\mathcal{C}_i^n$ of rate $R_i^n = \frac{1}{n}\log|\mathcal{C}_i^n|$. The code words from the L agents, $C_1^n,\ldots C_L^n$, are sent to a central estimator whose task is to recover the source message $x^n = (x(1),\ldots x(n))$ as accurately as possible in terms of the mean squared error defined as

$$D^n(X^n,\hat{X}^n) = \frac{1}{n}E\sum_{t=1}^{n}(X(t)-\hat{X}(t))^2 \tag{1}$$

where $\hat{X}^n$ is the estimate of the random message X^n made by the CEO. Denote the CEO's estimate by

$$\hat{X}^n = \Phi_L^n(C_1^n,\ldots C_L^n) \tag{2}$$

where C_i^n denotes the code word selected by Agent i ; C_i^n is random because of the joint randomness of the message and observation noise.

We study the tradeoff between the total rate, $R = \sum_{i=1}^{L} R_i^n$, and the mean squared error $D^n(X^n,\hat{X}^n)$ in the following format. For the given codes $\mathcal{C}_i^n, i = 1,\ldots L$ of block length n, let

$$D^n(\mathcal{C}_1^n,\ldots,\mathcal{C}_L^n) = \min_{\Phi_L^n} D^n(X^n,\Phi_L^n(C_1^n,\ldots,C_L^n)) \tag{3}$$

Define

$$D^n(L,R) \quad = \quad \min_{\sum_{i=1}^{L} R_i^n \le R} D^n(\mathcal{C}_1^n,\ldots,\mathcal{C}_L^n), \tag{4}$$

$$D(R) \quad = \quad \lim_{L\to\infty}\lim_{n\to\infty} D^n(L,R) \tag{5}$$

and

$$\beta(\sigma_X^2,\sigma_N^2) = \lim_{R\to\infty} R\frac{D(R)}{\sigma_X^2}. \tag{6}$$

III. Main Result

Theorem *Let $Q(u|y)$ be any conditional density on an arbitrary alphabet U, and let $\tilde{Q}(u|x) = \int_y W(y|x)Q(u|y)dy$. Then under the usual Cramer-Rao regularity conditions*

$$\beta(\sigma_X^2,\sigma_N^2) \quad \ge \quad \inf_{Q(u|y)} \frac{I(Y;U|X)}{\sigma_X^2 E[-\frac{\partial^2}{\partial X^2}\log\tilde{Q}(U|X)]} > 0$$

Also,

$$\beta(\sigma_X^2,\sigma_N^2) \le \frac{\sigma_N^2}{2\sigma_X^2}$$

We believe that the bounds are actually tight, but have been unable to establish this.

References

[1] Berger and Zhang " On the CEO Problem ", 1994 IEEE International Symposium on Information Theory , Trondheim, Norway.

Gaussian Multiterminal Source Coding

Yasutada Oohama
Department of Computer Science
and Communication Engineering
Faculty of Engineering, Kyushu University
6-10-1 Hakozaki, Higashi-ku
Fukuoka 812, Japan

Abstract — **We consider the problem of separate coding for two correlated memoryless Gaussian source. We determine the rate-distortion region in a special case that one source plays a role of partial side information to reproduce sequences emitted from the other source with a prescribed average distortion level. We also derive an explicit outer bound of the rate-distortion region, demonstrating that the inner bound obtained by Berger partially coincides with the rate-distortion region.**

I. Introduction

Let X and Y be Gaussian random variables with mean 0 and variance σ_X^2 and σ_Y^2, respectively. We denote by ρ the correlation coefficient between X and Y. We write n-independent copies of X, Y as $X^n = X_1, X_2, \ldots, X_n$, $Y^n = Y_1, Y_2, \ldots, Y_n$, respectively. Data sequences X^n and Y^n are separately encoded to $\varphi_1(X^n)$ and $\varphi_2(Y^n)$ and both are sent to the information processing center, where the decoder function ψ observes $\varphi_1(X^n)$ and $\varphi_2(Y^n)$ to output the estimation $(\hat{X}^n, \hat{Y}^n)$ of (X^n, Y^n). The encoder functions φ_i $(i = 1, 2)$ satisfy rate constraints $\frac{1}{n}\log\|\varphi_i\| \leq R_i + \delta$ $(i = 1, 2)$, where δ is an arbitrary prescribed positive number. For $(\hat{X}^n, \hat{Y}^n) = \psi(\varphi_1(X^n), \varphi_2(Y^n))$, define the mean square errors Δ_i $(i = 1, 2)$ by $\Delta_1 = E\frac{1}{n}\sum_{t=1}^n (X_t - \hat{X}_t)^2$, $\Delta_2 = E\frac{1}{n}\sum_{t=1}^n (Y_t - \hat{Y}_t)^2$. For given positive numbers D_i $(i = 1, 2)$, a rate pair (R_1, R_2) is *admissible* if for any $\delta > 0$ and any $n \geq n_0(\delta)$ there exists a triple $(\varphi_1, \varphi_2, \psi)$ such that $\Delta_i \leq D_i + \delta$ $(i = 1, 2)$. We denote by $\mathcal{R}(D_1, D_2)$ the set of all the admissible pair (R_1, R_2).

Our main goal is to determine rate-distortion region $\mathcal{R}(D_1, D_2)$. Berger [1] derived the inner bound of $\mathcal{R}(D_1, D_2)$. However, the optimality was not discussed in his paper.

In this paper, we determine the rate-distortion region for a certain special case, and show that the inner bound obtained by Berger partially coincides with $\mathcal{R}(D_1, D_2)$.

II. Statement of Main Results

If $D_2 \geq \sigma_Y^2$, there is substantially no constraint between Y^n and $\hat{Y}^n$. It means that Y^n works as an auxiliary information to reproduce X^n with a distortion level not greater than D_1, and that $\mathcal{R}(D_1, D_2)$ does not depend on D_2. We denote this region by $\mathcal{R}_1(D_1)$. Then the following theorem holds.

Theorem 1 : For every $D_1 > 0$

$$\mathcal{R}_1(D_1) = \left\{(R_1, R_2) \mid R_1 \geq \frac{1}{2}\log^*\left[(1-\rho^2)\frac{\sigma_X^2}{D_1}\left(1 + \frac{\rho^2}{1-\rho^2}\cdot\frac{s}{\sigma_Y^2}\right)\right], \; R_2 \geq \frac{1}{2}\log\left[\frac{\sigma_Y^2}{s}\right] \text{ for some } 0 < s \leq \sigma_Y^2 \right\},$$

where $\log^* x = \max\{\log x, 0\}$.

Wyner-Ziv [2] and Wyner [3] have determined the rate-distortion function for the case $s \to 0$ that the decoder can fully observe the side information Y^n. Theorem 1 is an extension of their results to the case that the decoder can observe *partial* side information. For the case that the random pair (X, Y) takes finite values the inner region of $\mathcal{R}_1(D_1)$ was derived by Berger et al. [4], but the determination problem of $\mathcal{R}_1(D_1)$ still remains open for this case.

Next, we derive an explicit outer bound of $\mathcal{R}(D_1, D_2)$. Let $\mathcal{R}_2(D_2)$ be the rate-distortion region for the case $D_1 \geq \sigma_X^2$. We obtain the following theorem.

Theorem 2 : For every $D_1, D_2 > 0$

$$\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}_{out}(D_1, D_2),$$

where

$$\mathcal{R}_{out}(D_1, D_2) = \mathcal{R}_1(D_1) \cap \mathcal{R}_2(D_2) \cap \hat{\mathcal{R}}_{12}(D_1, D_2),$$

$$\hat{\mathcal{R}}_{12}(D_1, D_2) = \left\{(R_1, R_2) \mid R_1 + R_2 \geq \frac{1}{2}\log\left[(1-\rho^2)\frac{\sigma_X^2\sigma_Y^2}{D_1 D_2}\right]\right\}.$$

The inner bound $\mathcal{R}_{in}(D_1, D_2)$ of $\mathcal{R}(D_1, D_2)$ according to Berger [1] is

$$\mathcal{R}_{in}(D_1, D_2) = \mathcal{R}_1(D_1) \cap \mathcal{R}_2(D_2) \cap \tilde{\mathcal{R}}_{12}(D_1, D_2),$$

where

$$\tilde{\mathcal{R}}_{12}(D_1, D_2) = \left\{(R_1, R_2) \mid R_1 + R_2 \geq \frac{1}{2}\log\left[(1-\rho^2)^{\frac{\beta}{2}}\cdot\frac{\sigma_X^2\sigma_Y^2}{D_1 D_2}\right]\right\},$$

$$\beta = 1 + \sqrt{1 + \frac{4\rho^2}{(1-\rho^2)^2}\cdot\frac{D_1 D_2}{\sigma_X^2\sigma_Y^2}}.$$

The boundary of $\mathcal{R}_{in}(D_1, D_2)$ consists of one straight line segment defined by the boundary of $\tilde{\mathcal{R}}_{12}(D_1, D_2)$ and two curved portions defined by the boundaries of $\mathcal{R}_1(D_1)$ and $\mathcal{R}_2(D_2)$. Hence, the inner bound established by Berger partially coincides with $\mathcal{R}(D_1, D_2)$ at two curved portions of its boundary. The gap between inner and outer bounds is the difference of the rate sum given by $\Delta R = \frac{1}{2}\log\left[\frac{\beta}{2}\right]$. We found that ΔR is negligible for relatively small values of D_1 and D_2. However, further discussions are still necessary for resolving this gap.

References

[1] T. Berger, "Multiterminal Source Coding," in *The Information Theory Approach to Communications*, G.Longo, Ed., CISM Courses and Lectures 229, New York : Springer, 1978.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol.IT-22, pp.1-10, 1976.

[3] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr.*, 38, pp.60 -80, 1978.

[4] T. Berger, K. B. Housewright, J. K. Omura, S. Tung, and J. Wolfowitz, "An upper bound on the rate-distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inform. Theory*, vol.IT-25, pp.664-666, 1979.

Multilevel Diversity Coding with Symmetrical Connectivity

James R. Roche[1], Raymond W. Yeung[2], and Ka Pun Hau[3]

I. Introduction

Multilevel diversity coding was recently introduced by Roche [1] and Yeung [2]. In a Multilevel Diversity Coding System, the decoders are partitioned into multiple *levels*. The reconstructions of the source by decoders within the same level are identical and are subject to the same distortion criterion. A comprehensive discussion of multilevel diversity coding is found in [2]. In particular, we refer the readers to [2] for the basic results and the notion of *superposition* in multilevel diversity coding.

In [2], a class of problems in multilevel diversity coding was suggested. In this paper we consider one such problem with symmetrical connectivity between the encoders and decoders (see Fig. 1). In this problem, there are three encoders and seven decoders. The source $\{X_k\}$ is an independent and identically distributed (i.i.d.) process. The seven decoders belong to three levels: Decoders 1, 2, and 3 belong to Level 1; Decoders 4, 5, and 6 belong to Level 2; and Decoder 7 belongs to Level 3. Note that each Level i decoder has access to i encoders, and $\{(\tilde{X}_i)_k\}$ is the reproduction of $\{X_k\}$ by a Level i decoder. We are interested in finding the trade-off between the rates of the encoders and the distortions of the reconstructions of the source by the decoders.

II. The Main Result

Defining rates and distortions in the usual way, we let R_i be the rate of Encoder i and let D_i be the maximum allowable distortion for each decoder in Level i, where in general each level has its own distortion function. Say that a sextuple $(R_1, R_2, R_3, D_1, D_2, D_3)$ is *admissible* if there exists a coding scheme with the given rates and expected distortions in the usual Shannon sense. Let

$$\mathcal{R} = \{(R_1, R_2, R_3, D_1, D_2, D_3) : (R_1, R_2, R_3, D_1, D_2, D_3) \text{ is admissible }\},$$

and let $\mathcal{R}^*$ be the set consisting of all $(R_1, R_2, R_3, D_1, D_2, D_3)$ satisfying the two conditions below:

1) for $l = 1, 2, 3$, there exists $\hat{X}_l$ such that

$$E\ d_l(X, \hat{X}_l) \le D_l; \tag{1}$$

2)

$$R_i \ge I(X; \hat{X}_1) \quad \text{for } 1 \le i \le 3 \tag{2}$$

$$R_i + R_j \ge 2I(X; \hat{X}_1) + I(X; \hat{X}_2|\hat{X}_1) \quad \text{for } 1 \le i < j \le 3 \tag{3}$$

$$2R_i + R_{i\oplus 1} + R_{i\oplus 2} \ge 4I(X; \hat{X}_1) + 2I(X; \hat{X}_2|\hat{X}_1) + I(X; \hat{X}_3|\hat{X}_1, \hat{X}_2) \quad \text{for } 1 \le i \le 3 \tag{4}$$

$$R_1 + R_2 + R_3 \ge 3I(X; \hat{X}_1) + \frac{3}{2}I(X; \hat{X}_2|\hat{X}_1) + I(X; \hat{X}_3|\hat{X}_1, \hat{X}_2), \tag{5}$$

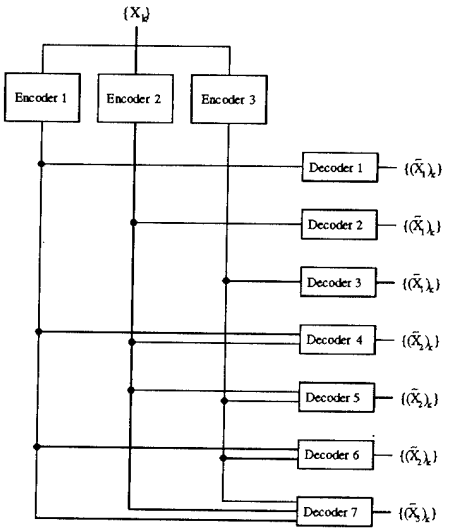


Figure 1: A symmetrical multilevel diversity coding system

where $\oplus$ is defined by

$$x \oplus y = \begin{cases} x + y & \text{if } x + y \le 3 \\ x + y - 3 & \text{if } x + y > 3. \end{cases}$$

It is shown in [3] that condition 2) is equivalent to

2′) For $i = 1, 2, 3$,

$$R_i = r_i^1 + r_i^2 + r_i^3, \tag{6}$$

where $r_i^1, r_i^2, r_i^3 \ge 0$, and

$$r_i^1 \ge I(X; \hat{X}_1) \quad \text{for } 1 \le i \le 3 \tag{7}$$

$$r_i^2 + r_j^2 \ge I(X; \hat{X}_2|\hat{X}_1) \quad \text{for } 1 \le i < j \le 3 \tag{8}$$

$$r_1^3 + r_2^3 + r_3^3 \ge I(X; \hat{X}_3|\hat{X}_1, \hat{X}_2). \tag{9}$$

We now state our main result.

Theorem 1 *$\mathcal{R}$ is the closure of con($\mathcal{R}^*$), where con($\mathcal{R}^*$) denotes the convex hull of $\mathcal{R}^*$.*

Proof The rate constraints in 2′) are used for proving the admissibility of $\mathcal{R}^*$, while the rate constraints in 2) are used for proving the converse. Please refer to [3] for the details of the proof. □

References

[1] J. R. Roche, "Distributed information storage," Ph.D. thesis, Stanford University, March 1992.

[2] R. W. Yeung, "Multilevel diversity coding with distortion," *IEEE Trans. Inform. Theory*, vol. 41, pp. 412-422, 1995.

[3] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical multilevel diversity coding," in preparation.

[1] Center for Communications Research, Thanet Road, Princeton, NJ 08540, USA; e-mail:roche@ccr-p.ida.org.

[2] Department of Information Engineering, the Chinese University of Hong Kong, N.T., Hong Kong; whyeung@ie.cuhk.hk

[3] Department of Information Engineering, the Chinese University of Hong Kong, N.T., Hong Kong; kphau3@ie.cuhk.hk

An Error Exponent for Lossy Source Coding with Side Information at the Decoder

Srikant Jayaraman and Toby Berger
School of Electrical Engineering, Cornell University, Ithaca, NY 14853, U.S.A.

I. Introduction

Consider discrete memoryless sources X and Y with joint pdf P_{XY} and let $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ be a single-letter distortion measure. Suppose the source X must be compressed at rate R with distortion no larger than D, when the side information Y is available only to the decoder. This is the Wyner-Ziv model for lossy source coding with side information at the decoder [1, 2]; one method of constructing good source codes is described below.

Let U be a r.v. taking values in $\mathcal{U}$ and assume that UXY have joint distribution satisfying $U \to X \to Y$. Define the following concepts:

1. An ordered collection C of $M = e^{n(I(X;U)+\eta)}$ n-tuples over the alphabet $\mathcal{U}$ is called a code book if each n-tuple is an element of the δ-typical set $T^n_{[U]_\delta}$; let $\mathcal{C}$ be the collection of all such code books.
2. For $R > 0$, a map $f : \{1, \ldots, M\} \to \{1, \ldots, e^{nR}\}$ is called a binning scheme; let $\mathcal{F}$ be the collection of all such binning schemes.

Given a code $(f, C) \in \mathcal{F} \times \mathcal{C}$, we consider standard *joint-typicality* encoding and *joint-typicality* bin decoding as described in [2]. It is well known that by suitable choice of $U \to X \to Y$, one can generate codes $(f, C) \in \mathcal{F} \times \mathcal{C}$ which operate over the *entire* region of achievable rate-distortion tuples. In other words, the family of coding strategies $\mathcal{F} \times \mathcal{C}$ contains schemes which are "optimal", in the rate-distortion sense.

This leads to the following question: for a fixed $U \to X \to Y$, is it possible to bound the error performance of the codes $(f, C) \in \mathcal{F} \times \mathcal{C}$? We can place meaningful exponential bounds on the error behavior of codes in the ensemble, if we use a *minimum entropy decoder* [3]. This decoder selects from the specified bin (say, bin k) any code word $\bar{u}_j$ which minimizes the empirical conditional entropy $H(\bar{u}_j|\bar{y})$.

II. The Error Exponent

There are two possible error events:

$$\begin{aligned} E'(f,C) &= \{(\bar{x},\bar{y}) : (\bar{x},\bar{u}_i) \notin T^n_{[XU]_\delta}, \forall \bar{u}_i \in C\}, \\ E(f,C) &= \{(\bar{x},\bar{y}) : \exists i\, (\bar{x},\bar{u}_i) \in T^n_{[XU]_\delta}, \\ &\qquad \exists j \neq i : f(i) = f(j), H(\bar{u}_j|\bar{y}) \leq H(\bar{u}_i|\bar{y})\}. \end{aligned}$$

Event $E'(f, C)$ is an encoder failure — $\bar{x} \in \mathcal{X}^n$ cannot be encoded into any $\bar{u}_i \in C$. Event $E(f, C)$ is a decoder failure; the decoder is, even with its side information $\bar{y}$, unable to extract the correct $\bar{u}_i$ from the specified bin.

Since $P^n_{XY}(E'(f, C))$ decays as $e^{-n\delta}$, we cannot determine a non-trivial exponential bound on $P^n_{XY}(E'(f, C))$. Our approach therefore will be to ignore the encoder error behavior entirely, and we shall only require that a code $(f, C) \in \mathcal{F} \times \mathcal{C}$ satisfy $P^n_{XY}(E'(f, C)) \to 0$. In this subclass of codes, we *define* an error exponent based on the decoder failure event $E(f, C)$:

$$\theta(R, UXY) \triangleq \sup_{\substack{(f,C)\in\mathcal{F}\times\mathcal{C} \\ P^n_{XY}(E'(f,C))\to 0}} \limsup_{n\to 0} -\frac{1}{n} \log P^n_{XY}(E(f,C)).$$

We believe that this definition is of significance because the event $E(f, C)$ arises in virtually every multiterminal source coding configuration. Moreover, as recently shown by Shimokawa, Han and Amari [4], a variation of $E(f, C)$ takes on particular importance in the multiterminal hypothesis testing problem. The lossy source coding model of Wyner and Ziv provides a canonical setup for studying this multiterminal error exponent.

III. Bounds on the Error Exponent

It is easy to see that $\theta(R, UXY) = 0$ for $R < I(X; U|Y)$ and $\theta(R, UXY) = \infty$ for $R > I(X; U)$. For rates $R \in [I(X; U|Y), I(X; U)]$, we define:

$$\theta_U(R, UXY) \triangleq \min_{\substack{\mathcal{U}\tilde{X}\tilde{Y} \\ U\to\tilde{X}\to\tilde{Y} \\ P_{\tilde{X}U}=P_{XU} \\ R\leq I(\tilde{X};\mathcal{U}|\tilde{Y})}} D(\tilde{U}\tilde{X}\tilde{Y}\|UXY),$$

and

$$\theta_L(R, UXY) \triangleq \min_{\substack{\mathcal{U}\tilde{X}\tilde{Y} \\ P_{\tilde{X}U}=P_{XU}}} D(\tilde{U}\tilde{X}\tilde{Y}\|UXY) + |R - I(\tilde{X};\tilde{U}) + I(\tilde{Y};\tilde{U})|^+.$$

Theorem 1

$$\theta_L(R, UXY) \leq \theta(R, UXY) \leq \theta_U(R, UXY).$$

The lower bound is proved by random selection of codes over the ensemble $\mathcal{F} \times \mathcal{C}$, and the upper bound follows by a sphere-packing argument [3]. As a check, we observe that $\theta_L(R, UXY)$ is strictly positive when $R > I(X; U|Y)$; accordingly, our result yields another proof of the direct part of the Wyner-Ziv theorem. We also note that the sphere-packing and random-coding bounds need not agree, even for $R \approx I(X; U|Y)$. We conjecture that the random-coding bound is tight near the lower rate boundary; a more clever application of the sphere-packing technique might close the gap and prove our conjecture.

References

[1] A. Wyner and J. Ziv. "The rate distortion function for source coding with side information at the receiver," *IEEE Trans. Inform. Theory*, IT-22:1-11, 1976.

[2] T.Berger. Multiterminal Source Coding, In G. Longo, editor, *The Information Theory Approach to Communications*, Springer-Verlag, New York, 1977.

[3] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.

[4] H. Shimokawa, T.S. Han, S. Amari. "Error bound of hypothesis testing with data compression," presented at the *IEEE International Symposium on Information Theory*, Trondheim, Norway, July 1994.

Error Exponents for Successive Refinement by Partitioning

Angelos Kanlis and Prakash Narayan[1]
Elect. Eng. Dept., Univ. of Maryland, College Park, MD, U.S.A.

Abstract — **We consider the rate-distortion function for successive refinement by partitioning and determine error exponents for two-step coding. It is seen that even when the rate-distortion functions for one- and two-step coding coincide, the error exponent in the former case may exceed those in the latter.**

I. Introduction

Given a discrete memoryless source (DMS) with probability mass function (pmf) P, and a suitable distortion measure, the minimum rate of coding at distortion Δ_1 is given by the rate distortion function $R(P, \Delta_1)$. If a finer description is required, say with distortion $\Delta_2 < \Delta_1$, additional information can be provided at rate $R_2 - R(P, \Delta_1)$. Clearly, $R_2 \geq R(P, \Delta_2)$. The minimum value of R_2 is the two-step rate-distortion function, $R(P, \Delta_1, \Delta_2)$ (Rimoldi [4]). The Markov condition under which $R(P, \Delta_1, \Delta_2) = R(P, \Delta_2)$ was determined independently by Koshélev [3] and Equitz-Cover [2].

II. Preliminaries

Let $\mathcal{X}$ be a finite set and $\{X_t\}_{t=1}^{\infty}$ be a $\mathcal{X}$-valued DMS with pmf P. Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be finite reproduction alphabets; and $d_i : \mathcal{X} \times \mathcal{Y}_i \to \mathbf{R}^+$, $i = 1, 2$, nonnegative-valued mappings that induce distortion measures on $\mathcal{X}^n \times \mathcal{Y}_i^n$, $i = 1, 2$, according to

$$d_i(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{t=1}^{n} d_i(x_t, y_t), \quad \mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}_i^n, \ i = 1, 2.$$

A two-step n-length block code consists of two encoder-decoder pairs $f_i^{(n)} : \mathcal{X}^n \to \mathcal{M}_i = \{1, \cdots, M_i\}$, $\phi_i^{(n)} : \prod_{j=1}^{i} \mathcal{M}_j \to \mathcal{Y}_i^n$, $i = 1, 2$.

For given rate $R_1 > 0$ and distortions $\Delta_1 > \Delta_2 \geq 0$ let $R(P, R_1, \Delta_1, \Delta_2)$ denote the minimum rate of the two-step code when the first-step code has rate R_1 and distortion Δ_1 and the two-step code has distortion Δ_2. It constitutes the rate-distortion function for the refining code and follows immediately from [4] :

$$R(P, R_1, \Delta_1, \Delta_2) = \inf_{\substack{P_X = P \\ \mathbf{E}d_1(X, Y_1) \leq \Delta_1 \\ \mathbf{E}d_2(X, Y_2) \leq \Delta_2 \\ I(X \wedge Y_1) \leq R_1}} I(X \wedge Y_1 Y_2).$$

III. Main Results

For convenience, we define

$$e_1 \triangleq \Pr\left(d_1(X^n, \phi_1(f_1(X^n))) > \Delta_1 \text{ or } d_2(X^n, \phi_2(f_1(X^n), f_2(X^n))) > \Delta_2\right),$$

$$e_1 \triangleq \Pr\left(d_2(X^n, \phi_2(f_1(X^n), f_2(X^n))) > \Delta_2\right),$$

where (f_1, ϕ_1) is of rate R_1 and distortion Δ_1. Further, let

$$F_1 \triangleq \inf_{\substack{Q: \ R(Q, \Delta_1) > R_1 \\ or \ \ R(Q, R_1, \Delta_1, \Delta_2) > R_2}} D(Q \| P),$$

$$F_1 \triangleq \inf_{Q: R(Q, R_1, \Delta_1, \Delta_2) > R_2} D(Q \| P).$$

Our main results establish that there exists a sequence of two-step codes of rates (R_1, R_2) with

$$\frac{1}{n} \log e_1 \leq -F_1 + \delta_1,$$
$$\frac{1}{n} \log e_2 \leq -F_2 + \delta_2,$$

for any $\delta_1, \delta_2 > 0$. Further, for any sequence of codes of rates (R_1, R_2)

$$\liminf_{n \to \infty} \frac{1}{n} \log e_1 \geq -F_1,$$
$$\liminf_{n \to \infty} \frac{1}{n} \log e_2 \geq -F_2.$$

Finally, even with the Markov condition [2, 3] in effect, so that

$$R(P, R_1, \Delta_1, \Delta_2) = R(P, \Delta_2), \ \ \Delta_2 < \Delta_1,$$

it is possible that $F_1 < F_2$. This is illustrated by the simple example of a DMS with Hamming distance distortion measure, where a simple necessary and sufficient condition for the error exponents to differ is

$$R_2 - R(P, \Delta_2) > R_1 - R(P, \Delta_1).$$

References

[1] I. Csiszár and J. Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems*. Academic Press, NY, 1981.

[2] W. H. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, no. 2, pp. 269-275, Mar. 1991.

[3] V. Koshélev, "Estimation of mean error for a discrete successive-approximation scheme," *Problemy Peredachi Informatsii*, vol. 17, No. 3, pp. 20-33, July-September, 1981.

[4] B. Rimoldi, "Successive refinement of information : characterization of the achivable rates," *IEEE Trans. Inform. Theory* vol. IT-40, no. 1, pp. 253-259, Jan. 94.

[1]This work was supported by NSF Grant OIR-85-00108

Remote Coding of Correlated Sources with High Resolution

Ram Zamir and Toby Berger

School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA . *e-mail: zamir / berger @ee.cornell.edu*

I. Direct Coding with High Resolution

Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of independent drawings of a pair of dependent continuous random variables. It is desired to block encode $\underline{X} = X_1 \dots X_n$ and $\underline{Y} = Y_1 \dots Y_n$ *separately*, and decode them *jointly*, such that $\frac{1}{n} E \sum_{i=1}^{n} (\widehat{X}_i - X_i)^2 \le D_x$, and $\frac{1}{n} E \sum_{i=1}^{n} (\widehat{Y}_i - Y_i)^2 \le D_y$. Let $\mathcal{R}(D_x, D_y) = \{(R_1, R_2)\}$ denote the set of rate pairs of X- and Y- encoders which satisfy these constraints for some n. Assume that the joint differential entropy $h(X,Y)$ exists and is finite, and let $\mathcal{R}^*(D_x, D_y)$ be the set of (R_1, R_2) pairs which satisfy

$$\begin{aligned} R_1 &\ge h(X|Y) - \tfrac{1}{2}\log 2\pi e D_x \\ R_2 &\ge h(Y|X) - \tfrac{1}{2}\log 2\pi e D_y \\ R_1 + R_2 &\ge h(X,Y) - \tfrac{1}{2}\log (2\pi e)^2 D_x D_y \,. \end{aligned} \tag{1}$$

Note that $\mathcal{R}^*(D_x, D_y)$ has the known "broken corner" structure of the Slepian-Wolf rate region.

Theorem 1 (Shannon Outer Bound) *For any D_x and D_y,*

$$\mathcal{R}(D_x, D_y) \subseteq \mathcal{R}^*(D_x, D_y) \,. \tag{2}$$

Furthermore, if $E\{X^2\} < \infty$, $E\{Y^2\} < \infty$ and $h(X,Y) > -\infty$, then, as $D_x, D_y \to 0$, the outer bound (2) becomes achievable, i.e.; $\mathcal{R}(D_x, D_y) \sim \mathcal{R}^*(D_x, D_y)$, where $\sim$ means that for any $\omega_1 > 0$ and $\omega_2 > 0$,

$$\min_{R_1, R_2 \in \mathcal{R}} \{\omega_1 R_1 + \omega_2 R_2\} - \min_{R_1, R_2 \in \mathcal{R}^*} \{\omega_1 R_1 + \omega_2 R_2\} \to 0 \,. \tag{3}$$

This theorem has a straightforward extension to general difference distortion measures.

II. Remote Coding with High Resolution

Now restrict our attention to the Gaussian case, but consider the following more general, *indirect* coding problem [1, pp. 78, 124]. We need to reconstruct a (memoryless) zero mean vector source $\underline{\theta} = \theta_1 \dots \theta_m$, jointly Gaussian with (X, Y), from separate encodings of X and Y, with averaged squared errors $D_1 \dots D_m$. We refer to $\underline{\theta}$ as the remote source, and to X and Y as its noisy measurements. Denote by $\mathcal{R}(D_1 \dots D_m) = \{(R_1, R_2)\}$ the set of admissible rate pairs of the X- and Y- encoders.

When X and Y are available with infinite resolution, the optimal reconstruction of $\underline{\theta}$ is the conditional mean $E(\underline{\theta}|X,Y) = H \cdot (X,Y)^t$, where H is some $m \times 2$ matrix. The mean squared errors are then the diagonal elements $D_1^{opt} \dots D_m^{opt}$ of the conditional covariance matrix $\mathrm{COV}(\underline{\theta}|X,Y)$. Clearly we can only satisfy distortions $D_i \ge D_i^{opt}$, and usually when $(D_1, \dots, D_m) \to (D_1^{opt}, \dots, D_m^{opt})$ the coding rates must go to infinity. We are interested in the asymptotic behavior in this limit. Let $\mathcal{K} = \{K\}$ be the set of all 2×2 nonnegative definite matrices which satisfy

$$\left(HKH^t\right)_{i,i} \le D_i - D_i^{opt} \,, \quad \text{for } i = 1 \dots m, \tag{4}$$

where H is the matrix associated with the optimal estimator defined above, and define

$$\mathcal{R}^{**}(D_1 \dots D_M) = \bigcup_{\{D_x, D_y \,:\, \mathrm{diag}[D_x, D_y] \in \mathcal{K}\}} \mathcal{R}^*(D_x, D_y) \,, \tag{5}$$

where the union is taken over all diagonal matrices in $\mathcal{K}$ whose diagonal elements are D_x and D_y, and the rate region $\mathcal{R}^*$ was defined in (1).

Theorem 2 *Assume that $E\{X^2\} < \infty$, $E\{Y^2\} < \infty$ and $h(X,Y) > -\infty$, and that H does not have all-zero columns. Then, as $(D_1, \dots, D_m) \to (D_1^{opt}, \dots, D_m^{opt})$, the region of admissible rate pairs satisfies*

$$\mathcal{R}(D_1 \dots D_m) \sim \mathcal{R}^{**}(D_1 \dots D_m) \,, \tag{6}$$

where the notation $\sim$ was defined in (3).

III. The Loss due to Separate Encoding

To gain some insight into these results, we examine below the total rate loss caused by the separation of the encoders in the high resolution limit. In the direct coding case (Section I) this loss is zero, since the rate in jointly encoding X and Y is given asymptotically by the Shannon lower bound [1, p. 92]), which coincides with the minimal rate sum of the separate encoders given in (1)-line 3.

In the indirect coding case (Section II), however, the loss may be positive. Let $\mathcal{K}^{**}$ denote the subset of *diagonal* matrices in $\mathcal{K}$, and let *det* and *det*** denote the maximum determinants over all matrices in $\mathcal{K}$ and $\mathcal{K}^{**}$, respectively. As $(D_1 \dots D_m) \to (D_1^{opt} \dots D_m^{opt})$, the rate sum of the X- and Y- encoders exceeds the rate of the joint encoder by

$$\frac{1}{2}\log\left(\frac{det}{det^{**}}\right) \,. \tag{7}$$

This loss is due to the fact that at high resolution the quantization errors made by the separate encodings of X and Y are effectively uncorrelated, and so we cannot take advantage of the *shaping gain*. When the number of remote sources is smaller than the number of measurements (i.e., when $m = 1$), the quantity (7) equals infinity. In fact, the rate sum of the separate encoders in this case is roughly *twice* the rate of the joint encoder (which diverges to infinity). This is because the separate encoders quantize two measurements, while the joint encoder effectively quantizes *one* continuous random variable, $E(\theta|X,Y)$, at about the same resolution.

Acknowledgements

The authors wish to thank helpful discussions with Raymond Yeung and Harish Viswanathan.

References

[1] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[0]This research was supported in part by the Wolfson Research Awards, administered by the Israel Academy of Science and Humanities, and by NSF Grants NCR-9216975 and IRI-9310670.

A Sliding Window Lempel-Ziv Algorithm for Differential Layer Encoding in Progressive Transmission

P.Subrahmanya and Toby Berger

Department of Electrical Engineering, Cornell University, Ithaca, NY 14853

I. Introduction

Differential layer encoding in progressive transmission refers to the process of generating additional bits which, in conjunction with a low resolution version of the image, enable the decoder to reconstruct the high resolution image. A practical algorithm for the progressive transmission of black and white images is the JBIG algorithm [1], which primarily uses an arithmetic coder for differential layer encoding. We consider differential layer encoding as an instance of coding with side information known to both the encoder and the decoder. Based on this idea, we present a sliding window Lempel-Ziv algorithm for differential layer encoding, and apply it to compress black and white images. The algorithm presented here can also be applied to other problems such as successive refinement of information [4].

II. Coding with Side Information

Consider a source coding situation where both encoder and decoder know a sequence X_1^N of letters drawn from a finite alphabet $\mathcal{X}$. The decoder now needs to transmit a sequence Y_1^N of letters drawn from an alphabet $\mathcal{Y}$. This is the problem of coding with side information known to both the encoder and the decoder. The sequence X_1^N is known as the side information. An algorithm for this data compression problem is as follows.

The Parsing

1. Initialization - First, we fix a window size n_w. Transmit the first n_w symbols of the Y sequence, $Y_1^{n_w}$ without any compression.
2. Matching - Let L^1 be the largest integer such that a copy of $(XY)_{n_w+1}^{n_w+L^1-1}$ begins in the current window $(XY)_1^{n_w}$. Let the copy begin at position *start*. Denote $Y_{n_w+1}^{n_w+L^1}$ as Y^1, the first phrase.
3. Sliding - Define the new window to be $(XY)_{L^1+1}^{n_w+L^1}$.

Repeat steps (2.) and (3.) as many times as necessary until the sequence is exhausted. Note that this parsing is identical to that produced by applying the sliding window Lempel-Ziv algorithm [2] to the XY sequence.

Representation of the phrases in binary

The phrase $Y^1 = Y_{n_w+1}^{n_w+L^1}$ consists of two parts, the matched portion $Y_{n_w+1}^{n_w+L^1-1}$, and the last symbol $Y_{n_w+L^1}$. We thus, need to specify three things.

1. The last symbol, represented using $\lceil \log(|\mathcal{Y}|) \rceil$ bits.
2. The length of the phrase, L^1, which can be represented using $\lceil \log L^1 \rceil + 2\lceil \log \log L^1 \rceil$ bits [3].
3. Starting point of the match - Let

$$N_{X^1} = |\{k \text{ s.t. } start \leq k \leq n_w \,,\, X_{k+1}^{k+L^1-1} = X_{n_w+1}^{n_w+L^1-1}|\}$$

Then, the starting point of the match can be specified using $\lceil \log N_{X^1} \rceil + 2\lceil \log \log N_{X^1} \rceil$ bits.

4. If the number of bits needed to represent the phrase using (1.), (2.), and (3.) exceeds $L^1\lceil \log(|\mathcal{Y}|) \rceil$ bits, then encode the phrase without any compression. The number of bits needed to do this is less than $\gamma_2 L^1$.

Let the total number of bits needed to represent the first phrase be denoted by $B(Y^1)$. Then,

$$B(Y^1) \leq \min\{\log N_{X^1} + 2\log\log N_{X^1} + \gamma_1 \log L^1, \gamma_2 L^1\}$$

Lemma 1 *Consider the sequence $(XY)_{-\infty}^{\infty}$ produced by a stationary ergodic source. Define for $l > 0$,*

$$\begin{aligned} W_l(xy) &= \min\{k : k > 0, (xy)_0^{l-1} = (xy)_{-k-l-1}^{-k}\} \\ N_X(l) &= |\{k : k \leq W_l(xy), (x)_0^{l-1} = (x)_{-k-l-1}^{-k}\}| \\ \text{Then,} \quad & Pr\{N_X(l) > 2^{l(H(Y|X)+\epsilon)}\} \to 0 \text{ as } l \to \infty \end{aligned}$$

where $H(Y|X)$ is the conditional entropy.

Based on this lemma, we conjecture that, if the above algorithm is used to parse an input of N symbols into exactly c phrases, then

$$\lim_{n_w \to \infty} \lim_{N \to \infty} E\left(\frac{1}{N} \sum_{j=1}^{c} B(Y^j)\right) = H(Y|X)$$

We are currently working to prove this conjecture.

III. Application to Compression of Black and White Images

We start out with a black and white image of size 1728 by 2376 pixels. A resolution reduction algorithm is applied to this image, resulting in an image of size 864 by 1188. This image is scanned in a raster scan fashion to produce a sequence X. Each pixel in X is associated with four pixels in the original. The values of these four pixels constitute the sequence Y. The above algorithm for coding with side information is then applied to these two sequences. Applying this algorithm to the CCITT facsimile test documents resulted in an average compression ratio of 21 : 1.

Acknowledgements

P. Subrahmanya acknowledges helpful discussions with Prof. Jacob Ziv and Dr. Dharmendra Modha.

References

[1] International Telephone and Telegraph Consultative Committee (CCITT), "Progressive Bi-level Image Compression," *Recommendation* T.82, 1993.

[2] A.D.Wyner and J.Ziv. "Asymptotic Optimality of the Sliding Window Lempel-Ziv Algorithm," *Proceedings of the IEEE* , 82(6):674–682, June 1994.

[3] P.Elias "Universal Codeword Sets and Representations of the Integers," *IEEE Transactions on Information Theory*, IT-21(2):194-202, March 1975.

[4] W.H.R.Equitz and T.M.Cover. "Successive Refinement of Information," *IEEE Transactions on Information Theory*, IT-37(2):269-275, March 1991.

On the Compression Dimension of Data Strings and Data Sets

John Kieffer and Greg Nelson[1]

Elec. Engr. Dept., University of Minnesota, Minneapolis, MN 55455, USA

Abstract — **A hierarchical lossless source code compresses data by means of a graph used to represent the data. We show that the hierarchical codes which perform best as the number of data samples grows have a compression performance that can be characterized via a notion of the dimension of the data which we call compression dimension.**

I. Data Representation via Graphs

Throughout this summary, we fix a finite set A as our source alphabet. Let x be a generic notation for a data string that can be formed from the symbols in A, whose length $|x|$ satisfies $2 \leq |x| < \infty$. The notation G shall be a generic notation for a finite directed acyclic rooted graph whose edges are ordered and whose terminal vertices are each labelled by a symbol from A. Each graph G gives rise to a data sequence x in a natural way, and we shall denote this state of affairs by the notation $G \to x$. We indicate how this is done with an example.

Example. We define a graph G with nine edges ordered as $e_1, e_2, \ldots, e_9$, and six vertices denoted $v_0, v_1, \ldots, v_5$, where v_0 is the root vertex and v_4, v_5 are the terminal vertices. Edges e_1, e_2 lead from v_0 to v_4; e_3 leads from v_0 to v_1; e_4, e_5 lead from v_1 to v_2; e_6, e_7 lead from v_2 to v_3; and e_8, e_9 lead from v_3 to v_5. Vertices v_4, v_5 are labelled with the symbols $0, 1$, respectively. (The alphabet A is $\{0, 1\}$ in this example.) Starting with the sequence $e_1e_2e_3$ of ordered edges emanating from the root vertex, we perform the following steps:

(1) $e_1e_2e_3 \to v_4v_4v_1$
(2) $v_4v_4v_1 \to v_4v_4e_4e_5$
(3) $v_4v_4e_4e_5 \to v_4v_4v_2v_2$
(4) $v_4v_4v_2v_2 \to v_4v_4e_6e_7e_6e_7$
(5) $v_4v_4e_6e_7e_6e_7 \to v_4v_4v_3v_3v_3v_3$
(6) $v_4v_4v_3v_3v_3v_3 \to v_4v_4e_8e_9e_8e_9e_8e_9e_8e_9$
(7) $v_4v_4e_8e_9e_8e_9e_8e_9e_8e_9 \to v_4v_4v_5v_5v_5v_5v_5v_5v_5v_5$

In the odd numbered steps, the sequence on the right is obtained by replacing each edge in the sequence on the left with the vertex to which that edge leads. In the even numbered steps, the sequence on the right is obtained by replacing each non-terminal vertex in the sequence on the left with the string of ordered edges emanating from that vertex. The final sequence on the right in (7) consists entirely of terminal vertices; to obtain the data string x such that $G \to x$, one replaces each terminal vertex in this final sequence by the label for that vertex. We see in this case that $x = 0011111111$.

II. Coding Problem for Hierarchical Codes

If G is a graph, we let $v(G), e(G)$ denote the number of vertices and the number of edges in G, respectively. In the following, all logarithms are to base two. For the purposes of this summary, we define a lossless source code α to be a one-to-one map which assigns to each data string x a binary codeword $\alpha(x)$. Informally, we want to think of a hierarchical code as a lossless source code in which the codeword for x is generated incrementally as one traverses the edges of some graph $G_x \to x$, each edge (or group of edges) contributing at least one bit to the codeword. Formally, we define a lossless source code α to be hierarchical if there exist positive real constants C_1, C_2 such that for each x,

$$C_1 e(G_x) \leq |\alpha(x)| \leq C_2 e(G_x) \log v(G_x)$$

for some graph $G_x \to x$. For example, finite-state sequential codes, the Lempel-Ziv code, and bintree codes are hierarchical by this definition. We are interested in the problem of characterizing those hierarchical lossless source codes α for which the codeword length $|\alpha(x)|$ grows most slowly as $|x| \to \infty$. Specifically, we characterize those hierarchical lossless source codes for which the "logarithmic compression rate" $\log|\alpha(x)|/\log|x|$ is minimized as $|x| \to \infty$. In the next section, we introduce the concept of compression dimension to solve our problem.

III. Solution via Compression Dimension

We define the compression dimension $\mathrm{Dim}(x)$ of the data string x as the ratio $\log e(G_x)/\log|x|$, where G_x is a graph with the minimal number of edges for which $G_x \to x$. If α is a hierarchical lossless source code, define $\mathrm{Dim}(x|\alpha)$ to be the ratio $\log|\alpha(x)|/\log|x|$.

Let S denote a data set consisting of infinitely many data strings x. We define the compression dimension $\mathrm{Dim}(S)$ of S to be the limit supremum of $\mathrm{Dim}(x)$ as $|x| \to \infty$ through members of S. If α is a hierarchical lossless source code, define $\mathrm{Dim}(S|\alpha)$ to be the limit supremum of $\mathrm{Dim}(x|\alpha)$ as $|x| \to \infty$ through members of S.

Theorem 1. For any hierarchical lossless source code α,
(i) $\liminf_{|x|\to\infty} \mathrm{Dim}(x|\alpha)/\mathrm{Dim}(x) \geq 1$.
(ii) $\mathrm{Dim}(S|\alpha) \geq \mathrm{Dim}(S)$ for any S.

Theorem 2. There exists a hierarchical lossless source code α^* such that
(i) $\lim_{|x|\to\infty} \mathrm{Dim}(x|\alpha^*)/\mathrm{Dim}(x) = 1$.
(ii) $\mathrm{Dim}(S|\alpha^*) = \mathrm{Dim}(S)$ for any S.

Remarks.
(1) Parts (i)-(ii) of Theorem 2 do not hold if α^* is the Lempel-Ziv code.
(2) There are several useful bounds on $\mathrm{Dim}(x)$, which we shall discuss in another work.
(3) For more on hierarchical lossless source codes, see [1].

Acknowledgements

The authors thank Dr. Xiaolin Wu and Dr. En-hui Yang for helpful discussions concerning this research.

References

[1] G. Nelson, J. Kieffer, and P. Cosman, "An interesting hierarchical lossless data compression algorithm," *Proceedings of the 1995 IEEE Information Theory Society Workshop (Rydzyna, Poland).*

[1]This work was supported by NSF Grant NCR-9304984

Optimal linear receivers for synchronizing pseudo random sequences

Anand G. Dabak
P.O. Box 655-474, M.S. 446, Texas Instruments, Dallas, TX 75082, USA
dabak@hc.ti.com

I. Introduction

Much research work has been done in finding sequences with good autocorrelation properties. The conventional receiver structure for synchronizing with such a sequence transmitted periodically is a matched filter matched to the sequence. However, synchronization performance of the receiver can be further improved if the receiver tries to *match* to the sequence and *dematch* the circular shifts of the sequence. In [1], the authors solve the problem of finding the optimal receiver for synchronization in satellite systems, when the preamble is preceded by *random* data. However, the problem at hand is different since the data preceding the synchronization sequence is known. We derive the optimal *linear* receiver to synchronize with a given sequence and demonstrate the synchronization gain achieved by employing such a receiver over the conventional receiver. This gain is obtained by simply changing the receiver coefficients, leaving the receiver structure *unchanged.*

II. The optimal linear receiver

Let $S = \{s_0, s_1, s_2, ..., s_{n-1}\}$; $s_i \in [-1, 1]$ be a sequence and $S_K = \{s_K, s_{1+K}, s_{2+K}, ..., s_{n+K-1}\}$ denote a circular shift of S by K. All the additions in the above equation are *modulo n.* Let $r_{ss}(K) = \frac{\langle S_K, S\rangle}{\sqrt{n}}$ denote the autocorrelation of S and $r_{ss}^{max} = \max_{K\in[1,..,n-1]} r_{ss}(K)$ be its maximum off peak auto correlation. Assume that the sequence S is being transmitted periodically and the receiver is trying to synchronize with it. This can be modeled by letting the received sequence be $\{r_i;\ i = -\infty, .., 0, ..\}$, where $r_i = s_{i'} + n_i$; n_i are samples of white Gaussian noise process and $i' = (i - k')$*modulo n.* This problem occurs in the synchronization of spread spectrum systems [2] and in CDMA systems[4]. For a sequence $X = \{x_0, x_1, .., x_{n-1}\}$; $x_i \in \mathbb{R}$, $\langle X, X\rangle = 1$, consider the *linear* receiver; $L_x(k) = \langle S_{k-k'}, X\rangle + N_k$. Substituting $x_i = \frac{s_i}{\sqrt{n}}$ gives us the conventional receiver, which matches the incoming data to the sequence S. A measure of goodness of the code S is the difference $r_{ss}(0) - r_{ss}^{max} = \sqrt{n} - r_{ss}^{max}$. The larger this difference, we can see that in an additive white Gaussian noise scenario the better is the estimate of k'. Without loss of generality, let $r_{SX}(0)$ denote the maximum correlation between the sequence S and X and let $r_{SX}^{max} = \max_{K\in[1,..,n-1]} r_{SX}(K)$. Now consider the following optimization problem;

Problem 1

$$\max_{X\in\mathbb{R}^n} \{r_{SX}(0) - r_{SX}^{\max}\};\ \|X\| = 1$$

$$\Rightarrow \max_{X\in\mathbb{R}^n} \min_{K\in[1,..,n-1]} \langle S - S_K, X\rangle; \|X\| = 1 \qquad (1)$$

Geometrically speaking, equation (1) finds that unit vector $X \in \mathbb{R}^n$ which is closest to the collection of vectors $\{S - S_K; K = 1, .., n-1\}$, in the sense that the minimum of the projections of the vectors $S - S_K$ on X is maximized. We now give the solution to the optimization problem.

Proposition 1 There $\exists \tilde{X} \in \mathbb{R}^n$, $\tilde{X}$ not necessarily a unit vector satisfying the following conditions

1. It is a linear combination of the vectors $\{S - S_K;\ K = 1, .., n-1\}$ that is; $\tilde{X} = \sum_{K=1}^{n-1} \alpha_K (S - S_K)$.
2. The solution lies within the *cone* of the vectors $\{S - S_K;\ K = 1, .., n-1\}$, implying that the coefficients $0 \leq \alpha_K \leq 1$.
3. Let each $\alpha_K \neq 0$ be denoted by a variable $\tilde{\alpha}_l$. The collection $\{\tilde{\alpha}_l;\ l = 1, ..m\}$, $m \leq n-1$ thus denotes the set of all the non-zero α_K's. Let $\tilde{S}_l$ represent the vector S_K corresponding to $\tilde{\alpha}_l$. Then, the vectors $\{\tilde{S}_l;\ l = 1, .., m\}$ are *linearly independent.*
4. Finally, let $\gamma = \min_{K=1,..,n-1} \langle S - S_K, \tilde{X}\rangle$, then $\forall l\ \langle S - \tilde{S}_l, \tilde{X}\rangle = \gamma$

The vector $X_{opt} = \frac{\tilde{X}}{\|\tilde{X}\|}$ *uniquely* solves the optimization problem.

III. Examples

Consider the length 31 Gold sequences, there are 33 in all [3]. Two of these are the pseudo random sequences. Let S denote the remaining 31 Gold sequences. The $\sqrt{n}(r_{SS}(0) - r_{SS}^{max})$ for these sequencs can be seen [3] to be $(31 - 7) = 24$. On the other hand, computing the optimal vector X for these sequences, the $\sqrt{n}(r_{SX}(0) - r_{SX}^{max})$ turns out to be 26.3, 26.4, 26, 27, 27, 27, 27, 25.9, 25.6, 25.6, 26.4, 25.6, 25.7, 26.4, 26.5, 25.6, 25.7, 25.7, 25.8, 26.4, 25.8, 26.5, 26.9, 25.6, 26.4, 25.7, 26.5, 25.6, 25.7, 25.7, 26.9. The gain in terms of the signal to noise ratio can be seen to be between $20log\left(\frac{25.6}{24}\right) = 0.56$ dB to $20log\left(\frac{27}{24}\right) = 1$ dB. This gain is also confirmed by plotting the probability of false alarm against the signal to noise ratio for the conventional and the optimal linear detectors. Since the auto correlation function of the pseudo noise sequences is a delta function, we do not expect to get much gain for pseudo noise sequences.

We can thus conclude that for sequences with large off peak correlation values, like the Gold and Kasami sequences, substantial synchronization gain can be achieved by simply changing the matched filter coefficients of the receiver. This gain is expected to decrease as the length n of the sequence employed increases. However, for applications involving short length pseudo random sequences, the gain could be significant.

References

[1] J. L. Massey, "Optimum frame synchronization.", *IEEE Trans. Comm. Tech.*, 115:118, Apr., 1972.

[2] Andreas Polydoros & Charles Weber, "Worst case considerations for coherent serial acquisition of pseudo noise sequences", *Proceedings of NTC'80*, 24.6.1-24.6.4, Houston, Texas.

[3] Dilip V. Sarwate & Michael B. Pursley, "Acquisition in direct-sequence spread spectrum communication networks: An asymptotic analysis.", *IEEE Trans. on Information Theory*, 39:903-912, 1993

[4] Upamanyu Madhow & Michael B. Pursley, "Acquisition in Direct-Sequence Spread Spectrum Communication Networks: An Asymptotic Analysis", *IEEE Transactions on Information Theory*, vol.39, no. 3, pp. 903-912, 1993.

Maximum Likelihood Synchronization and Frequency Measurements

O. M. Collins and Z. Zhuang
University of Notre Dame
South Bend, IN 46556

All coherent communications systems need a way to regenerate the transmitter's carrier at the receiver. Most current communications systems do not send short messages and so this regeneration is performed by a phase locked loop.(Ref. 6) However, many communications systems now being considered, e.g., networks of small instruments on Mars or large deployments of free floating oceanographic sensors, will transmit short messages separated by long periods of transmitter shutdown.

The first section of this talk presents a maximum likelihood algorithm for estimating the phase and frequency of a carrier coming from one of these burst transmitters, when the carrier is observed against a background of white Gaussian noise and for enough time for the variance of the maximum likelihood estimate to be low. The algorithm avoids the "uncountable infinity of devices" which caused Reference 1 to conclude the maximum likelihood algorithm was "clearly unrealizable". The talk then analyzes the performance of this algorithm and, in the process, not only provides a much more compact proof of some of the classic results in Reference 1 through 4, but also strengthens them.

Simulations of the algorithm's operation at a moderate signal ratio are compared with the high SNR bound. The talk then outlines a similar algorithm to estimate the phase and frequency of the decaying sinusoids characteristic of physical measurements. Both algorithms have important roles in making physical measurements, e.g., the proton's gyromagnetic ratio. (Ref. 5)

Although the algorithms are truly maximum likelihood only for asymptotically high signal to noise ratios, their performance is imperceptibly different from this bound at all SNR's of practical interest, i.e., having a good algorithm for phase and frequency estimation is not helpful, if the data is insufficient for a fairly accurate measurement to be made. The input to both algorithms is a coarse frequency estimate and a set of samples representing a digitized segment of spectrum. The coarse frequency estimate can easily be generated by taking an FFT of the samples and determining the largest bin. For convenience in the derivations, the talk assumes that the amplitude of the constant sine wave is known and that both the amplitude and decay constant of the decaying sine wave also are known. This knowledge is, however, not necessary for either algorithm.

The talk concludes with a short section contrasting the two maximum likelihood algorithms with conventional frequency measurement techniques. This final section demonstrates that the results in this talk can greatly improve most frequency measurements that are signal to noise ratio limited .

1) A. J. Viterbi, *Principles of Coherent Communication*, McGraw Hill Book Company, 1966

2) C. W. Helstrom, *Statistical Theory of Signal Detection*, New York, Pergamon Press, 1960

3) P. M. Woodward, *Probability and Information Theory*, with Applications to Radar, London, Pergamon Press Ltd, 1953

4) Rife, D. C. and Boorstyn, R. R., "Single-Tone Parameter Estimation from Discrete-Time Observations", *IEEE Transactions on Information Theory*, Vol. IT-20, No. 5, September 1974, pp.591-598

5) L. Marton, *Advances in Electronics and Electron Physics*, Vol. 23, New York, Academic Press, 1967

Multilevel Coding to Combat Quantization of the Sum of the Transmitted Signal, a Noise and a Known Interference

Hanan Herzberg and Burton R. Saltzberg
AT&T Bell Laboratories, 200 Laurel Avenue, Middletown, NJ 07748

Abstract — **Encoding and decoding schemes, presented in this paper, are aimed at enabling transfer of data through a channel in which two types of interference are added to the transmitted signal and the sum is quantized. One of these interferences is known (note that the input of the quantizer is not accessible), whereas the second is an AWGN. An upper bound on the error rate, contributed by the component codes of a multilevel code, has been developed for multistage decoding.**

Summary

The encoding and decoding schemes presented in this paper are aimed at enabling transfer of data through a channel that is different from the conventional additive white Gaussian noise (AWGN) channel. Here we are interested in a channel where two types of interference are added to the transmitted signal and the sum is quantized. One of these interferences is known (or can be approximated), denoted by α, whereas the second is an AWGN, denoted by n. Since the input of the quantizer is not accessible, the known interference can not be removed from the received signal. We will show that the error rate for an uncoded transmission through this channel is unacceptably large, even for low noise levels and linear quantization. It will be shown that the problem becomes even more severe when a non-linear quantization is present. Therefore, coding is essential and huge coding gain is achievable in this application.

Investigation of the uncoded error events leads to the conclusion that a multilevel coding (see e.g., [1]) is an efficient solution. Note that coded-modulation structures, including multilevel coding schemes, have been designed mostly for AWGN channels. The component codes of a multilevel code, employed over an AWGN channel, should be selected such that the minimum Euclidean distance between the transmitted sequences would increase. However, this design rule is not applicable for our channel.

We derived a new metric, required for maximum likelihood decoding of data received over the foregoing channel. However, an important parameter of a coded system is the computational complexity of the decoder. The multistage decoder (see e.g., [1]) is an efficient scheme for decoding multilevel codes. The decoder employs a separate binary decoder for each component code. In order to decode a component code, maximum likelihood decoding is performed under the assumption that the bits related to higher partition levels are uncoded, and that the data transferred from the decoders for the codes related to the lower partition levels are correct. However, it can be shown that the reduction in the coding gain, due to multistage decoding, is very small, whereas the reduction in complexity is substantial.

Let r_j be one of the quantization levels. Let s_i be the $i-th$ element in the alphabet of the transmitted symbols. Let s_{i_0} and s_{i_1} be the closest symbols to the value $r_j - \alpha$, where i_0 and i_1 are even and odd labels, corresponding to the least significant bit of the symbol being 0 or 1, respectively. Note that if the $k-th$ input to the receiver is r_j, the $k-th$ output of the decoder for C_1 (the code related to the first partition level) should be either s_{i_0} or s_{i_1}. The metric related to the symbol s_i is $\log(Pr[r_j|s_i,\alpha])$, where s_i can be substituted by s_{i_0} or s_{i_1}. The even/odd characteristic of the sequence at the output of this decoder should construct a codeword in C_1, which is applied to the least significant bits of the symbol sequence. The decoding of the other component codes, corresponding to other bits in the symbol label, is performed in a similar fashion.

Let $\delta_{ij\alpha}$ be the distance between the sum $s_i + \alpha$ and the threshold of the decision region of r_j, where a negative value of $\delta_{ij\alpha}$ indicates that the sum is outside the decision region. The conditional probability of the channel output is given by

$$Pr[r_j|s_i,\alpha] = Q(-\delta_{ij\alpha}/\sigma),$$

where σ^2 is the variance of n and $Q(x) \equiv \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy$. Based on the conditional probability it can be shown that the metrics corresponding to a maximum likelihood decoding of a component code is a function of σ^2. However, in many cases the level of the noise is unknown. Therefore, a metric for a suboptimal decoding, in which the noise level is not required, has been derived. It can be shown that for moderate, as well as low, noise levels the performance of the suboptimal decoding is very close to that of the maximum likelihood decoding. The latter statement is supported by computer simulation.

An upper bound on the error rate at the output of the component code's decoder was derived. For instance, let the component code C_1 be a convolutional code. Let $\{a_d\}$ be the set of error coefficients, used for evaluating the bit error rate of a convolutional code [2]. It was proved that the average bit error rate, contributed by the decoder for C_1, is bounded by the following upper bound

$$P_b \le \sum_d a_d D^d,$$

where $D =$

$$\int_{-\infty}^{\infty} f_\alpha(\alpha) \sum_j \sqrt{\sum_{i_0} Pr[s_{i_0}]Q(\frac{\delta_{i_0 j\alpha}}{\sigma}) \sum_{i_1} Pr[s_{i_1}]Q(\frac{\delta_{i_1 j\alpha}}{\sigma})}\, d\alpha,$$

i_0 and i_1 are even and odd labels, respectively, and $f_\alpha(\alpha)$ is the probability density function of the interference α. Note that D can easily be evaluated numerically.

References

[1] G. J. Pottie and D. P. Taylor, "Multilevel codes based on partitioning," *IEEE Trans. Information Theory*, vol. IT-35, pp. 87-98, January 1989.

[2] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, New York: McGraw-Hill, 1979.

Probability Distribution of Differential Phase Perturbed by Tone Interference and Its Application

Mao ZENG and Qiang WANG
Dept. of Elect. & Comp. Eng., University of Victoria
Victoria, B.C., Canada V8W 3P6

Abstract

Differential phase shift keying (DPSK) is widely used in communication systems where simplicity and robustness are desired. One such system is slow frequency hopped DPSK (SFH/DPSK) which can sustain a much higher data rate than a fast frequency hopped system while having the same hop rate.[1], [2]

In the detection of SFH/DPSK, differentially coherent detection is often employed. This is because it is impossible to maintain the phase coherence between different hops. Differentially coherent detection can take advantage of phase coherence within a hop and thus outperforms noncoherent detection. In this paper we present a study of the probability distribution of a received differential phase perturbed by tone jamming and Gaussian noise. The intent is to study the effects of jamming against SFH/DPSK and to provide some tools for the analysis such a system.

In much previous work, the performance of SFH/DPSK has been considered[1-5]. Simon[4] has analyzed the performance of SFH/DPSK under multiple continuous tone jamming for a specific set of signal phases and equally spaced decision regions. The analytical results were obtained by ignoring the system thermal noise so that the derivation relied largely on geometric relation. Gong analyzed the performance of a specific binary SFH/DPSK scheme in both tone and noise interference[3]. In [1], [2], Wang, *et al*, presented a method to derive the general probability distribution for arbitrary DPSK signals. In this paper, we give an alternative but simple expression of the general probability distribution of a received differential phase corrupted by continuous tone jamming and Gaussian noise.The probability distributions of the received differential phase corrupted by either continuous tone jamming or Gaussian noise is the special form of it. Thus this result is a generalization of the previous well known results by Pawula, Rice and Roberts[5].These results are derived by making use of an unconventional approach which relates the desired probability to a functional of the joint characteristic function of narrowband waveform. Our starting point is the basic relation

$$\Gamma(\Theta) = \int_0^\infty \int_0^\infty \frac{\partial}{\partial\Theta} J_0\left(\sqrt{x^2+y^2+2xy\cos\Theta}\right)\frac{1}{xy}dxdy \quad (1)$$

where $\Gamma(\Theta)$ is a periodic sawtooth function of period 2π defined as

$$\Gamma(\Theta) = \Theta \qquad -\pi < \Theta \le \pi$$

Then the relation between the joint characteristic function and the probability $P\{\psi_1 \le \psi \le \psi_2\}$ (see [5] for definition) can be derived. The final result can be given by a simple expression in terms of Marcum's Q-function as follows.

$$P\{\psi_1 \le \psi \le \psi_2\} = G(\psi_2) - G(\psi_1) \quad (2)$$

The auxiliary function $G(\psi)$ has the form:

$$G(\psi) = \frac{\psi}{2\pi} - \frac{1}{4\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin(\Delta\Phi-\psi)}{T} q\left(\sqrt{\frac{\sigma_N}{\sigma_J}}S, \sqrt{\sigma_N T}\right) - \frac{\sin\psi}{S} q\left(\sqrt{\sigma_N T}, \sqrt{\frac{\sigma_N}{\sigma_J}}S\right) d\theta \quad (3)$$

where

$$S = 1 - \cos\theta\cos\psi; \qquad T = 1 - \cos\theta\cos(\Delta\Phi - \psi)$$

$$q(a,b) = 1 - Q(a,b) \quad (4)$$

To illustrate the application of these results, we analyze the error probability performance of a general uncoded SFH/DPSK signal under worst case tone jamming and Gaussian noise. Skewed differential phases with unequal decision regions and the error performance when a frequency offset between the jamming tone and DPSK carrier is also considered.

Reference

[1] Q. Wang, T. A. Gulliver, and V. K. Bhargava, "Performance of SFH/MDPSK in Tone Interference and Gaussian Noise", *IEEE Trans. Commun.*, vol. COM-42, Feb/Mar/Apr, 1994, pp. 1450-1454

[2] Q. Wang, T. A. Gulliver, and V. K. Bhargava, "Probability Distribution of DPSK in Tone Interference and Applications to SFH/DPSK", *IEEE Journal on Selected Areas in Communications,* vol. 8, June 1990, pp. 895-906

[3] K. S. Gong, "Performance analysis of FH/DPSK in additive white Gaussian noise and multitone jamming", in *Proc. IEEE MILCOM*, 1988, pp. 53.4.1-53.4.7.

[4] M. K. Simon, *et al, Spread Spectrum Communications,* Volume II, Computer Science press, 1985

[5] R. F. Pawula, S. O. Rice and J. H. Roberts, "Distribution of the Phase Angle Between Two Vectors Perturbed by Gaussian Noise", *IEEE Trans. Commun.*, vol. COM-30, Aug., 1982

Cyclotomic Cosets and Steady State Solutions to a Dynamic Jamming Game

Ranjan K. Mallik

Dept. of E & ECE, Indian Institute of Technology, Kharagpur 721302, West Bengal, India

Robert A. Scholtz

CSI, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565, U.S.A.

Abstract — **For an on-off slotted dynamic jamming game, a relation between the steady state solutions and cyclotomic cosets is established.**

I. The Dynamic Jamming Game Model

An on-off slotted communication jamming game is modeled as a *two-person zero-sum non-cooperative dynamic game* [1] played over T uniform time slots between a *communicator* (a transmitter-receiver pair) and a *jammer*. In slot t, the *communicator's power level* X_t is randomly distributed over $0, P$, and the *jammer's power level* Y_t over $0, J$ ($P, J > 0$). Each player has knowledge of its own and the opponent's previous plays. The *payoff function* $\mathcal{J}$ is given by $\mathcal{J} \triangleq \frac{1}{T}\sum_{t=1}^{T} E[f(X_t, Y_t)]$, where $f(X_t, Y_t)$ is the normalized payoff to the communicator.

Let Z_t be a measure of the *communicator's past energy accumulation at the beginning of slot* t, and δ_1 $(0 < \delta_1 < 1)$ the *communicator's thermal memory constant*. The relation $Z_t = X_{t-1} + \delta_1 Z_{t-1}$, for $t = 2, \ldots, T$, with $Z_1 = 0$, holds. There is a *power constraint* $X_t + \delta_1 Z_t \leq \hat{P}$ for all t. The jammer is subject to similar constraints determined by analogous quantities W_t, δ_2 and $\hat{J}$. The transmitter parameters $\frac{\hat{P}}{P}$, δ_1, $\frac{\hat{J}}{J}$, δ_2 and the payoff matrix are known to both players. The *strategies* are defined as

$$p_t(x|z,w) \triangleq \text{Prob}(X_t = x | Z_t = z, W_t = w)\,, \tag{1a}$$

$$q_t(y|z,w) \triangleq \text{Prob}(Y_t = y | Z_t = z, W_t = w)\,. \tag{1b}$$

The *optimal strategies* can be found by *dynamic programming*. Denoting $S_t^*(z,w)$ to be the *optimum accumulated payoff at time t given the past energy accumulations z and w*, we obtain an evolution equation which gives $S_{t-1}^*(z,w)$ in terms of $S_t^*(z,w)$ for $t = T$ down to 2.

II. $M \times N$ Grid Solutions

On the (z,w) plane $S_T^*(z,w)$ changes its value at most once along the z-axis at $z = \frac{\hat{P}-P}{\delta_1}$, and along the w-axis at $w = \frac{\hat{J}-J}{\delta_2}$. We call $\frac{\hat{P}-P}{\delta_1}$ a *critical point (c.pt.) on the z-axis*, and $\frac{\hat{J}-J}{\delta_2}$ a *c.pt. on the w-axis*. As time goes backward in the evolution equation, the operating points $(\delta_1, \frac{\hat{P}}{P})$ and $(\delta_2, \frac{\hat{J}}{J})$ for which the c.pts. of $S_t^*(z,w)$ do not increase indefinitely with reverse time give rise to *steady state solutions*.

Consider the communicator's case. Let $\mathcal{U}(T) = \left\{\frac{\hat{P}-P}{\delta_1}\right\}$ denote the *c.pt. set at time T along the z-axis*. Using the power constraints, an operator $\mathcal{O}$ which maps the communicator's energy accumulation z at time t to that at $t-1$ is defined as

$$\mathcal{O}(z) \triangleq \begin{cases} \frac{z}{\delta_1} & \text{if } 0 \leq z \leq P, \\ \frac{z-P}{\delta_1} & \text{if } P < z \leq \hat{P}. \end{cases} \tag{2}$$

Then $\mathcal{U}(t-1) = \mathcal{U}(T) \bigcup \mathcal{O}(\mathcal{U}(t)) \bigcap [0, \hat{P}]$. For a steady state solution with $M-1$ c.pts., we force the condition that $\mathcal{U}(t) = \{a_1, \ldots, a_{M-1}\} = \mathcal{U}(t-1)$ which gives the *c.pt. generation system*

$$\begin{aligned} a_{c_1} &= \mathcal{O}(\hat{P})\,, \\ a_{c_i} &= \mathcal{O}(a_{c_{i-1}})\,, \quad i = 2, \ldots, M-1. \end{aligned} \tag{3}$$

Here $[c_1, \ldots, c_{M-1}]$, the *c.pt. index vector*, is a permutation of $[1, \ldots, M-1]$. The *operating condition* is given by $\frac{a_{c_{M-1}}}{\delta_1} > \hat{P}$, $\frac{a_{c_{M-1}} - P}{\delta_1} < 0$. If the jammer also has $N-1$ c.pts. on the w-axis, then $S_t^*(z,w)$ will have a $M \times N$ grid structure on the (z,w) plane and the game will have a *steady state $M \times N$ grid solution*. The number of such solutions is related to cyclotomic cosets.

A full cyclotomic coset mod $(2^M - 1)$ can be written as a M-tuple $(\nu_1, \ldots, \nu_M)$, where $\nu_1 < \cdots < \nu_M$, or, alternatively, as $(\nu_M, \nu_{c_1}, \ldots, \nu_{c_{M-1}})$, where the *coset index vector* $[c_1, \ldots, c_{M-1}]$ is a permutation of $[1, \ldots, M-1]$ for which we obtain the *coset generation system*

$$\begin{aligned} \nu_{c_1} &= 2\nu_M \bmod (2^M - 1)\,, \\ \nu_{c_i} &= 2\nu_{c_{i-1}} \bmod (2^M - 1)\,, \quad i = 2, \ldots, M-1, \end{aligned} \tag{4}$$

which has the same form as (3). Let an operator $\mathcal{T}_M$ be defined as

$$\mathcal{T}_M(\nu) \triangleq \begin{cases} 2\nu & \text{if } 1 \leq \nu \leq \frac{2^M-2}{2}, \\ 2\nu - (2^M - 1) & \text{if } \frac{2^M}{2} \leq \nu \leq (2^M - 2). \end{cases} \tag{5}$$

Comparing (3) with (4) we find that the following isomorphisms hold for each index vector $[c_1, \ldots, c_{M-1}]$:

$$\mathcal{T}_M \longleftrightarrow \mathcal{O}\,, \quad (\nu_M, \nu_{c_1}, \ldots, \nu_{c_{M-1}}) \longleftrightarrow (\hat{P}, a_{c_1}, \ldots, a_{c_{M-1}})\,.$$

Thus *the number of c.pt. generation systems for any natural number $M \geq 2$ equals the number $h(M)$ of full cyclotomic cosets mod* $(2^M - 1)$ given by $h(M) = \frac{1}{M}\sum_{d|M} \mu(d) 2^{\frac{M}{d}}$, where μ is the *Möbius function of number theory*. The jammer's case is analogous.

For given M and N, there are $h(M)$ c.pt. generation systems for the communicator and $h(N)$ for the jammer. Therefore *the game has $h(M).h(N)$ different $M \times N$ grid solutions*. Since $h(2) = 1$, there is an unique 2×2 grid solution [2].

References

[1] R. K. Mallik, R. A. Scholtz, and G. Papavassilopoulos. A Simple Dynamic Jamming Game. In *ISIT Proceedings*, pp. 383, 1994.

[2] R. K. Mallik, R. A. Scholtz, and G. P. Papavassilopoulos. On the Steady State Solution of a Two-by-two Dynamic Jamming Game with Cumulative Power Constraints. In *ACSSC Proceedings, vol. 2*, pp. 888–892, 1991.

Duration of a Search for a Fixed Pattern in Random Data: the Distribution Function

Dragana Bajić[1], Dušan Drajić[1], Ognjen Katić[2]

[1]Faculty of Electrical Engineering,Bulevar Revolucije 73, Belgrade,Yugoslavia
[2]Glenayre R&D Inc, 1570 Kootenay Street, Vancouver, B.C, Canada

Abstract - **A probability distribution function of the duration of a search for a fixed pattern in random data is derived, in terms of bifix analysis of a pattern.**

I. Introduction

According to the classical well-known paper [1], the expected duration of a search for a fixed L-ary pattern of length n in a sequence of random L-ary equiprobable data is:

$$E\{x\} = \sum_{i=0}^{n} h_i \cdot L^i - n \qquad (1)$$

where h_i, $i = 0,\ldots,n$ represent bifix indicator with the following meaning: $h_i = 1$ if a bifix (a sequence that is both prefix and suffix) of length i exists; otherwise, $h_i = 0$ and by convention $h_0 = h_n = 1$. This formula is unavoidable for any research considering synchronization processes (e.g. [2, 3, 4, 5]).

This paper presents an extension to the research given in [1], as it gives the formula for probability distribution function upon which the expected duration and variance of the same process, as well as the higher moments, can be evaluated.

II. Results

The probability that the n-digit pattern will occur for the first time at the k^{th} position within the stream of random data equals to:

$$\Pr\{x = k\} = a_k \cdot p^{k+n-1} \qquad (2)$$

where a_k is expressed using a recursion:

$$a_k = \sum_{i=1}^{\min(n,k)} (L \cdot h_{n+1-i} - h_{n-i}) \cdot a_{k-i}. \qquad (3)$$

and where $p = 1/L$ is the probability of a random equiprobable digit.

Expression (2) is the probability distribution function so it satisfies the condition:

$$S\{x\} = \sum_{i=1}^{\infty} \Pr\{x = i\} = \sum_{i=1}^{\infty} a_i \cdot p^{i+n-1} = 1. \qquad (4)$$

Variance of a duration of a search for the fixed pattern in random data can be found by statistical methods:

$$\sigma^2 = \sum_{i=1}^{\infty} i^2 \cdot \Pr\{x=i\} - E^2\{x\} = (E\{x\}+n)\cdot(E\{x\}+n-1) - 2\cdot\sum_{i=0}^{n} i \cdot h_i \cdot L^i, \qquad (5)$$

while performing the summation $E\{x\} = \sum_{i=1}^{\infty} i \cdot \Pr\{x = i\}$, formula (1) is obtained.

The expressions (4) and (5) can be easily proven using (2) and (3).

The probability distribution functions for the bifix-free binary pattern 01011 and "all zeros" pattern of the same length (for which $h_i = 1$, $i = 0,\ldots,n$) is plotted in Fig. 1, dashed lines representing the simulation study, simulation being performed over the sample of 100000 searches.

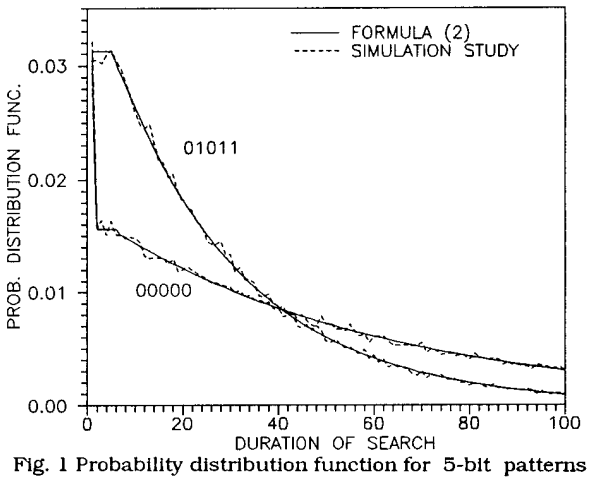


Fig. 1 Probability distribution function for 5-bit patterns

III. Conclusion

The probability distribution function derived in this paper might be useful for all the researches considering the search problem. For some previous researches it has been obtained either using the simulation study (e.g. [5]), or by visual inspection for each particular pattern (e.g. [6]).

References

[1] T. Nielsen: "On the Expected Duration of a Search for a Fixed Pattern in Random Data", *IEEE Trans. on Information Theory*, vol. IT-19, pp. 702-704, Sept. 1973.

[2] R. A. Sholtz: "Frame Synchronization techniques", *IEEE Trans. on Communications*, vol. COM-28, pp. 1204-1212, Aug. 1980.

[3] S. Patarsen and C. Georghiades: "Frame Synchronization for Optical Overlapping Pulse-Position Modulation Systems", *IEEE Trans. on Communications*, vol. 40. pp. 783-794, Apr.1992.

[4] C. Georghiades and D. Snyder: Locating Data Frames in Direct-Detection Optical Communication Systems, *IEEE Trans. on Communications*, vol. COM-32, pp. 118-723, Feb. 1984.

[5] M.N.Al-Subbagh and E. Jones: "Optimum Patterns for Frame Alignment", *IEE Proceedings*, vol. 135, part F, pp. 594-603, Dec. 1988.

[6] H. Haeberle: "Frame Synchronizing PCM Systems", Electrical Communications, pp. 280-287, Vol. 44, 1969.

New Codes with the Same Weight Distributions as the Goethals Codes and the Delsarte-Goethals Codes

Tor Helleseth, P. Vijay Kumar and Abhijit G. Shanbhag[1]

Department of Informatics, University of Bergen, Høyteknologisenteret, N-5020 Bergen, Norway and Communication Sciences Institute, EE-Systems, University of Southern California, Los Angeles, CA 90089-2565, USA

Abstract — **The Goethals code is a binary nonlinear code of length 2^{m+1} which has $2^{2^{m+1}-3m-2}$ codewords and minimum Hamming distance 8 for any odd $m \geq 3$. We construct new codes over Z_4 such that their Gray maps lead to codes with the same weight distribution as the Goethals codes and the Delsarte-Goethals codes.**

I. Introduction

Let Z_4 denote the integers modulo 4 and let R be a Galois ring of characteristic 4 with 4^m elements. The multiplicative group of units in R contains a unique cyclic group of order $2^m - 1$. Let β be a generator for this subgroup and let $\mathcal{T} = \{0, 1, \beta, \beta^2, \cdots, \beta^{2^m-2}\}$.

The Gray map ϕ is defined by $\phi(0) = 00$, $\phi(1) = 01$, $\phi(2) = 11$ and $\phi(3) = 10$. From any (n, M) code over Z_4 the Gray map gives in a natural way a binary $(2n, M)$ code. The Lee weight distribution of the code over Z_4 equals the Hamming weight distribution of its binary Gray map.

Let C_1 be the binary code defined by $C_1 = \phi(\mathcal{C}_1)$, where $\mathcal{C}_1$ is the linear code over Z_4 with parity-check matrix given by

$$H_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & \beta & \beta^2 & \cdots & \beta^{2^m-2} \\ 0 & 2 & 2\beta^3 & 2\beta^6 & \cdots & 2\beta^{3(2^m-2)} \end{bmatrix}.$$

Hammons, Kumar, Sloane, Calderbank and Solé [1], have shown that if m is odd, then C_1 is a nonlinear binary $(2^{m+1}, 2^{2^{m+1}-3m-2}, 8)$ code. This code has the same weight distribution as the Goethals code. They also showed that $\phi(\mathcal{C}_1^{\perp})$ is a Delsarte-Goethals code.

II. Main Results

The main result is to show that we can construct many codes $\mathcal{C}_k$ over Z_4 with the same Lee weight distribution as $\mathcal{C}_1$. In particular, this implies that the Hamming weight distribution of $C_k = \phi(\mathcal{C}_k)$ is the same as for C_1 and therefore identical to the weight distribution of the Goethals code. From the MacWilliams identities and from the results of Hammons, Kumar, Sloane, Calderbank and Solé [1] it follows that $\phi(\mathcal{C}_k^{\perp})$ has the same Hamming weight distribution as the Delsarte-Goethals code $\phi(\mathcal{C}_1^{\perp})$.

Theorem Let $m \geq 3$ be odd and $\gcd(k, m) = 1$. Then any code $\mathcal{C}_k$ with parity-check matrix H_k given by

$$H_k = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & \beta & \beta^2 & \cdots & \beta^{2^m-2} \\ 0 & 2 & 2\beta^{2^k+1} & 2\beta^{(2^k+1)2} & \cdots & 2\beta^{(2^k+1)(2^m-2)} \end{bmatrix}.$$

has the same Lee weight distribution as $\mathcal{C}_1$ (i.e., independent of k).

Sketch of proof. We give an explicit derivation for the weight distribution of these codes via exponential sums. In the following we will give a brief sketch of the main ideas behind the proof. It turns out to be natural to study the Lee weight distribution of $\mathcal{C}_k^{\perp}$.

The Lee weight of $a \in Z_4$ and the real part of i^a is related by $w_L(a) = 1 - \Re(i^a)$, where $i = \sqrt{-1}$. Hence, the Lee weight of $\mathbf{c} = (c_0, c_1, \cdots, c_{n-1}) \in Z_4^n$ is related to (the real part of) an exponential sum as follows

$$d_L(\mathbf{c}) = n - \Re(\sum_{t=0}^{n-1} i^{c_t}).$$

Let T be the trace mapping from the Galois ring R to Z_4. Let $\mathbf{c}(a, b)$ be a vector of length $n = 2^m$ indexed by $x \in \mathcal{T}$ such that

$$\mathbf{c}(a, b) = T(ax + 2bx^{2^k+1}), \quad a \in R, b \in \mathcal{T}.$$

Let $\mathbf{1}$ denote the all-one vector of length 2^m. Then

$$\mathcal{C}_k^{\perp} = \{u\mathbf{1} + \mathbf{c}(a, b) \mid \quad a \in R, b \in \mathcal{T}, u \in Z_4\}.$$

Let $u \in Z_4$, $a \in R$ and $b \in \mathcal{T}$ and define

$$S(a, b, u) = i^u \sum_{x \in \mathcal{T}} i^{T(ax + 2bx^{2^k+1})}.$$

The main part of the proof is to determine the values and the number of occurrences of each value for the real part of this exponential sum. It turns out that this distribution is independent of k when $\gcd(k, m) = 1$. Hence from the relation between the exponential sum and the Lee weight of the codewords in $\mathcal{C}_k^{\perp}$, we conclude that the Lee weight distribution is independent of k and coincides with the distribution for the Delsarte-Goethals codes that can be found in Chapter 15 in MacWilliams and Sloane [3].

In Kumar, Helleseth, Calderbank and Hammons [2] large families of quaternary sequences with good correlation properties were constructed from the codes $\mathcal{C}_1^{\perp}$. We can also construct families of quaternary sequences with similar properties from $\mathcal{C}_k^{\perp}$.

[1] This work was supported in part by the Norwegian Research Council under Grant Numbers 107542/410 and 107623/420 and the National Science Foundation under Grant Number NCR-9016077

References

[1] R. Hammons, P.V. Kumar, N.J.A. Sloane, R. Calderbank and P.Solé, "The Z_4-Linearity of Kerdock, Preparata, Goethals, and Related Codes," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 301–319, 1994.

[2] P.V. Kumar, T. Helleseth, R. Calderbank and R. Hammons, "Large Families of Quaternary Sequences with Low Correlation," to appear in the IEEE Trans. on Inform. Theory.

[3] F.J. MacWilliam and N.J.A. Sloane, The Theory of Error Correcting Codes, (North-Holland, New York, 1977).

A Cyclic [6,3,4] Group Code and the Hexacode Over GF(4)

Moshe Ran and Jakov Snyders[1]

Dept. of Electical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, Israel.

Abstract — **A [6,3,4] code H_6 over an Abelian group $\mathcal{A}_4$ with four elements is presented. H_6 is cyclic, unlike the [6,3,4] hexacode E_6 over GF(4). However, H_6 and E_6 are isomorphic when the latter is viewed as a group code. H_6 is the smallest member of a class of $[2k, k, 4]$ cyclic and reversible codes over $\mathcal{A}_4$.**

I. Summary

A group code C of length n over an Abelian group $\mathcal{A}$ is a subgroup of $\mathcal{A}^n$, the n–fold direct product of $\mathcal{A}$. The rate $k(C)$ is defined by $k(C) = \log_{|\mathcal{A}|}|C|$, where $|X|$ stands for cardinality. A group code C of length n with rate k and minimum Hamming distance d_H is called an $[n, k, d_H]$ code. A linear code C over a field F is also a group code over the additive group of F. It has been shown in [2] that many of the important structural properties of codes over F are associated with the additive and not the multiplicative group properties of F.

We present a [6,3,4] group code H_6 over $\mathcal{A}_4$ with $|\mathcal{A}_4| = 4$. Let $\mathcal{A}_4 = \{a, b, c, d\}$ be the additive group of GF(4), where a is the identity element. The elements of $\mathcal{A}_4$, are called *symbols*. For the purpose of describing some binary codes with the aid of E_6, we use various binary representations for symbols, e.g., $a = 0000$, $b = 0101$, $c = 0011$, $d = 0110$.

Let H_6 be the code that comprises the (symbolwise) cyclic shifts, and their sums, of $(cabbba)$. H_6 is obviously a $[6, 3, 4]$ group code, hence it is an MDS code. Consequently, every three coordinates in H_6 constitute an information set, whereby every three symbols occur exactly once (in any three fixed positions), every two symbols occur 4 times and every symbol $4^2 = 16$ times. H_6 is the smallest member of a class of $[2k, k, 4]$ cyclic and reversible codes over $\mathcal{A}_4$.

There is a unique formally self dual [6,3,4] code over GF(4) (see [1, pp. 301–303]), called hexacode and denoted E_6. No version of E_6 maps onto H_6 under any bijection f: GF(4) $\mapsto$ $\mathcal{A}_4$. In fact, no cyclic [6,3,4] code over GF(4) exists. Nonetheless, if E_6 is viewed as a group code then it is isomorphic to H_6. Since $(aaaaaa) \in H_6$, H_6 and E_6 have identical coset Hamming weight distributions.

Some properties of H_6 are the following.

1) H_6 is invariant under replacement of a by d and b by c.

2) H_6 is invariant under cyclic permutation.

3) H_6 is invariant under reversal of the symbols.

4) H_6 is representable by a 4-section 16-states non-symmetric trellis diagram, and also by a 3-section 16-states symmetric trellis diagram.

5) Let

$$\begin{aligned} C_0 &= \{aaaaaa, cccdbd, abccba, cdabad\} \\ C_0' &= \{aaaaaa, bccbaa, ddbcaa, cbddaa\} \\ C_0'' &= \{aaaaaa, aabccb, aacbdd, aaddbc\} \end{aligned}$$

and $C_1 = C_0' + C_0''$. Then C_1 is [6,2,4] code. We have $H_6 = C_0 + C_1$ and, using standard notations for group partitioning,

$$H_6 = [C_2/C_1] + [C_1/C_0''] + C_0'' = C_0 + C_0' + C_0''.$$

6) H_6 consists of the blocks of $3-(24, 6, 1)$ constrained design (see [3]). H_6 can also be represented as the union of four $2-(12, 3, 1)$ constrained designs. (A constrained design may exit for parameters values for which no conventional t-design exits. In particular, neither a 3-(24,6,1) nor a 2-(12,3,1) t-design exits.)

For E_6 representations similar to those of 4) – 6) apply. Also, there exists a self-dual version of E_6, for which a property similar to 1) holds. However, properties 2) and 3) are unshared by (all versions of) E_6.

We present several constructions for binary codes of length 24 derived from H_6. In particular, the MOG (Miracle Octad Generator) construction, by which the [24,12,8] Golay codewords are described as some set of binary images of E_6, applies also with E_6 replaced by H_6.

An approach to fast maximum likelihood decoding of some binary codes of length 24 may be based on H_6. The binary codewords are regarded as images of H_6. Let $\mathbf{z} \in \mathcal{A}^n$ be the vector obtained by symbol–by–symbol soft decoding. The neighborhood of $\mathbf{z}$ is examined in order to identify the most likely codeword. A small list of candidate codewords is prepared by employing certain elimination rules. A substantial reduction of computational complexity is achieved in maximum likelihood soft decoding by intensively exploiting the structure of H_6 in the decoding procedures.

[1] This work was supported in part by the Israel Science Foundation administrated by the Israel Academy of Science and Humanities.

References

[1] J.H. Conway, and N.J.A. Sloane, *Sphere Packing Lattices and Groups,* New York: Springer–Verlag, 1988.

[2] G.D. Forney Jr. and M.D. Trott, "The dynamics of group codes: state space, trellis diagram, and the canonical encoders," *IEEE Trans. Inform. Theory,* vol. IT–39, pp. 1491–1513, Sept. 1993.

[3] M. Ran and J. Snyders, "Constrained designs for maximum likelihood soft decoding of RM(2,m) and the extended Golay codes," *IEEE Trans. Communications*, to be published.

Decoding Binary Expansions of Low-Rate Reed-Solomon Codes Far Beyond the BCH Bound

Charles T. Retter
U.S. Army Research Laboratory, AMSRL-IS-TP, Aberdeen Proving Ground, MD 21005

Abstract — **Binary expansions of low-rate Reed-Solomon codes typically are capable of correcting far more binary errors than guaranteed by the BCH bound on the Reed-Solomon code. Practical decoding algorithms that often correct beyond the true minimum distances of the binary codes are described.**

I. Minimum Distances

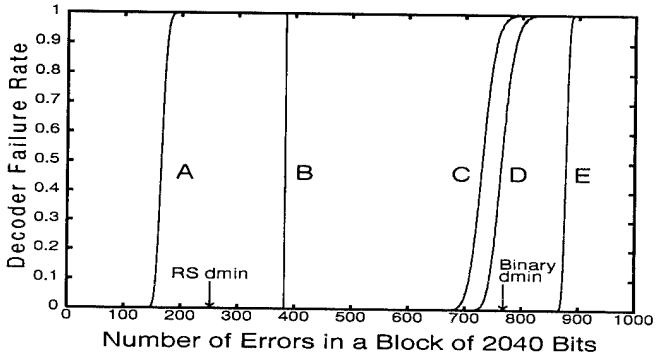
The minimum distance of an (N,K) Reed-Solomon code is given exactly by the BCH bound (N+1–K). Since low-rate Reed-Solomon codes have no binary subfield-subcodes, provided that 1 is a root of the generator polynomial, the binary expansions of many of these codes have surprisingly high minimum distances. To explore the properties of these codes, a large number of weight distributions were computed by generating codewords on a KSR1 supercomputer. All Reed-Solomon codes with parameters (31,7), (63,6), (127,5), and (255,4) were expanded using all normal bases. Then the most promising codes with parameters (31,8), (63,7), (127,6), and (255,5) were examined. A total of 3064 codes containing almost 50 trillion codewords were generated.

RS Codes	Binary Codes	Worst d_{min}	Average d_{min}	Best d_{min}	BCH Bound
(31,7)	(155,35)	40	40.944	44	25
(63,6)	(378,36)	84	123.690	136	58
(127,5)	(889,35)	320	359.405	368	123
(255,4)	(2040,32)	680	863.402	920	252
(31,8)	(155,40)	32	32.000	32	24
(63,7)	(378,42)	128	128.000	128	57
(127,6)	(889,42)	352	352.000	352	122
(255,5)	(2040,40)	884	884.000	884	251

II. Decoding

A conventional decoder for these codes would map binary m-tuples into symbols in GF(2^m) and decode using one of the standard Reed-Solomon decoding algorithms. This approach will decode correctly only if the number of symbol errors is less than (N+1-K)/2. Although not every bit error becomes a symbol error, this approach cannot take advantage of the true capabilities of these binary codes.

At AAECC-3, Bossert and Hergert [1, 2] suggested a simple approach to decoding linear codes, based on a very large syndrome formed by using all of the minimum-weight codewords in the dual as parity checks. The observation that the weight of this large syndrome increases with the number of errors suggests various simple algorithms to search for the nearest codeword by reducing the weight of the syndrome. The most important requirement for this algorithm is that the dual must contain a large number of codewords with very low weights, which is the case for the codes described above. For example, the weight distributions of the duals of 2304 binary (2040,32) codes were computed, and 99% of the duals were found to have $d_{min} = 5$, with one code having 1142808 words of that weight. However, a direct implementation of the Bossert-Hergert algorithm tends to stop at local minima when all of the minimum-weight words are used as checks. By varying the set of checks on successive passes, the local minima can be avoided. Simulations of several variations of this modified algorithm showed that far more errors can be corrected than with conventional Reed-Solomon decoders. The figure below shows the failure rates of five decoders for the code described above ($d_{min} = 768$).



A A conventional Reed-Solomon decoder, which fails 50% of the time with 165 bit errors.

B A hypothetical binary bounded-distance decoder, which fails with 384 or more errors.

C A four-pass threshold decoder, which starts by using all weight-5 checks, but changes the set of checks to avoid local minima. Its 50% failure rate occurs with 730 errors. Since the code is quasi-cyclic, a hardware implementation of this decoder is reasonably simple and very fast.

D A similar decoder, which changes only the best bit on each of 1500 passes. The 50% failure rate for this decoder is reached at 763 errors.

E A maximum-likelihood decoder, which has a 50% failure rate at 878 errors.

References

[1] Bossert, M., and G. Hergert, "A Decoding Algorithm for Linear Codes," *Algebraic Algorithms and Error-Correcting Codes*, Proceedings of AAECC-3, LNCS 229, Springer-Verlag, pp. 150-155, 1985.

[2] Bossert, M., and G. Hergert, "Hard- and Soft-Decision Decoding Beyond the Half Minimum Distance — An Algorithm for Linear Codes," *IEEE Transactions on Information Theory* IT-32(5), pp. 709-714, September 1986.

Multisequence Generation and Decoding of Cyclic Codes over $\mathbb{Z}_q$

José Carmelo Interlando and Reginaldo Palazzo Jr.[1]

Dept. of Communications, State University of Campinas - UNICAMP - P.O. Box: 6101, Campinas, SP 13083-970 - Brazil
email: carmelo@decom.fee.unicamp.br, palazzo@decom.fee.unicamp.br

Abstract — **We propose an algorithm for linear feedback shift-register (LFSR) synthesis in the case of multiple sequences belonging to a commutative ring with identity. It is also shown how this algorithm can be applied to the decoding of cyclic codes defined over an integer residue ring $\mathbb{Z}_q$, where q is a power of a prime.**

I. Introduction

It is well known the practical and theoretical importance of cyclic codes defined over a finite field F_q. Recently, cyclic codes over integer rings $\mathbb{Z}_M$ have also been receiving special attention; reasons for that are, for instance, i) the mapping of cyclic codes over $\mathbb{Z}_4$ into nonlinear binary codes with excellent error-correcting capabilities, and ii) the matching of these codes to MPSK modulation schemes. In this paper we extend a set of results of the paper by Feng and Tzeng [1], culminating in a decoding procedure for cyclic codes over integer rings $\mathbb{Z}_q$, with q a power of a prime. We shall be considering only those cyclic codes over $\mathbb{Z}_q$ whose generator polynomials divide $x^n - 1$, where n denotes the code length. Let $\beta \in GR(q,r)$ (the r-dimensional Galois extension ring of $\mathbb{Z}_q$) denote a primitive root of $x^n - 1$. Suppose further that $\beta^{b+ic_1+hc_2}$ are roots of the generator polynomial $g(x)$ of a cyclic code C over $\mathbb{Z}_q$, for $i = 1, 2, ..., d_0 - 1$, $h = 1, 2, ..., s+1$, where $\gcd(c_1, n) = \gcd(c_2, n) = 1$. Then, $d_{\min}(C) \geq d_0 + s$ (Hartmann-Tzeng (HT) bound for cyclic codes over $\mathbb{Z}_q$).

II. Modified Fundamental Iterative Algorithm

Let R be a commutative ring with identity (CRI). Given an $M \times N$ matrix A with entries in R, and with rank less than N, find the smallest ℓ such that the $(\ell+1)$-th column in A can be expressed as a linear combination of the previous ℓ columns. The solution to this problem, when the entries of A lie in a field F is given by the *Fundamental Iterative Algorithm* (FIA), as proposed in [1]. Henceforth, we follow the notation of [1]. By extending Lemma 1 [1] to the ring case, we have devised a *Modified FIA*, which is similar to the original FIA, except for step 4), namely,

4) if $d_{r,s} \neq 0$, then,

- a) if there exists a $d_{r,u} \in D$, for some $1 \leq u < s$ and a y (over R) satisfying $d_{r,s} - y \cdot d_{r,u} = 0$, then $C^{(r-1,s)}(x) \leftarrow C^{(r-1,s)}(x) - y \cdot C^{(u)}(x) \cdot x^{s-u}$, and return to 3a);
- b) if either there is no such a $d_{r,u} \in D$, for some $1 \leq u < s$ or if $d_{r,s} - y \cdot d_{r,u} = 0$ does not have a solution in y (over R), then: i) if column s is LI on the previous $s-1$ columns (up to row r), then $d_{r,s}$ is stored in Table D, $C^{(s)}(x) \leftarrow C^{(r-1,s)}(x)$, $C^{(0,s+1)}(x) \leftarrow C^{(s)}(x)$, $s \leftarrow s+1$, $r \leftarrow 1$, and return to 2) else, ii) if $a_{h,s} + \alpha_1 a_{h,s-1} + \ldots + \alpha_{s-1} a_{h,1} = 0$, $1 \leq h \leq r$, for some choice of coefficients α_i, then $C^{(r-1,s)} \leftarrow 1 + \alpha_1 x + \ldots + \alpha_{s-1} x^{s-1}$, and return to 2).

[1]This work has been supported by CNPq, under grant 301416/85-0, and FAPESP, under grant 92/4845-7, Brazil.

Theorem 1 *The final s and $C^{(r-1,s)}(x)$ obtained from the* Modified FIA *is the solution to the problem with minimum s.*

III. Extended Berlekamp-Massey Algorithm

Given t sequences over a CRI R, find a shortest LFSR that is capable of generating them, i.e., solve the linear system of equations (1) (in [1]) over R. This is equivalent to finding the minimum ℓ such that the $(\ell+1)$-th column in matrix S, as in [1], can be expressed as a linear combination of the previous ℓ columns. Here, our main result was to extend Theorems 2 and 3, from [1], to the case when the sequences lie in a CRI, and incorporate it in the *Generalized Berlekamp-Massey (BM) Algorithm for Multiple Sequences.* The obtained algorithm is similar to the original one, except for Steps 2) and 3). We describe Step 2 a) and b) below, which refer to the computation of $\sigma^{(n+1,1)}(x)$ from $\sigma^{(n,t)}(x)$ when $d_n^{(1)} \neq 0$; Step 3) works in a similar way. For more details, see [2].

2b) if $d_n^{(1)} \neq 0$ then find an m_t such that the equation $d_n^{(1)} - y \cdot d_{m_t}^{(1)} = 0$ has a solution in y (over R). Then, $\sigma^{(n+1,1)}(x) = \sigma^{(n,t)}(x) - y \cdot \sigma^{(m_t+1,t)}(x) \cdot x^{n-m_t}$ and $l_{n+1}^{(1)} = \max\{l_n^{(t)}, n - m_t + l_{m_t}^{(t)}\}$;

2c) if $l_{n+1}^{(1)} = \max\{l_n^{(1)}, n+1-l_n^{(1)}\}$ then go to 3); else search for a solution $D^{(n+1,t)}(x)$ with minimum possible degree l in the range $\max\{l_n^{(1)}, n+1-l_n^{(1)}\} \leq l < \max\{l_n^{(1)}, l_{m_t}^{(t)} + n - m_t\}$ such that the polynomial defined by $D^{(n+1,t)}(x) - \sigma^{(n,t)}(x) = x^{n-m_t} \cdot \sigma^{(m_t,t)}(x)$ is a solution for the first m_t power sums, $d_{m_t}^{(1)} = -d_n^{(1)}$, and $\sigma_0^{(m_t)}$ is a zero divisor in R. If such a polynomial is found, then $\sigma^{(n+1,t)}(x) \leftarrow D^{(n+1,t)}(x)$; and $l_{n+1}^{(1)} \leftarrow l$;

IV. Decoding of Cyclic Codes over $\mathbb{Z}_q$

The error-location numbers are calculated by solving the linear system of equations (20) (in [1]) with minimum possible ν, via the *BM Algorithm for multisequences over a CRI.* In general, one has more than one minimal solution satisfying equations (20) (in [1]). However, we have shown that when $\rho(z) = z^\nu \sigma(z^{-1})$ has ν or more roots z_i (note that $\rho(z)$ is a polynomial with coefficients in $GR(q,r)$), then these roots are related to the *correct* error location numbers by $z_i - x_i^{c_1} =$ zero divisor in $GR(q,r)$, for $1 \leq i \leq \nu$, and are uniquely determined. The error magnitudes are still computed using Forney's procedure with minor changes.

References

[1] 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.

[2] J.C. Interlando, *A contribution to the construction and decoding of linear codes over abelian groups via concatenation of codes over integer residue rings,* Ph.D. Dissertation, State University of Campinas - UNICAMP, Brazil, Dec. 1994.

On the maximality of BCH codes

Françoise Levy-dit-Vehel[1] and Simon Litsyn[2]

[1] Institut National de Recherche en Informatique et Automatique (INRIA), Domaine de Voluceau, Rocquencourt, B.P.105, 78153 Le Chesnay Cedex, FRANCE

[2] Tel Aviv University, Dept. of Electrical Engineering-Systems, Ramat Aviv 69978, ISRAEL

Abstract — **We present new bounds for the minimal length starting from which BCH codes of given minimal distance $2t+1$ have covering radius at most $2t$.**

I. Introduction

A linear $[n,k,d]$ code C is said to be maximal if, for all linear $[n,k+1]$ codes C',

$$C' \supset C \Rightarrow d(C') < d,$$

where $d(C')$ is the minimum distance of C'. In other words, one cannot add a coset to C without decreasing its minimum distance.

The problem of determining the length starting from which the t-error correcting BCH code is maximal amounts to the one of finding the smallest length for which its covering radius is strictly less than $2t+1$.

The first result of this kind was derived by Tietäväinen [2], following a paper of Helleseth [1]. His bound guarantees maximality for the t-error correcting BCH code of length $n = \frac{2^m-1}{N}$, provided $2^m \geq ((2t-1)N)^{4t+2}$. Here we sharpen this bound.

For asymptotic results on covering radius of BCH codes, see also the paper of Skorobogatov and Vlăduts [3].

II. The results

Theorem 1 *The t-error correcting BCH-code of length 2^m-1 over $\mathbf{F}_2$ is maximal provided*

$$2^m \geq 4(1+\varepsilon(t))(t-1)^2(t!)^2,$$

where $\varepsilon(t)$ is a decreasing function of t, $\varepsilon(4) < 0.581$, $\varepsilon(5) < 0.138$, and $\varepsilon(t) < \frac{e^{2t}}{(t-1)^{2(t-1)}}$ for $t \geq 5$.

Theorem 2 *The t-error correcting BCH-code of length $n = \frac{2^m-1}{N}$ over $\mathbf{F}_2$ is maximal provided*

$$2^m \geq (1+\varepsilon_N(t))((2t-1)N-1)^2(t!)^2,$$

where $\varepsilon_N(t)$ is a decreasing function of t satisfying, for $N \geq 2$, $\varepsilon_N(4) < 0.347$, $\varepsilon_N(5) < 0.008$, and $\varepsilon_N(t) < \frac{4e^{2t}}{((2t-1)N-1)^2(t-1)^{2(t-2)}}$ for $t \geq 5$.

III. Sketch of the proof

Let $BCH(2t+1)$ stand for the t-error correcting BCH code of length $n = (2^m-1)/N$.

Consider the system

$$\begin{cases} x_1^N + \ldots + x_i^N = b_1 y^N \\ x_1^{3N} + \ldots + x_i^{3N} = b_2 y^{3N} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ x_1^{(2t-1)N} + \ldots + x_i^{(2t-1)N} = b_t y^{(2t-1)N} \end{cases} \tag{1}$$

Let $\mathcal{N}_i$ be the number of solutions $(x_1,\ldots,x_i,y) \in (\mathbf{F}_{2^m}^*)^{i+1}$ of system (1), with $x_j \neq x_k$ for $j \neq k$. If, for all $(b_1,\ldots,b_t) \in \mathbf{F}_{2^m}^t \setminus \{\mathbf{0}\}$, there exists (at least one) i, $1 \leq i \leq 2t$, such that $\mathcal{N}_i \neq 0$, then the covering radius of $BCH(2t+1)$ is less than or equal to $2t$.

To prove the maximality of $BCH(2t+1)$ it is sufficient to prove that, starting from a suitable length, its covering radius is less than or equal to $2t$. We are done if we can prove that, for m large enough, the sum

$$\sum_{i=0}^{t} a_i \mathcal{N}_i \neq 0,$$

for some $(2t+1)$-tuple $(a_0,\ldots,a_{2t})$.

Choosing the $(2t+1)$-tuple $(a_0,\ldots,a_{2t})$ to be the coefficients of the expansion of a properly chosen polynomial of degree $2t$ in the basis of Krawtchouk polynomials, we obtain the aforementioned result.

References

[1] T. HELLESETH: On the covering radius of cyclic linear codes and arithmetic codes, *Discrete Applied Mathematics*, vol. 11, pp. 157–173, 1985.

[2] A. TIETÄVÄINEN: On the covering radius of long binary BCH codes, *Discrete Applied Mathematics*, vol. 16, pp. 75–77, 1987.

[3] S. G. VLĂDUTS and A. N. SKOROBOGATOV: Covering radius for long BCH codes, *Problemy Peredachi Informatsii*, vol. 25, pp. 38–45, 1989. Translated in: *Problems of Inform. Transm.*, vol. 25, No. 1, pp. 28–34, 1989.

Weights of Long Primitive Binary BCH–Codes Are Not Binomially Distributed

Dejan E. Lazic, Hakam Kalouti, Thomas Beth
Universität Karlsruhe, Fakultät für Informatik,
Institut für Algorithmen und Kognitive Systeme, D–76 128 Karlsruhe, Germany
e-mail: lazic@ira.uka.de, kalouti@ira.uka.de

Abstract — **Primitive binary BCH–codes were supposed to have binomial weight distributions for all code rates $R \leq 1$ when $N \to \infty$. Here it is shown that this is only true if $R \to 1$.**

I. Introduction

The first bound on the differences between weight distributions and the binomial distribution was given in [1] for primitive binary BCH–codes of length $N = 2^m - 1$ and $d_{min} = 2t+1$. There it was shown that for

$$0 < t < 0.1 \cdot \sqrt{N} \tag{1}$$

and any weight (distance) d_H satisfying the inequalities $2t + \nu(t) \leq d_H \leq N - 2t + \nu(t)$ with

$$\nu(t) = \left\lceil \frac{2t \ln t + 4.5\, t + 0.1 \ln N}{0.5 \ln N - \ln t - 2.25} \right\rceil$$

the number of codewords A_{d_H} of weight d_H is given by

$$A_{d_H} = 2^{-(N-K)} \binom{N}{d_H} (1+\varepsilon(N)), \quad |\varepsilon(N)| < \text{const} \cdot N^{-0.1} \tag{2}$$

This bound was improved by several authors and lead to the common opinion that the weights of long primitive binary BCH–codes are binomially distributed for all code rates $R \leq 1$ with $N \to \infty$. In this paper it is shown that this is only true if $R \to 1$.

II. Binary Block Codes with Binomial Distance Distributions

Using the distance distribution method in [2] it was shown that fixed rate code sequences of binary block codes with asymptotic (in N) binomial Hamming distance distribution have a cutoff rate that is equal to the channel cutoff rate of the BSC and thus are asymptocially optimal according to Massey's cutoff rate criterion. Furthermore, in [3] it was shown that the error exponent of such a code family attains the BSC error exponent, if the code rate R lies in the interval between the critical rate R_{crit} and channel capacity R_C. Thus, binary codes with binomial distance distribution for $N \to \infty$ have a positive error exponent for all rates up to the channel capacity of the BSC. This argument and (2) lead to the conclusion that primitive binary BCH–codes are asymptotically optimal on the BSC, i. e., have a positive error exponent $E(R)$ in the interval $(0, R_C)$, if they are decoded by a Maximum–Likelihood decoder. This conclusion is shown to be false by a contradiction based on results from [3] and [4].

III. The Contradiction

For binary codes with $R \to 1$ we obviously have

$$N \to \infty \text{ and } R \to 1 \implies E(R) = 0\,. \tag{3}$$

Another result from [2] yields that for a fixed rate code sequence of linear block codes with vanishing normalized minimum distance the error exponent cannot be different from zero:

$$\lim_{N\to\infty} \frac{d_{Hmin}}{N} = 0 \implies E(R) = 0\,, \quad 0 < R \leq 1\,. \tag{4}$$

Using the result from [4], one obtains for primitive binary BCH–codes (p. b. BCH) that their normalized minimum distances vanish for $N \to \infty$. Thus, we arrive at a contradiction: From the conclusion in Section II we have

$$\text{p. b. BCH} \implies E(R) > 0\,, \qquad R < R_C < 1\,, \tag{5}$$

and from (4) and Berlekamp's result in [4] follows

$$\text{p. b. BCH} \implies E(R) = 0\,. \qquad 0 < R \leq 1\,, \tag{6}$$

There is no contradiction, if we compare expression (6) with (3)

$$\text{p. b. BCH} \implies E(R) = 0\,, \qquad R \to 1\,. \tag{7}$$

In fact, the solution of the contradiction between (5) and (6) can be obtained by analyzing the code rate R of the code sequences used in [1]. Using equation (1), $m = \log_2(N+1)$, and the well known inequality $N - K \leq mt$ for BCH–codes we have

$$R \geq 1 - \frac{0.1 \cdot \log_2(N+1)}{\sqrt{N}}\,. \tag{8}$$

For long codes, i. e. $N \to \infty$, this result leads to $R \to 1$. Thus, only a comparison of expression (6) with (7) is possible. But for $R = 1$ all binary codes have binomially distributed weights. We conclude that the results obtained in [1] are only valid for code rate $R \to 1$ when $N \to \infty$. Furthermore, from expression (6) follows that the weights of these codes cannot be binomially distributed for $R < 1$ and $N \to \infty$.

References

[1] **Sidelnikov V. M.:** Weight Spectrum of Binary BCH–Codes, Problemy Peredachi Informatsii, Vol. 7, pp. 14–22, January–March, 1971.

[2] **Lazic D. E., Senk V.:** 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 Transactions on Information Theory, Vol. 38, No. 5, pp. 1548–1559, 1992.

[3] **Beth Th., Lazic D. E., Senk V.:** A family of binary codes with asymptotically good distance distribution, International Symposium on Coding Theory and Applications, Udine, Italy, November 5 - 9, 1990, Proceedings, Springer - Verlag.

[4] **Berlekamp E. R.:** Long Primitive Binary BCH–Codes Have Distance $d \sim 2n \ln R^{-1} / \log n \ldots$, IEEE Transactions on Information Theory, Vol. 18, No. 3, pp. 415–426, 1972.

On Decoding Doubly Extended Reed-Solomon Codes

Jørn M. Jensen
Mathematical Institute, Technical University of Denmark,
DK-2800 Lyngby, Denmark

Abstract — **In this talk we shall discuss the algebraic decoding of doubly extended Reed-Solomon codes.**

I. Introduction

Recently, it has been shown that some doubly extended Reed-Solomon (DRS) codes have a simple encoder [1]. One way to decode a DRS code is to use a standard decoder twice, see [2, sec. 9.3]. An extension of the Berlekamp-Massey algorithm has been published, which can decode DRS codes using only one trial, [3]. The aim of the talk is to demonstrate that any t-error correcting RS decoder - such as the PGZ-, the BM- or the Euclidean algorithm - easily can be extended to be a decoder for a DRS code. In the following we shall show this for the decoder based on the Euclidean algorithm.

II. Decoding

Let $\underset{\sim}{c} = (c_-, c_0, \ldots, c_{n-1}, c_+), n \leq q-1$, be a codeword in a $(n, k, d = n-k+1)$ DRS defined over $GF(q)$. The two extended symbols are denoted c_- and c_+. A parity check matrix for the code is

$$\underset{\sim}{H} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \alpha & & \alpha^{n-1} & 0 \\ \vdots & & & & & \\ 0 & 1 & \alpha^{d-2} & \cdots & (\alpha^{d-2})^{n-1} & 1 \end{bmatrix},$$

where α is a primitive element in $GF(q)$. Let $\underset{\sim}{r} = (r_-, r_0, \ldots, r_{n-1}, r_+)$ be the received vector and $\underset{\sim}{e} = \underset{\sim}{r} - \underset{\sim}{c}$ the error vector. The $d-1$ syndromes $S_0, S_1, \ldots, S_{d-2}$ are calculated as $\underset{\sim}{H}\underset{\sim}{r}^T = (S_0, S_1, \ldots, S_{d-2})^T$. Let $S(X) = S_0 + S_1 x + \cdots + S_{d-2}x^{d-2}$ be the syndrome polynomial. Assume that $w(\underset{\sim}{e}) = s \leq \lfloor (d-1)/2 \rfloor$. An error-locator polynomial $\lambda(x) = \lambda_0 + \lambda_1 x + \cdots + \lambda_s x^s$ is defined as

$$\begin{aligned} \lambda(\alpha^{-i}) &= 0 \text{ if } e_i \neq 0, 0 \leq i \leq n-1 \\ \lambda_0 &= 0 \text{ if } e_+ \neq 0 \\ \lambda_s &= 0 \text{ if } e_- \neq 0 \end{aligned} \quad (1)$$

An error-evaluator polynomial $w(x)$ is defined as

$$w(x) = e_- \cdot \lambda(x) + \sum_{i \in I} e_i \frac{\lambda(x)}{1 - \alpha^i x}, \quad (2)$$

where $I = \{i | e_i \neq 0, 0 \leq i \leq n-1\}$. It can be verified that $S(x)$, $\lambda(x)$, and $w(x)$ satisfy a key-equation, that is

$$\lambda(x)S(x) \equiv w(x) \bmod x^{d-1} \quad (3)$$

and that

$$\begin{aligned} &\deg \lambda(x) \leq t, \deg w(x) \leq t-1, \deg w(x) \leq \deg \lambda(x), \\ &\deg \lambda(x) + \deg w(x) < d-1. \end{aligned} \quad (4)$$

Also $\lambda(x)$ is a polynomial of lowest degree, satisfying (3) and (4).

A set of polynomials $(\lambda(x), w(x))$ satisfying (3) and (4) can therefore in the usual way be determined by the Euclidian algorithm. Based on the polynomials $\lambda(x)$ and $w(x)$ the non-zero elements of the error vector $(e_0, \ldots, e_{n-1})$ can be estimated, since

$$e_i = -\frac{w(\alpha^{-i})}{\alpha^{-i}\lambda'(\alpha^{-i})} \text{ if } \lambda(\alpha^{-i}) = 0 \text{ and } 0 \leq i \leq n-1 \quad (5)$$

which is the usual formula for calculating the error symbols.

The errors e_- and e_+ can now be determined by using the syndrome equations, that is

$$\begin{aligned} e_- &= S_0 - \sum_{i=0}^{n-1} e_i \\ e_+ &= S_{d-2} - \sum_{i=0}^{n-1} e_i (\alpha^{d-2})^i \end{aligned} \quad (6)$$

Hence, once the syndromes have been calculated a standard RS decoder based on the Euclidian algorithm needs only to be extended by (6) in order to be a decoder for a DRS code.

Example. Consider a $(17, 9, 9)$ DRS defined over $GF(16)$. Let α be a primitive element satisfying $\alpha^4 = 1 + \alpha$. And let the syndromes be $(S_0, \ldots, S_7) = (\alpha^{11}, \alpha^{14}, \alpha^{12}, \alpha^{12}, \alpha^5, \alpha^3, \alpha^3, \alpha^5)$. Applying Euclid's algorithm to $S(x)$ and x^8, $\lambda(x)$ and $w(x)$ are estimated to $\lambda(x) = \alpha^{10}x^3 + \alpha^{13}x^2 + \alpha x$ and $w(x) = \alpha^{11}x^3 + \alpha^7 x^2 + \alpha^{12}x$. The non-zero roots of $\lambda(x)$ are α^{-2} and α^{-7} which from (5) implies that $e_2 = \alpha^5$ and $e_7 = \alpha^9$. Using (6) $e_- = \alpha$ and $e_+ = \alpha^3$.

In this manuscript we have only considered random error correction. The conclusion given here can be extended to include erasure decoding as well.

References

[1] J. M. Jensen, "A Class of Constacyclic Codes", *IEEE Trans. Inform. Theory*, vol. 40, pp. 951–954, May 1994.

[2] R. E. Blahut, "Theory and Practice of Error Control Codes", Addison-Wesley 1983.

[3] A. Dür, "The decoding of extended Reed-Solomon codes", *Discrete Mathematics* 90(1991), pp. 21–40.

The Generalized Hamming Weight of Some BCH Codes and Related Codes

Tor Helleseth and Eli Winjum[1]

Department of Informatics, University of Bergen, Høyteknologisenteret, N-5020 Bergen, Norway

Abstract — **We determine the generalized Hamming weights d_r for $1 \le r \le h+2$ of a binary primitive BCH code with minimum distance $d = 2^h - 1$. This extends a result of van der Geer and van der Vlugt [2], [3] who determined d_r for $1 \le r \le 5$ for the triple error-correcting primitive BCH code. We also consider the weight hierarchy of some codes with parity-check polynomial which are the product of two primitive polynomials of the same degree. In particular we have studied some of the codes with few nonzero weights studied by Niho.**

I. Introduction

Let C be an $[n,k,d]$ binary linear code. The support $\chi(D)$ of a subcode D of C is defined as the number of coordinates which are not identically zero. The r-th generalized Hamming weight of a code C is defined as

$$d_r = \min\{|\ \chi(D)\ ||\ D \text{ is an r-dimensional subcode of C}\}.$$

The weight hierarchy of the code C is d_r, $r = 1, 2 \cdots, k$. The weight hierarchy is an important parameter for the code in particular for estimating the trellis complexity of the code.

To find the weight hierarchy of a code is in general a very hard problem. For BCH codes some partial results are known. For the double error-correcting BCH code of length $n = 2^m - 1$, it is known that $d_1 = 5$, $d_2 = 8$ and $d_3 = 10$. For the triple error-correcting BCH code van der Geer and van der Vlugt [2], [3] proved that $d_1 = 7, d_2 = 11, d_3 = 13, d_4 = 14$ and $d_4 = 15$. Our first result is a generalization of this result and has a simple and direct proof.

Theorem 1 Let C be a primitive BCH code of length $n = 2^m - 1$ and designed distance $d = 2^h - 1$. Then

$$d_r = 2^{h+1} - 2^{h+1-r} - 1 \text{ for } r = 1, 2, \cdots, h+1$$

and

$$d_{h+2} = 2^{h+1} - 1.$$

Proof. The positions of a primitive BCH code can be indexed by the nonzero elements of $GF(2^m)$. For a primitive BCH code of designed distance $d = 2^h - 1$, it is well known that codewords with ones in the locations which correspond to the nonzero vectors of a h-dimensional subspace of $GF(2^m)$ (considered as an m-dimensional vectorspace) have minimum weight.

It is well known that $d_1 = 2^h - 1$ and $d_r \ge \sum_{i=0}^{r-1} \lceil d_1/2^i \rceil = 2^{h+1} - 2^{h+1-r} - 1$ for $1 \le r \le h+1$. Hence it is sufficient to find an r-dimensional subcode with the support given in the theorem when $1 \le r \le h+2$.

Let U be a subspace of dimension $h+1$. We let $\mathbf{c}_i$ for $i = 1, 2, \cdots, h+2$ denote minimum weight codewords with nonzero locations corresponding to subspaces V_i, $i = 1, 2, \cdots, h+2$ of U. We will show how these can be chosen such that the subcode D_r generated by $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_r$ has the support d_r. To select the subspaces V_i for $i = 1, 2, \cdots, h+2$ we first select V_1 to be any h-dimensional subspace of U. Suppose $V_1, V_2, \cdots, V_{i-1}$ have been selected, then select V_i to be any h-dimensional subspace of U not contained in $V_1 \cup V_2 \cup \cdots \cup V_{i-1}$. This is always possible as long as $i \le h+2$. Then it is easy to verify that $|\ \chi(D_r)\ | = |\ V_1 \cup V_2 \cup \cdots \cup V_r\ | = d_r$, which completes the proof.

II. On the weight hierarchy of a Niho code

We have studied the weight hierarchy of some codes with parity-check polynomial which are the product of two binary primitive polynomials of the same degree m. This is also in general a hard problem since it includes the dual of the double error-correcting codes BCH codes as a special case, where only partial results are known. Let $m_i(x)$ denote the minimum polynomial of a α^i, where α denotes an element of order $2^n - 1$.

As an example of our results on the weight hierarchy of these codes, we present good upper bounds on the complete weight hierarchy of a 4-weight code of length $2^n - 1$ where $n = 2m \equiv 0 \pmod 4$ (whose weight distribution was determined by Niho [1]).

Theorem Let $h(x) = m_1(x)m_d(x)$ be the parity-check polynomial of a $[2^{2m} - 1, 4m, 2^{2m-1} - 2^m]$ code C where $m \ge 2$ is an even integer. If $d = 2^{m+1} - 1$ then $\gcd(d, 2^{2m} - 1) = 1$ and

$$d_r \le \begin{cases} (2^r - 1)(2^{2m-r} - 2^{m+1-r}), & 1 \le r \le m \\ (2^m - 1)(2^m - 2) + (2^{r-m} - 1)2^{2m-r}, & m+1 \le r \le 2m \\ (2^m - 1)(2^m - 1) + (2^{r-2m} - 1)2^{3m-r}, & 2m+1 \le r \le 3m \\ (2^m - 1)2^m + (2^{r-3m} - 1)2^{4m-r}, & 3m+1 \le r \le 4m. \end{cases}$$

We also give lower bounds and show that equality holds in many cases.

References

[1] Y. Niho, Multi-valued cross-correlation functions between two maximal linear recursive sequences, Ph.d thesis, University of Southern California, USCEE report 409, Los Angeles, USA.

[2] G. van der Geer and M. van der Vlugt, On generalized weights of BCH codes, IEEE Trans. on Inform. Theory, vol. 40, pp. 543-546, 1994.

[3] G. van der Geer and M. van der Vlugt, Generalized Hamming weights of BCH(3) Revisited, IEEE Trans. on Inform. Theory, vol. 41, pp. 300-301, 1995.

[1]This work was supported in part by the Norwegian Research Council under Grant Numbers 107542/410 and 107623/420

Cyclic Codes and Quadratic Residue Codes over Z_4

Vera Pless[1] and Zhongqiang Qian
Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago
Chicago, IL, USA

Abstract — **We prove that any $Z_4 - cyclic$ code has generators of the form $(fh, 2fg)$ where $fgh = x^n - 1$ over Z_4. From this we can easily find the order of the code and generators of the dual. A particular interesting family of $Z_4 - cyclic$ codes are quadratic residue codes. We define such codes in terms of their idempotent generators and show that these codes also have many good properties which are analogous in many respects to properties of quadratic residue codes over a field.**

I. $Z_4 - cyclic$ CODES

Let Z_4 denote the integer residues modulo 4. Z_4 is a ring which has 2 as a zero divizor. A set of $n - tuples$ over Z_4 is called a code over Z_4 or a Z_4 *code* if it is a Z_4 *module*.
Let μ: $Z_4[x] \rightarrow Z_2[x]$ be the map which sends 0, 2 to 0; 1, 3 to 1 and x to x.

Definition: A polynomial f in $Z_4[x]$ is *basic irreducible* if μf is irreducible in $Z_2[x]$; f is *primary* if (f) is a primary ideal.
Lemma: If $x^n - 1 = f_1 f_2 \cdots f_r$, where the f_i are basic irreducible and pairwise coprime, then this factorization is unique.

Theorem 1: Let all f_i be as above for an odd n, and let $\hat{f}_i$ denote the product of all f_j except f_i, then the ideals $(\hat{f}_i)$ and $(2\hat{f}_i)$, for $i = 1, 2, \cdots r$, generate all ideals of $Z_4[x]/(x^n - 1)$.

If f is a polynomial, f^* denotes its reciprocal.

Theorem 2: Suppose C is a $Z_4 - cyclic$ code of odd length n, and $x^n - 1 = f_1 f_2 \cdots f_r$, where the f_i are basic irreducible and pairwise coprime, then $C = (fh, 2fg)$, where g and h are coprime and $fgh = x^n - 1$, $|C| = 4^{n-degf-degh} 2^{n-degf-degg}$, and $C^{\perp} = (g^* h^*, 2g^* f^*)$.

Theorem 3: Let C be as in theorem 2, if $C=(f)$, then C has an idempotent generator in Z_4; if $C=(2f)$, then $C=(2e)$, where e is an idempotent generator in Z_2; if $C=(fh, 2fg)$, then $C=(e, 2\nu)$ where $fgh = x^n - 1$, e is an idempotent in Z_4 and ν is an idempotent in Z_2.

Theorem 4: If $C = (e(x))$ where e is an idempotent in $Z_4[x]$, then $C^{\perp} = (1\text{-}e(x^{-1}))$.

II. QUADRATIC RESIDUE CODES

Quadratic residue codes are cyclic codes which can be defined in terms of their idempotent generators [5].
Let $e_1 = \sum_{i \in Q} x^i$ and $e_2 = \sum_{i \in N} x^i$, where Q is the set of quadratic residues and N is the set of non quadratic residues for a prime $p \equiv \pm 1 \pmod 8$. Then e_1 and e_2 are idempotents of binary $Q.R$ $[p, p + 1/2]$ codes.

Theorem 5: Let p be a prime $\equiv \pm 1 \pmod 8$ such that $p + 1$ (or $p - 1$) $= 8r$. If r is odd then $e_i + 2e_j$ and $1 + 3e_i + 2e_j$ are idempotents over Z_4, where $i, j = 1, 2$ and $i \neq j$.

If r is even then $3e_i$ and $1 + e_i$ are idempotents over Z_4, where $i = 1, 2$.

Definition: A $Z_4 - cyclic$ code is a $Z_4 -$ *quadratic residue (Q.R) code* if it is generated by one of the idempotents in theorem 5.

Theorem 6: Let p be a prime and $p + 1 = 8r$ for odd r, if $Q_1 = (e_1 + 2e_2)$, $Q_2 = (e_2 + 2e_1)$, $Q_1' = (1 + 3e_2 + 2e_1)$ and $Q_2' = (1 + 3e_1 + 2e_2)$ are $Z_4 - Q.R$ codes, then
(a) Q_1 and Q_2 are equivalent, Q_1' and Q_2' are equivalent;
(b) $Q_1 \cap Q_2 = (3h)$ and $Q_1 + Q_2 = R_p = Z_4[x]/(x^p - 1)$, where h is all 1 vector;
(c) $|Q_1| = 4^{p+1/2} = |Q_2|$;
(d) $Q_1 = Q_1' + (h)$, $Q_2 = Q_2' + (h)$;
(e) $|Q_1'| = |Q_2'| = 4^{p-1/2}$;
(f) Q_1' and Q_2' are self-orthogonal and $Q_1^{\perp} = Q_1'$, $Q_2^{\perp} = Q_2'$.
Note: If r is even or $p \equiv 1 \pmod 8$, there are similar results.

Theorem 7: Let $\overline{Q}$ be an extended $Z_4 - Q.R$ code. Then the group of $\overline{Q}$ contains a subgroup which is isomorphic to $PSL_2(p)$.

Theorem 8: The extended $Z_4 - Q.R$ code of length 32 has minimum Lee weight 14, minimum Euclidean weight 16 and minimum Hamming weight 8.

The extended $Z_4 - Q.R$ code of length 48 has minimum Lee weight 18 , minimum Euclidean weight 24 and minimum Hamming weight 12.

Their images under the Gray map are non-linear and have better minimum Hamming weight than any known binary linear codes [3].

ACKNOWLEDGEMENTS

We thank A.R.Calderbank for sending us copies of [2] and [4]. Some of results in [1], [2] and [4] are similar to ours, but ours were done differently and independently.

REFERENCES

[1] A.Bonnecaze, P.Solé, "Quaternary constructions of formally self-dual binary codes and unimodular lattices", *Lecture Notes in Computer Science*, 781(1994), pp. 194–205.

[2] A.Bonnecaze, P.Solé and A.R.Calderbank, "Quaternary quadratic residue codes and unimodular lattices", *IEEE Trans. Inform. Theory*, vol.41, No.2, pp.366-377, March 1995.

[3] A.E.Brouwer and Tom Verhoeff, "An updated table of minimum-distance bounds for binary linear codes", *IEEE Trans. Inform. Theory*, vol.39, pp. 662-675, March 1993.

[4] A.R.Calderbank and N.J.A.Sloane, "Modular and p-adic cyclic codes", preprint.

[5] J.S.Leon, J.M.Masley, and V.Pless, "Duadic codes", *IEEE Trans.Inform.Theory*, vol.30, pp. 709-714, September 1984.

[1]This work was supported in part by NSA grant MDA 904-91-H-0003.

Improved Estimates for the Minimum Distance of Weighted Degree Z_4 Trace Codes[1]

Tor Helleseth
Dep. of Informatics
Univ. of Bergen
N-5020, Bergen
Norway

P. Vijay Kumar
EEB 534, EE-Systems
Univ. South. Calif.
Los Angeles
CA 90089-2565

Oscar Moreno
Dep. of Mathematics
Univ. Puerto Rico
Rio Piedras
PR 00931

Abhijit G. Shanbhag
EEB 522, EE-Systems
Univ. South. Calif.
Los Angeles
CA 90089-2565

Abstract — **A recently derived upper bound for Weil-type exponential sums over Galois rings leads directly to an estimate for the minimum Lee distance of Z_4-linear trace codes. In this paper, an improved minimum distance estimate is presented. The improved estimate is tight for the Kerdock code as well as for the Delsarte-Goethals' codes.**

I. INTRODUCTION

Let $R_m := GR(4, m)$ denote the Galois ring (char. 4) of 4^m elements. Let β be an element in R_m of order $2^m - 1$ and set $\mathcal{T}_m = \{0, 1, \beta, \beta^2, \ldots, \beta^{2^m-2}\}$. Let $f(x) \in R_m[x]$ be *nondegenerate* with *weighted degree* D_f [1]. We define a Z_4-linear *weighted degree* trace code $C(m, D)$ via

$$C(m, D) = \{\theta + Tr(f(x)) \mid D_f \leq D, \theta \in Z_4\}_{x \in \mathcal{T}_m}.$$

The minimum Lee distance $d_{\min}$ of the codes $C(m, D)$ can be shown to be

$$d_{\min} = \min \quad \{2^m - \Re\{\textstyle\sum_{x \in \mathcal{T}_m} \omega^{\theta + Tr(f(x))}\} \mid \theta \in Z_4, \quad f(x) \text{ nondegenerate}, \quad D_f \leq D\}$$

where $\Re(x)$ denotes the real part of x.

In [1], Kumar et al. prove

Theorem 1

$$\left| \sum_{x \in \mathcal{T}_m} \omega^{Tr(f(x))} \right| \leq (D_f - 1) 2^{\frac{m}{2}}.$$

Thus, $|\Re(\sum_{x \in \mathcal{T}_m} \omega^{Tr(f(x))})| \leq (D_f - 1) 2^{\frac{m}{2}}$. In this paper, we show that this estimate can in some cases, be strengthened upto a factor of $\sqrt{2}$.

II. IMPROVED ESTIMATES

Define $D_1 = D$ or $D - 1$ when D is odd or even respectively. With the code $C(m, D)$, we associate the sets $S_1 = \{\beta^{2^i a} \mid 1 \leq a \leq \lfloor \frac{D}{2} \rfloor, 0 \leq i \leq m - 1\}$, and $S_2 = \{\beta^{2^i b} \mid 1 \leq b \leq D_1, 0 \leq i \leq m - 1\}$. Also let $S_1^2 = \{\beta_1 \cdot \beta_2 \mid \beta_1, \beta_2 \in S_1\}$. Define $S_D = S_1^2 \cup S_2$. Using McEliece's theorem on divisibility of binary cyclic codes, one can show that

Theorem 2 *Let l be the smallest integer for which the product of terms in S_D yields 1. Then 2^{l-1} divides the Lee weight of every codeword in $C(m, D)$.*

For a more general version of Theorem 2, see [3].

We denote

$$\rho_{f,m} = \Re(\sum_{x \in \mathcal{T}_m} \omega^{Tr(f(x))}).$$

Let $f(x) = a(x) + 2b(x)$, $a(x), b(x) \in \mathcal{T}_m[x]$. For any positive integer j, let $w_2(j)$ denote the Hamming weight of the binary expansion of j. For $g(x) = \sum_{j=0}^{n} g_j x^j \in \mathcal{T}_m[x]$, we define $w_2(g(x)) = \max\{w_2(j) \mid g_j \neq 0, 0 \leq j \leq n\}$. We then define $w_2(f(x)) = \max\{2 \cdot w_2(a(x)), w_2(b(x))\}$. It can be shown that l in Theorem 2 satisfies $l \geq \lceil \frac{m}{w_2(f(x))} \rceil$. Thus,

$$2^{\lceil \frac{m}{w_2(f(x))} \rceil} | 2 \cdot \rho_{f,m}. \tag{1}$$

Let $\rho_{f,ms} = \Re(\sum_{x \in \mathcal{T}_{ms}} \omega^{Tr(f(x))})$. In a similar manner, $2^{\lceil \frac{ms}{w_2(f(x))} \rceil} | 2 \cdot \rho_{f,ms}$ so that

$$2^{\lfloor \frac{m}{w_2(f(x))} \rfloor s} | 2 \cdot \rho_{f,ms}. \tag{2}$$

Using a result of Ax and Moreno and Moreno's adaptation [2] of Serre's technique, we have

Theorem 3 *Let $h_f = \lfloor \frac{m}{w_2(f(x))} \rfloor$, $e_f = \lceil \frac{m}{w_2(f(x))} \rceil$. Then*

$$|\rho_{f,m}| \leq 2^{e_f - 1} \lfloor \frac{2^{h_f - 1}(D_f - 1)\lfloor 2^{1-h_f}\sqrt{q} \rfloor}{2^{e_f - 1}} \rfloor.$$

Further,

Corollary 4 *Let d_{min} be the minimum Lee distance of the code $C(m, D)$. Let $e = min\{e_f | D_f \leq D\}$ and $h = min\{h_f | D_f \leq D\}$. Then*

$$d_{min} \geq 2^m - 2^{e-1} \lfloor \frac{2^{h-1}(D - 1)\lfloor 2^{1-h}\sqrt{q} \rfloor}{2^{e-1}} \rfloor.$$

The bound in Theorem 3 and Corollary 4 are infact sharp when applied to Kerdock and Delsarte-Goethals' codes.

[1] The work was supported in part by the Norwegian Research Council for Science and the Humanities and in part by the National Science Foundation under Grant Number NCR-93-05017.

REFERENCES

[1] P.V. Kumar, T. Helleseth, and A.R. Calderbank, "An Upper Bound for Weil Exponential Sums over Galois Rings and Applications", *IEEE Trans. Inform. Theory*, vol. IT-41, pp. 456–468, 1995.

[2] O. Moreno and C. Moreno, "The Macwilliams-Sloane Conjecture on the Tightness of the Carlitz-Uchiyama Bound and the Weights of Duals of BCH Codes", *IEEE Trans. Info. Theory*, vol. IT-40, pp. 1101–1113, 1994.

[3] B. Poonen et al., "Notes on a 2-adic Proof of McEliece's Theorem and Generalizations", preprint provided by Robert Calderbank.

Representation of Nonlinear OQPSK-Type Modulated Waveforms as a Sum of Linear OQPSK-Type Components

A.Gusmão [1] and V.Gonçalves [2]

[1] CAPS, I.S.Técnico, Lisboa, Portugal; [2] ENIDH, Paço D'Arcos, Portugal

Abstract — **We point out that a large variety of nonlinear OQPSK-Type waveforms can be exactly represented as a sum of linear OQPSK-type components. A similar representation, with an increased number of components, can be adopted for any signal obtained by filtering and nonlinear amplification of the above mentioned waveforms.**

I. INTRODUCTION

OQPSK-type modulation schemes are known to be well suited for radio applications where nonlinear power amplifiers operating close to saturation are employed. In most cases, a linear modulation scheme is assumed and, hence, the complex envelope of the modulated waveform, at the amplifier input, can be given by $s_{in}(t) = \sum_n c_n x(t-nT)$, where $x(t)$ denotes the modulation pulse, T is the duration of the bit interval and, according to the data sequence, $c_{2i} = \pm 1$ and $c_{2i+1} = \pm j$. Whenever the envelope $\mid s_{in}(t) \mid$ has fluctuations, the nonlinear characteristics of the amplifier lead to some distortion. On the other hand, if a nonlinear, constant - envelope modulation scheme is adopted, no signal distortion appears at the amplifier output. It is well-known that a Binary-CPM scheme with $h = 1/2$ can also be regarded as a member of a wide OQPSK-type class, even for the "partial - response" case (e. g., GMSK, TFM, etc.), since the corresponding modulated waveforms are similar to those resulting for the linear OQPSK-type schemes. In [1] Laurent has shown that any Binary-CPM signal can be represented as a sum of several linearly modulated signals, each of them characterized by a specific modulation pulse, $x^{(k)}(t)$. The required number of linear components depends on the duration of the "frequency pulse", $g(t)$, in the standard CPM representation. If this duration is LT (for integer L), the complex envelope can be written as

$$s(t) = \sum_{k=0}^{2^{L-1}-1} \sum_n c_n^{(k)} x^{(k)}(t-nT) \qquad (1)$$

with pulses $x^{(k)}(t)$ and sequences $\{c_n^{(k)}\}$ depending on $g(t)$, h and the data sequence. For $h = 1/2$, all the linear components of the CPM signal belong to the OQPSK-type class: this means that, if $c_m^{(k)} = \pm 1$, then $c_{m+1}^{(k)} = \pm j$.

II. GENERALIZED REPRESENTATION OF NONLINEAR OQPSK-TYPE WAVEFORMS

Similarly to the Binary-CPM signals with $h = 1/2$, many other OQPSK-type signals can be exactly represented as a sum of linear OQPSK-type components. This is the case with the signal at the output of a nonlinear power amplifier, for an input belonging to the OQPSK-type class, characterized by the above-mentioned complex envelope $s_{in}(t)$. Eqn (1) can still be valid, provided that $x(t)$ has duration (L+1) T and the power amplifier is modelled as a bandpass memoryless nonlinearity; moreover, all the linear components belong to the OQPSK-type class. In this case, $c_n^{(0)} = c_n$ and, for $L \geq 2$, we have to define sequences $\{c_n^{(k)}\}, k = 1, .., 2^{L-1}-1$, according to

$$c_n^{(k)} = c_n^{(0)} \prod_{l=1}^{L-1} \beta_{n,l}^{(k)} \qquad (2)$$

where $\beta_{n,l}^{(k)} = 1$ if $\alpha_{k,l} = 0$ and $\beta_{n,l}^{(k)} = c_{n-l-1}^{(0)} c_{n-l}^{(0)*}$ if $\alpha_{k,l} = 1$, when $(\alpha_{k,L-1}\alpha_{k,L-2}\ldots\alpha_{k,1})$ is taken as the binary representation of k. The calculation of the 2^{L-1} pulses $x^{(k)}(t)$ can easily be done by taking advantage of the correlation properties of the sequences $\{c_n^{(k)}\}$; for an i.i.d. input sequence, $E[c_n^{(i)} c_m^{(j)*}] = 1$ if $n = m$ and $i = j$, and *zero* otherwise. Hence

$$x^{(k)}(t) = E[c_0^{(k)*} s(t)], k = 0, 1, \ldots, 2^{L-1} - 1, \qquad (3)$$

where $s(t)$ can be obtained from $s_{in}(t)$ by taking into account the AM/AM and AM/PM conversion functions of the amplifier. If $x(t)$ occupies the interval $[0, (L+1)T]$, the resulting "output pulses" $x^{(k)}(t)$ will occupy the following intervals: $[0, (L+1)T]$, for $k = 0$; $[0, (L-1)T]$, for $k = 1$; $[0, (L-2)T]$, for both $k = 2$ and $k = 3$; ...; $[0, T]$, for $2^{L-2} \leq k \leq 2^{L-1} - 1$. We stress the connections between the pulses $x^{(k)}(t)$ and the low pass equivalent Volterra kernels which characterize the nonlinear transmission system [2]. Additionally, we point out the following: if any OQPSK-type signal given by (1) ($\{c_n^{(k)}\}$ sequences and pulse durations defined as previously) is filtered and then power amplified by a bandpass memoryless nonlinearity, the resulting output signal can also be represented as a sum of linear OQPSK-type components; the number of output components is 2^{L+H-1} when the filter impulse response has duration HT (integer H).

REFERENCES

[1] P. A. Laurent, "Exact and Approximate Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses", *IEEE Trans.*, COM-34, pp. 150-160, Febr. 1986.

[2] S. Benedetto et al., "Modelling and Performance Evaluation of Nonlinear Satellite Links - A Volterra Approach", *IEEE Trans. Aerosp. Electron. Syst.*, Vol. AES -15, pp. 494-507, July 1979.

Properties of Guided Scrambling Encoders and their Coded Sequences

I.J. Fair, V.K. Bhargava, Q. Wang

Department of Electrical and Computer Engineering, University of Victoria
P.O. Box 3055 M.S. 8610 Victoria BC, Canada, V8W 3P6

I. Introduction

Guided Scrambling (GS) line codes augment the source bit stream prior to self-synchronizing scrambling to ensure that the scrambling process generates an encoded bit sequence with good line code characteristics [1]. With arithmetic from the ring of polynomials over GF(2), self-synchronizing scrambling can be interpreted as division of the source bit sequence by the scrambling polynomial and transmission of the resulting quotient. When augmenting bits are inserted in fixed, periodic positions, GS codes can be interpreted as block line codes which encode source words to quotients. In particular, Block Guided Scrambling (BGS) generates a transmitted bit stream which is a concatenation of finite-length quotients chosen from sets of quotients which represent each source word. Alternatively, in Continuous Guided Scrambling (CGS), the transmitted sequence appears to be a continuous quotient due to the fact that the encoder shift registers are updated following quotient selection to contain the remainder associated with the selected quotient. The quotient selection mechanisms of both BGS and CGS encoders can be modeled as finite state machines with quotient sets as input and the selected quotient as output. In CGS encoding, the selection mechanism also outputs the remainder associated with the selected quotient.

In this paper we describe several characteristics of GS encoders and their coded sequences. We begin by defining required terms.

II. Definitions

Let the complement of a polynomial $p(x)$ be the polynomial that contains the coefficient one in every position that $p(x)$ contains the coefficient zero, and contains a zero in every position that $p(x)$ contains a one.

If a quotient set Q_i exists such that each quotient contained in this set has a complement in set Q_j, and each quotient in Q_j has a complement in Q_i, we say that quotient sets Q_i and Q_j are complementary. Note that a quotient set can be its own complement.

We also say that states in the Mealy machine model of the selection mechanism are complementary if complementary quotients are selected from these states whenever the input quotient selection sets are complementary.

Finally, we denote a remainder that is generated from a particular state with non-zero probability after an undetermined period of encoding to be a recurrent remainder for that state.

III. Properties of GS Encoders

We now state three propeties of GS encoders. Complete derivation of these properties can be found in [2].

Property E1: Every quotient selection set has a complement.

Property E2: In all selection mechanisms proposed in [1 - 3] which enforce symmetrical bounds on the running digital sum (RDS) of the encoded bit sequence or do not restrict RDS, every state has a complement. In general, this holds for all GS encoders where there is symmetry in quotient selection. If the selection mechanism can be modeled with a single state, it is its own complement. Finally, when complementary quotients are selected from complementary states, the next states are complementary.

Property E3: When all source words occur with non-zero probability regardless of the encoding interval, complementary states in the CGS encoder selection mechanism have the same number of recurrent remainders.

IV. Properties of GS Encoded Sequences

GS coded sequences exhibit the following properties whenever the source bit stream is stationary and is comprised of words of any length in which the words are independent and all words have non-zero probability of occurrence. These properties also hold in many instances when one or more of the source words do not occur.

Property S1: In CGS sequences encoded using even-weight scrambling polynomials, zeros and ones occur with equal probability in all bit positions. Consequently, the power spectral density of the coded sequence can contain discrete components only at frequencies $f = m/T$ for integers m. The discrete power spectrum has the form

$$W_D(f) = |P(f)|^2 \frac{|\eta|^2}{T^2} \sum_{m=-\infty}^{\infty} \delta\left(f - \frac{m}{T}\right), \quad (1)$$

where $P(f)$ is the Fourier transform of the pulse shape, T is the duration of each coded symbol, η is the average code symbol amplitude given by

$$\eta = \frac{V_0 + V_1}{2}, \quad (2)$$

and V_0 and V_1 are the values with which the symbols zero and one are represented.

Property S2: If the scrambling polynomial has odd weight, the power spectra of CGS coded sequences resulting from source sequences with complementary statistics are identical. Further, when source words are equiprobable, the discrete spectrum is given by Equation 1.

Property S3: When the source words are equiprobable, the power spectral density is not affected by the block or continuous nature of the code, and the discrete spectrum is given by Equation 1.

Acknowledgments

This work was supported by NSERC, the British Columbia Advanced Systems Institute, and the University of Victoria.

References

[1] I.J. Fair, W.D. Grover, W.A. Krzymien, R.I. MacDonald, "Guided Scrambling: A New Line Coding Technique for High Bit Rate Fiber Optic Transmission Systems," *IEEE Transactions on Communications*, vol. 39, no. 2, pp. 289 - 297, February 1991.

[2] I.J. Fair, "Further Characterization of Guided Scrambling Line Codes," *Ph.D. Thesis*, University of Victoria, 1994.

[3] I.J. Fair, Q. Wang, V.K. Bhargava, "Polynomials for Guided Scrambling Line Codes," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 3, pp. 499 - 509, April 1995.

Channel Codes That Exploit the Residual Correlation in CELP-Encoded Speech[1]

Fady Alajaji[†], *Nam Phamdo*[‡] and *Tom Fuja*[*]

[†] Department of Mathematics and Satistics, Queen's University, Kingston, Ontario K7L 3N6, Canada

[‡] Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350, USA

[*] Department of Electrical Engineering, University of Maryland, College Park, MD 20742, USA

Abstract — **This paper describes methods by which the residual correlation in CELP-encoded speech can be exploited by an appropriately designed channel decoder. Specfically, the LSP redundancy in FS 1016 CELP is quantified and used to effect near-MAP decoding of Reed-Solomon and convolutional codes. Coding gains of up to 3.5 dB are obtained over conventional ML algorithms.**

Summary of Results

We consider the problem of reliably transmitting speech compressed with codebook-excited linear predictive (CELP) coding over a noisy channel. The particular implementation we consider is Federal Standard 1016 4.8 kbit/s CELP.

Like all practical speech encoders, CELP does not eliminate *all* the redundancy in speech samples; what remains is the "residual redundancy". In this work, we consider methods by which channel codes can use this redundancy to enhance the performance of CELP-encoded speech over very noisy channels. Specifically, we describe ways the residual redundancy in CELP's line spectral parameters (LSP's) can be quantified and exploited. We begin by proposing two models for LSP generation; the first model incorporates only the non-uniformity of the LSP's and their correlation within a CELP frame, while the second provides for correlation between frames as well. When these models are "trained" using an actual CELP bitstream they show that as many as 12.5 of the 30 high-order LSP bits in each frame may be redundant.

We next present decoding algorithms that exploit that redundancy via both convolutional and Reed-Solomon codes. For convolutional codes, we employ three soft-decision decoding schemes, all based on the Viterbi algorithm:

- ML – the "usual" maximum likelihood algorithm;
- MAP 1 – a MAP algorithm that exploits the redundancy due to the non-uniformity of the LSP's and their correlation *within* a frame – about 10 bits/frame;
- MAP 2 – which exploits the redundancy from the non-uniform distribution of the LSP's and their correlation within *and between* frames – about 12.5 bits/frame.

For block codes, we present four soft-decision decoding (SDD) algorithms:

- SDD 1 – which approximates "traditional" maximum likelihood decoding and does not attempt to exploit any of the residual redundancy;
- SDD 2 – which exploits only the redundancy due to the ordered nature of the LSP's – about 4.4 bits/frame;
- SDD 3 – which like MAP 1 exploits the redundancy due to the non-uniform distribution of the LSP's and their correlation within a frame;
- SDD 4 – which like MAP 2 exploits both the inter- and intra-frame correlation and the redundancy due to the non-uniform distribution.

Figures 1 and 2 display the simulated performance of these decoders in terms of average spectral distortion; the channel is AWGN and the modulation is BPSK. Clearly, MAP 2 and SDD 4 provide exceptional performance at very low E_b/N_0.

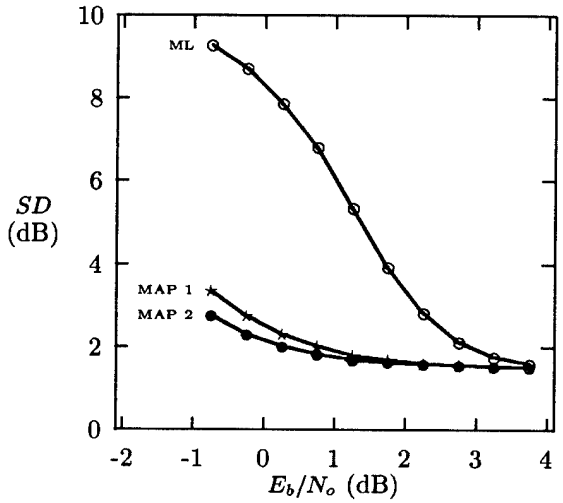


Fig. 1: Average spectral distortion – convolutional code.

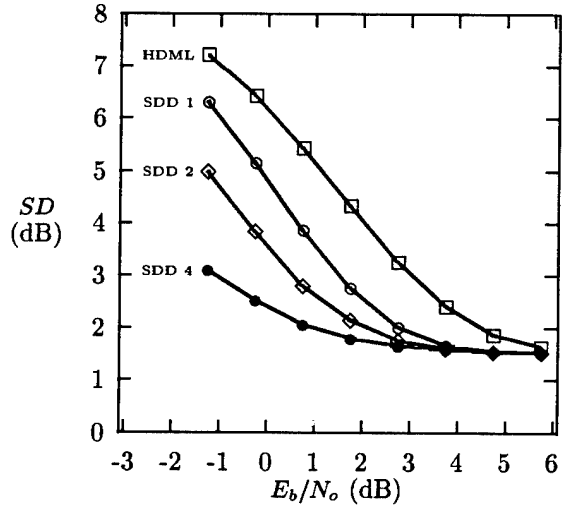


Fig. 2: Average spectral distortion – Reed-Solomon code. (HDML = hard-decision ML.)

[1]The work of Alajaji and Fuja has been supported by the U.S. Department of Defense under grant MDA 904-94-3008. The work of Phamdo has been supported by NTT Corporation.

Modal Analysis of Linear Nonbinary Block Codes Used on Stochastic Finite State Channels

Hans-Jürgen Zepernick[1]
Australian Telecommunications Research Institute,
Perth, Western Australia, Australia

Abstract – **An analytical method for an exact evaluation of the coset probabilities of algebraically decoded linear nonbinary block codes used on stochastic finite state channels is presented. The analysis can be performed in a transform domain showing an easy computational structure. The transform coefficients are connected with the coset probabilities by a complex generalization of the Walsh-Hadamard-Transformation.**

I. Introduction

A performance assessment of block codes can be achieved by rating the decoding events, where the channel has to be included. In this paper an analytical method for evaluating coset probabilities of algebraically decoded linear block codes over prime fields is presented. The codes are used on burst error channels. The ideas explained are part of a general modal approach to coding schemes [1].

II. Stochastic Finite State Channel

Input, output and state of the channel are described by a finite alphabet $\mathcal{X}$ of input symbols x, a finite alphabet $\mathcal{Y}$ of output symbols y and a finite set $\mathcal{S}$ of S states s. The conditional probability $p(y, s'|s, x)$ is the probability, when the channel is in state s, that the input x results in an output y and the next state will be s'. They are the elements $d_{ss'}(y|x) = p(y, s'|s, x)$ of a family of input-output-matrices $\mathbf{D}(y|x) = [d_{ss'}(y|x)]$. The sets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{S}$ and the matrix family $\{\mathbf{D}(y|x)\}$ define the stochastic sequential machine $\mathcal{D}$. The machine $\mathcal{D}$ together with the initial probability distribution $\boldsymbol{\sigma}_0$ on the state set $\mathcal{S}$ form the stochastic automaton $\underline{\mathcal{D}}$. Later, the symbols x, y are considered as elements of a prime field $F = GF(p)$. For the assumed symmetric channels, the matrix family is represented by the finite set $\{\, \mathbf{D}_f := \mathbf{D}(y = x+f|x) \mid f \in F \,\}$.

III. Modal analysis of nonbinary codes

We consider linear (n, k) block codes V_0 over the prime field $F = GF(p)$ with $m = n - k$ check digits. The decoder is modeled as a deterministic acceptor processing the output symbols of the channel. The syndrome $\mathbf{s} = \sum_{\nu=1}^{n} y^{(\nu)} \mathbf{h}_\nu^T$ is regarded as the decoder state, where each $\mathbf{h} = [h_{m-1}, \cdots, h_1, h_0]^T$ is a column of the parity check matrix $\mathbf{H}$. It is useful to evaluate the syndrome step-by-step leading to partial syndromes and the recursion $\mathbf{s}^{(\nu)} = \mathbf{s}^{(\nu-1)} + y^{(\nu)} \mathbf{h}_\nu^T$; $\nu = 1, \cdots, n$, where $\mathbf{s}^{(0)} = \mathbf{0}$. To each recursions step a section of the code trellis is assigned. The trellis can be analytically described by trellis matrices in the form of the Kronecker product $\mathbf{M}_\mathbf{h}(y) = \mathbf{M}_{h_{m-1}}(y) \otimes \cdots \otimes \mathbf{M}_{h_1}(y) \otimes \mathbf{M}_{h_0}(y)$ of elementary trellis matrices $\mathbf{M}_h(y) = circ(0 \cdots 10 \cdots 0)$. The one element in the first row of the circulant matrix $\mathbf{M}_h(y)$ is in the column $t = y \cdot h \ mod\ p$. The eigenvectors of $\mathbf{M}_\mathbf{h}(y)$ are the columns of the modal matrix $\mathbf{W}_m = \mathbf{W}_1 \otimes \mathbf{W}_{m-1}$, where $\mathbf{W}_1 = [w^{i \cdot j}]$ and $w = e^{-j\frac{2\pi}{p}}$. Using $\mathbf{W}_m$ for similarity transformation of the trellis matrices into the transform domain, we obtain the spectral matrices in the form of the Kronecker product $\mathbf{\Lambda}_\mathbf{h}(y) = \mathbf{W}_m^{-1} \mathbf{M}_\mathbf{h}(y) \mathbf{W}_m = \mathbf{\Lambda}_{h_{m-1}}(y) \otimes \cdots \otimes \mathbf{\Lambda}_{h_1}(y) \otimes \mathbf{\Lambda}_{h_0}(y)$ of elementary spectral matrices $\mathbf{\Lambda}_h(y) = diag\{w^{i \cdot t}\}$, where $t = y \cdot h \ mod\ p$.

IV. Modal Analysis of the coded system

The channel-decoder-cascade can be represented by a weighted trellis. An analytical description of the weighted trellis can be obtained by a weighted trellis matrix $\mathbf{U}_\mathbf{H} = \prod_{\nu=1}^{n} \mathbf{U}_{\mathbf{h}_\nu}$, where $\mathbf{U}_\mathbf{h} = \sum_{y \in GF(p)} \mathbf{M}_\mathbf{h}(y) \otimes \mathbf{D}_y$. Then, the row vector of the $T = p^m$ coset probabilities P_t is $\mathbf{P} = [P_0, P_1, \cdots, P_{T-1}] = (\boldsymbol{\tau}_0 \otimes \boldsymbol{\sigma}_0) \mathbf{U}_\mathbf{H} (\mathbf{I} \otimes \mathbf{e})$, with $\boldsymbol{\tau}_0 = [1, 0, \cdots, 0]$, $\mathbf{e} = [1, \cdots, 1]^T$ and the identity matrix $\mathbf{I}$. The mapping into the transform domain can be achieved by using $\mathbf{T} = \mathbf{W}_m \otimes \mathbf{I}$ and results in the weighted spectral matrix $\mathbf{\Theta}_\mathbf{H} = \prod_{\nu=1}^{n} \mathbf{\Theta}_{\mathbf{h}_\nu} = \prod_{\nu=1}^{n} diag\{\mathbf{\Theta}_{i\mathbf{h}_\nu}\}$, where $\mathbf{\Theta}_\mathbf{h} = \mathbf{T}^{-1} \mathbf{U}_\mathbf{h} \mathbf{T}$ and $\mathbf{\Theta}_{i\mathbf{h}} = \sum_{y \in GF(p)} \mathbf{D}_y w^{<\mathbf{i}, y \cdot \mathbf{h}^T>}$; $< \mathbf{a}, \mathbf{b} > = \sum_{\mu=0}^{m-1} a_\mu b_\mu \ mod\ p$, $\mathbf{a} = vec_p(a)$, $\mathbf{b} = vec_p(b)$. Premultiplying $\mathbf{\Theta}_\mathbf{H}$ by $(\boldsymbol{\iota}_0 \otimes \boldsymbol{\sigma}_0) = (\boldsymbol{\tau}_0 \otimes \boldsymbol{\sigma}_0)\mathbf{T}$ and postmultiplying the result by $\mathbf{I} \otimes \mathbf{e}$ yields the transform coefficients Q_i given in the vector $\mathbf{Q} = [Q_0, Q_1, \cdots, Q_{T-1}] = (\boldsymbol{\iota}_0 \otimes \boldsymbol{\sigma}_0) \mathbf{\Theta}_\mathbf{H} (\mathbf{I} \otimes \mathbf{e})$. The coefficients Q_i and probabilities P_t are connected via the complex Walsh-Hadamard-Transformation $P_t = \frac{1}{T} \sum_{i=0}^{T-1} Q_i \cdot w^{-<\mathbf{i}, \mathbf{t}>}$, where $T = p^m$.

V. Conclusion

The automata model of the channel-decoder-cascade allows an analytical evaluation of the coset probabilities of nonbinary block coded systems. By means of the modal analysis the task can be shifted into a transform domain of easier computational structure and reduced storage requirements. The domains are connected by a complex Walsh-Hadamard-Transformation. The results are exact within the framework of the model. The proposed fast algorithm is suitable for implementation on a computer.

References

[1] H.-J. Zepernick, "Modal analysis of coding schemes used on channels with and without memory,"(in German), Ph.D. dissertation, FernUniversity of Hagen, 1994.

[1] The author is on leave from the FernUniversity of Hagen

Decoding Procedure Capacities for the Gilbert-Elliott Channel [1]

Gunilla Bratt and Rolf Johannesson
Dept. of Information Theory
Lund University
P.O. Box 118
S-221 00 LUND, Sweden

Kamil Sh. Zigangirov
Dept. of Telecommunication Theory
Lund University
P.O. Box 118
S-221 00 LUND, Sweden

Abstract — **Sequential decoding for the Gilbert-Elliott channel is considered. The decoding procedure capacity C_D is defined to be the supremum of the rates for which there exists a code that gives arbitrarily small decoding error probability. For different assumptions of the decoder's knowledge of the channel states expressions for C_D are derived.**

I. Introcuction

Assume that a tree code is used together with sequential decoding to communicate over the Gilbert-Elliott channel. Let $P(\mathcal{E})$ denote the average probability of decoding error over the ensemble of random, infinite depth tree codes. In this paper we address the question: "When will $P(\mathcal{E}) \to 0$?".

Consider the Gilbert-Elliott channel model and denote the error probabilities in the Good and Bad states by e_G and e_B, respectively. Furthermore, let P_G and P_B denote the fraction of time spent in the Good and Bad states, respectively.

II. Decoding procedure capacity

Let us define the *decoding procedure assumptions*, D. The optimistic assumption, $D = o$, assumes that the decoder has a complete knowledge of the channel state, which could be given by a genie. The pessimistic assumption, $D = p$, assumes that the decoder neither is given any channel state information nor tries to make any estimate of it. Given the decoding procedure assumption D and the use of the Gilbert-Elliott channel, let C_D denote the supremum of the rates for which we can guarantee that there exists a code that gives an arbitrarily small decoding error probability $P(\mathcal{E})$. We will call C_D the *decoding procedure capacity.*

We have proved that the decoding procedure capacities are given by

$$C_o = P_G \cdot C_{BSC}(e_G) + P_B \cdot C_{BSC}(e_B)$$

and

$$\begin{aligned} C_p &= P_G \cdot (C_{BSC}(e_G) - h(b)) + P_B \cdot (C_{BSC}(e_B) - h(g)) \\ &= C_o - (P_G \cdot h(b) + P_B \cdot h(g)), \end{aligned}$$

where b and g denote the transition probabilities from Good to Bad and from Bad to Good, respectively, in the channel model.

Theorem 1 *Given the Gilbert-Elliott channel and the decoding procedure assumptions, the use of a rate R random, infinite depth tree code with the stack decoder, then for any rate $R < C_D$ and $\eta \in \mathbb{Z}^+$,*

$$P(N \geq \eta) \to 0 \text{ if } \eta \to \infty,$$

where N is the number of computations in an incorrect *subtree.*

When we wish to transmit over an ordinary Discrete Memoryless Channel at rates (above R_{comp} and) close to its capacity, it is sufficient to allow the number of computations of sequential decoding to go to infinity to be able to guarantee that $P(\mathcal{E})$ can be chosen arbitrarily small. We will show that this is also sufficient for transmission close to rates C_D, which is the motivation why we call these rates "decoding procedure *capacities*".

Theorem 2 *Given the assumptions of Theorem 1, then for any rate $R < C_D$ the average probability of decoding error*

$$P(\mathcal{E}) \to 0,$$

if the number of computations, N, is allowed to go to ∞.

Since the important condition in Theorem 2 is that $R < C_D$, it is clear that the theorem's statement, *given* the decoding procedure assumptions, is equivalent to stating that the maximal transmission rate over the Gilbert-Elliott channel is at least the rate C_D.

In the pessimistic case we can interpret this as follows. For arbitrarily small $P(\mathcal{E})$, there exists a code such that the transmission rate will be (at least) C_p, *even without* any knowledge of the channel state or any attempt to estimate it.

III. Channel capacity

A common method to lowerbound C_{GE} is to calculate $C_{BSC}(\bar{e})$, where $\bar{e} = P_G \cdot e_G + P_B \cdot e_B$, but it turns out that C_p is a better lower bound for channels with a stable behaviour. The optimistic case helps us to find a stronger result:

Theorem 3 *Given that the receiver has a complete channel state knowledge, then the channel capacity for the Gilbert-Elliott channel C_{GE}^R is equal to*

$$C_{GE}^R = C_o.$$

From the proof of Theorem 3 follows immediately

Corollary 4 *Given that* both *transmitter and receiver have complete knowledge of the channel state* sequence *then for the channel capacity of the Gilbert-Elliott channel C_{GE}^{TR} we have*

$$C_{GE}^{TR} = C_{GE}^R.$$

It should be noted that the capacities C_{GE}^{TR} and C_{GE}^R, in contradiction to what is the case for C_D, are parameters purely dependent of the channel's properties and that nothing is assumed about the decoding method. In the derivations of C_D we assume sequential decoding, but by deriving them we show that they are achievable rates as such, given the decoding procedure assumptions.

[1] This research was supported in part by the Royal Swedish Academy of Sciences in liaison with the Russian Academy of Sciences, and in part by the Swedish Research Council for Engineering Sciences under Grant 91–91.

References

[1] Gunilla Bratt: "Sequential Decoding for the Gilbert-Elliott Channel — Strategy and Analysis". Ph.D. Thesis. Dept. of Information Theory, Lund University, Lund, Sweden, 1994.

Real-Time Channel Estimation Using Fuzzy Logic

F. ARANI*, R. SMIETANA**, B. HONARY*

*Communications & Signal Processing
Research Consortium
Lancaster University
Lancaster, LA1 4YR UK.
Tel: +44 1524 594204
Fax: +44 1524 594207
Email: F.Arani@lancaster.ac.uk

**University of Ulm,
Abteilung Informationstechnik,
Albert-Einstein-Allee 43,D89081
Ulm, Germany

Abstract - **The use of real time channel estimation information is known to result in significant performance advantages in coded systems operating on fading channels. Little work has been done in fast and accurate channel signal to noise ratio estimation in a very noisy channels (SNR< 0 dB). In this paper, a new real time channel estimation technique using the Viterbi algorithm and Fuzzy logic concepts is presented.**

I- INTRODUCTION

The distortion imposed by the channel on the transmitted data stream in a digital communication system is normally observed in the form of errors at the receiver. The main objectives of any communication system are to minimise the number of these errors and to maximise the throughput of the system. In order to optimise the system performance adaptively in response to channel conditions, an estimate of the receiver's error rate is required to initiate control actions.

Real time channel estimation techniques [1] are useful tools for obtaining an on-line estimate of the channel state. Previous work [2] in this area could not accurately estimate channel SNR fast under very noisy condition.

As a by-product of the Viterbi decoding algorithm, the cumulative metric of the most likely path through the decoder trellis is available as an additional information besides the decoded output symbol. This information may be interpreted as a measure for the signal-to-noise ratio (SNR) in the transmission channel [3] and consequently the error probability of the decoded sequence could be estimated.

In the channel estimation scheme described here, the path metric values at the output of the Viterbi decoder are applied to a Fuzzy-logic unit, which retrieves this information by means of post-processing and mapping into membership functions (MF). Channel SNR estimation is made after a fixed number of decoding steps.

II- THE FUZZY-LOGIC UNIT

After a fixed number of decoding steps the Fuzzy-Logic unit (FLU) reads the transformed trellis side-information and computes the membership-values for each SNR-membership function in steps of 1dB in a range between -7dB and 27dB. The rule base consists of a small look-up table, which contains the mean values of the input information obtained during off-line training for SNRs between -10 to 30dB in steps of 1 dB. Each membership function is triangular shaped, where the highest membership-value is assigned to the Fuzzy input being equal to the stored mean value for the kth dB step, with $k \in \{-7,-6, \ldots , 27\}$. After comparing the input values with the rule-base, a vector of membership-values is obtained, which represents a fuzzy description for the SNR estimation. In order to reduce the variance of the channel SNR estimate, the membership-values are defuzzificated by the Centroid inference method [4,5].

III. SIMULATION RESULTS

Simulation results have shown that after receiving 2 kbit of decoded output symbols (i.e. eight taps), the estimated value for transmission Eb/No- between 0 dB and 25 dB can be obtained with 100% certainty with variance of 0.25 dB

IV. REFERENCES

[1] F.Zolghadr, B.Honary, M.Darnell, "Statistical real-time channel evaluation (STRCE), techniques using variable rate T-codes, IEE Proc., vol. 136, Pt. I, No.4, 1989.

[2] M.Shaw, B.Honary,B., M.Darnell, " An RTCE assisted ARQ transmission Scheme: Design and Implementation", IEE Conf. Proc. on HF Communication, vol.284, pp.43-53, 1988.

[3] J.G. Proakis,"Digital Communications", 2nd ed., McGraw-Hill 1989.

[4] H.J.Zimmerman, "Fuzzy set theory and its applications", 2nd ed., Kluwer Academic Publishers, 1992.

[5] H.H Bothe, "Fuzzy Logic", Springer Verlag, 1993.

Diversity Waveform Sets for High-Resolution Delay-Doppler Imaging

Jiann-Ching Guey and Mark R. Bell
School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-1285

Abstract — **Fundamental properties of the ambiguity function and the uncertainty relation of Fourier transforms assert a fundamental limitation on the ability of any single radar waveform to simultaneously resolve targets closely spaced in both time-delay and Doppler-shift. In this paper, a method of using multiple waveform sets to make high-resolution delay-Doppler measurements is proposed. The fundamental theorem that supports this method is established. Explicit optimal phase, frequency, and joint phase-frequency coded waveform sets having constant amplitude are presented, as well as algorithms for the construction of such sets of arbitrary size.**

I. Introduction

A radar or other pulse-echo delay-Doppler measurement system can be viewed as an imaging system that forms a delay-Doppler image of the illuminated environment. When a radar system is viewed in this way, it becomes clear that its delay-Doppler resolution is determined by its imaging point-spread function or ambiguity function of the illuminating signal $s(t)$, defined as

$$\chi_s(\tau,\nu) = \int_{-\infty}^{\infty} s(t)s^*(t-\tau)e^{-j2\pi\nu t}\,dt.$$

We have some control in selecting this point-spread function. However, some fairly strong constraints on the mathematical form of the ambiguity function prohibit the ability to simultaneously achieve high-resolution in both delay and Doppler. For example, the total volume under the squared modulus of the ambiguity function of a signal with energy E is always E^2, while the peak of the ambiguity function always has height E. This is true for any single waveform $s(t)$ and cannot be changed by any modulation scheme.

One way around this delay-Doppler resolution constraint is to make multiple pulse-echo measurements using waveforms having sufficiently different ambiguity functions and then process and combine the individual waveform returns in to form a high-resolution delay-Doppler image. This leads to the introduction of the *composite ambiguity function* of a set of signals. A main theorem on the composite ambiguity function is established to support the validity of our idea.

II. Main Theorem on Composite Ambiguity Function

Theorem 1 *For a set of signals* $\{s_0(t), s_1(t), \cdots, s_{K-1}(t)\}$ *with total energy*

$$E_T = \sum_{i=0}^{K-1} \int_{-\infty}^{\infty} |s_i(t)|^2 dt\,,$$

the volume V *under their associated composite ambiguity function* $C(\tau,\nu)$ *defined as*

$$V = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\sum_{i=0}^{K-1} \chi_{s_i}(\tau,\nu)\right|^2 d\tau d\nu$$

satisfies

$$\frac{E_T^2}{K} \le V = \sum_{i=0}^{K-1}\sum_{j=0}^{K-1}\left|\int_{-\infty}^{\infty} s_i(t)s_j^*(t)\,dt\right|^2 \le E_T^2\,.$$

Furthermore, the minimum is achieved when $\{s_0(t), s_1(t), \cdots, s_{K-1}(t)\}$ *is a set of equal-energy orthogonal signals.*

This theorem provides a general rule of selecting signal set for waveform-diverse measurements. However, a point-spread function with a small volume is not sufficient for obtaining high-resolution image. It has been shown [1, Ch. 3] quantitatively that in addition to small ambiguous volume, an ideal point-spread function for delay-Doppler radar imaging should have a thumbtack shape. This is achieved by appropriately selecting modulation schemes for the coded waveform set.

III. Coded Waveforms Design

We will study only coded waveforms because the structural constraints of these waveforms result in designs that can be easily implemented in real systems. Particular families of waveforms that are investigated include

1. Phase-modulated signals;
2. Frequency-modulated signals;
3. Frequency and phase modulated signals.

The coded waveform sets we've investigated contain K signals $\{s_0(t), s_1(t), \cdots, s_{K-1}(t)\}$ where

$$s_i(t) = \sum_{n=0}^{N-1} \psi_{i,n}(t-nT)e^{j2\pi\frac{d_{i,n}}{T}t}e^{j\phi_{i,n}} \qquad (1)$$

consists of a sequence of N baseband pulses of length T with finite energy. Each pulse is modulated by an integral frequency modulating index $d_{i,n}$ and a phase modulating index $\phi_{i,n}$ that can take on any real number.

The modulating patterns of a set of coded waveforms determine the distribution of the ambiguity sidelobes of its resulting composite ambiguity function. Phase modulating pattern controls the polarities of the ambiguity sidelobes while frequency modulating pattern determines their locations. The examples that will be shown demonstrate that by appropriately selecting the phase modulating pattern, it is possible to cancel the ambiguity sidelobes, and by selecting the frequency modulating pattern, we can spread out the ambiguity sidelobes so that the resulting composite ambiguity function resembles a thumbtack. The combination of both phase and frequency modulations gives the best result.

References

[1] J. C. Guey, *Sequence and Waveform Set Design for Radar and Communication Systems,* Ph.D. Dissertation, Purdue University, West Lafayette, IN, 1995.

Reduced Complexity Symbol-by-Symbol Demodulation

Michael P. Fitz[1] and Saul B. Gelfand
School of Electrical & Computer Engineering, Purdue University,
West Lafayette, IN, USA 47907-1285

Abstract — **Reduced complexity symbol-by-symbol demodulation is examined. We examine the performance with standard complexity reduction techniques (e.g., M-algorithm and T-algorithm) and then derive a reduced state symbol-by-symbol demodulation algorithm which makes symbol-by-symbol demodulation performance and complexity competitive with sequence estimation.**

I. Introduction

Symbol-by-symbol demodulation (SYD) structures, (e.g.,[1]) while optimum in terms minimizing symbol error probability, typically have a complexity greater than sequence demodulation (SED) techniques (e.g.,the Viterbi algorithm) for a fixed decoding lag. Consequently when only hard decision outputs are required SED techniques are invariably used in practice. However, soft decision metrics are often needed (e.g., interleaved or concatenated coding schemes), and hence reduced complexity high performance SYD structures are of interest. In this paper we propose a new algorithm that produces symbol-by-symbol metrics at roughly the same complexity as SED without a significant loss in performance and examine methods to significantly reduce the complexity of SYD.

II. Overview of Optimum Recursive Estimation

Consider a modulation with memory described by a time invariant Markov chain transmitting m bits of information per symbol corrupted by an AWGN. Define K as the decoding lag, σ_k to be the modulation state, $\|\sigma_k\|$ to be the cardinality of the modulation state, and $\mathbf{w}(k)$ to be all the observations until time k. We also use Ω_I as the transmitted symbol space and

$$\mathbf{I}_n(k) = \{\mathbf{I}_{k-n}, \mathbf{I}_{k-n+1}, \cdots, \mathbf{I}_k\}$$

to represent the last $n+1$ transmitted symbols.

Assuming equally likely transmitted symbols, recursive symbol-by-symbol and sequence estimation algorithms have the same three part structure: 1) measurement update, 2) metric production, and 3) sufficient statistic update. The measurement update takes the sufficient statistics from the previous time iteration and the latest measurement and computes an updated information state. From this information state the output metric and the sufficient statistic for the current iteration can be produced. The forward recursion optimum SYD has sufficient statistics $p(\sigma_k | \mathbf{w}(k-1))$ (the posterior probability mass function (pmf) of the modulation state) and $p(\mathbf{I}_{k-i} | \sigma_k, \mathbf{w}(k-1)), i = 1, K$ (the conditional posterior pmfs of the transmitted symbols). Similarly the sufficient statistics for optimum SED are $\max_{\mathbf{I}_{K-1}(k-1) \in \Omega_I^K} p(\mathbf{I}_{K-1}(k-1), \sigma_k | \mathbf{w}(k-1))$ (the largest posterior pmf for each modulation state) and $\arg \max_{\mathbf{I}_{K-1}(k-1) \in \Omega_I^K} p(\mathbf{I}_{K-1}(k-1), \sigma_k | \mathbf{w}(k-1))$ (the sequence that achieves the maximum). It should be noted that SED can operate on log-likelihood functions while SYD cannot, but the exponential function evaluation needed in the measurement update for SYD could easily be done with a lookup table. The complexity of optimum SYD is $O(KM^2 \|\sigma_k\|)$ where $M = 2^m$ and the complexity of SED is $O(M \|\sigma_k\|)$.

III. Complexity Reduction Techniques

Often in practice the complexity of an optimal demodulator is prohibitive and reduced complexity demodulation techniques need to be used. Since the structure of SYD is so similar to SED the best complexity reduction techniques are also similar. Two of the most applicable techniques are the M-algorithm [2] which saves the M most likely sequences and the T-algorithm [3] which saves and processes only the statistics or posterior likelihood values which break a threshold each iteration. The T-algorithm version of the SYD provides the best average complexity performance tradeoff and the threshold for this algorithm can be chosen in a principled fashion. Conversely, the T-algorithm version of the SYD has the disadvantage of having variable complexity and memory requirements.

Additionally a reduced state SYD algorithm analogous to the RSSE [4] can be derived using the approximation:

A1: $\{I_{K-1}(k-1), \sigma_k\}$ is deterministic given $\{I_{K-1}(k-1), \sigma_k\} \in \tilde{\sigma}_k$ and $\mathbf{w}(k-1)$

where $\tilde{\sigma}_k$ is the reduced state partition. The recursion resulting from (A1) has a similar form as the optimal algorithms and the sufficient statistic is $p(\tilde{\sigma}_k | \mathbf{w}(k-1))$ (the posterior pmf of the state partition) and $\hat{\mathbf{I}}_{k-i}(\tilde{\sigma}_k, \mathbf{w}(k-1)), i = 1, K$ (the conditional decisions). The complexity of this reduced state demodulator is $O(KM \|\tilde{\sigma}_k\|)$. For medium to high SNR and when $\tilde{\sigma}_k = \sigma_k$, this reduced state SYD has roughly the same complexity as SED and produces performance almost indistinguishable from the optimum estimator.

IV. Conclusions

The combination of the reduced state symbol-by-symbol demodulation and the T-algorithm provides a demodulation algorithm that maximizes average performance versus computational complexity while still maintaining a reasonable maximum complexity and memory requirement.

[1] This work was supported by NSF under Grant NCR-9406073

References

[1] Y. Li, B. Vucetic, and Y. Sato, "Optimum Soft-Output Detection for Channels in the Presence of Intersymbol Interference," *IEEE Trans. Info. Theory*, vol. IT-41, May 1972, pp. 363-378.

[2] J. B. Anderson and S. Mohan, "Sequential Coding Algorithms: A Survey and Cost Analysis," *IEEE Trans. Commun.*, vol. COM-32, February 1984, pp. 169-176.

[3] S.J. Simmons, "Breadth-First Trellis Decoding with Adaptive Effort," *IEEE Trans. Commun.*, vol. COM-38, January 1990, pp. 3-12.

[4] V.M. Eyuboglu and S.U.H. Qureshi, "Reduced-State Sequence Estimation with Set Partitioning and Decision Feedback," *IEEE Trans. Commun.*, vol. COM-36, January 1988, pp. 13-20.

The Representation of Multicomponent Chirp Signals Using Frequency-Shear Distribution

S. Yao and Z.Y. He[1]
Department of Radio Engineering, Southeast University
Nanjing 210018, P.R. China

Abstract - **We propose a new representation method for multicomponent chirp signals. This representation is based on the 2-D frequency-shear plane. Analytical results for the chirp signals are presented.**

I. INTRODUCTION

The conventional tools for analysis of this class signal are time-frequency distributions (TFD) which can be interpreted as a smoothed version of the Wigner distribution (WD) of signal to be analyzed [1]. Much current efforts have put on designing nice kernel functions to achieve better performance in suppressing cross terms while retaining high time and frequency resolutions of auto terms. However, those kernel functions are based on rectangular tessellaion of the time-frequency plane and therefore they may not suit for representing a certain types of signals such as chirping signals. In this paper, we introduce a frequency-shear distribution that maps a signal onto frequency-shear plane. The properties of this distribution are investigated and analytical results are presented.

II. THE FREQUENCY-SHEAR DISTRIBUTION

Let us define a transform of signal $x(t)$ as follows

$$Q_x(\nu,q)=\int_{-\infty}^{\infty}x(t)g(t)e^{j(\nu t+\frac{q}{2}t^2)}dt \tag{1}$$

where, ν and q denote frequency and shear, respectively. $g(t)$ is a weighted window function. The frequency-shear distribution (FSD) is defined as the squared magnitude of $Q_x(\nu,q)$:

$$\begin{aligned}\Theta_x(\nu,q)&=|Q_x(\nu,q)|^2\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W_x(t,\omega)W_g(t,\omega-\nu-qt)dtd\omega\end{aligned} \tag{2}$$

where, $W(t,\omega)$ is the so-called Wigner distribution. From the definition, we know that the time-frequency function of the signal is weighted with a chirplet function, which corresponds to the local structure of the chirp signal. The weighting function has an oblique analysis cell on the time-frequency plane that is suitable for analyzing multicomponent chirp signals.

III. THE REPRESENTATION OF MULTICOMPONENT CHIRP SIGNALS

In this section, we will consider several typical signals and analytically calculate their FSDs.

(1) Single chirp signal

A linear chirp signal with constant magnitude has the following WD $W_{x_1}(t,\omega)=A^2 2\pi\delta(\omega-\omega_0-\gamma t)$ which is highly concentrated about the chirp's linear instantaneous frequency. In the FSD, we chose a Gaussian signal as the weighted function. The FSD of this chirp signal is calculated and given by

$$\Theta_{x_1}(\nu,q)=\frac{2|A|^2\sqrt{\pi}}{|\sigma(q-\gamma)|}e^{-\frac{(\nu-\omega_0)^2}{\sigma^2(q-\gamma)^2}},\quad \text{if } \sigma>>1 \tag{3}$$

The chirp signal is located at the point (ω_0,γ) on the frequency-shear plane as expected.

(2) The signal of two chirp components

Assuming that the signal is consisting of the sum of two chirp signals

$$x_2(t)=A_1e^{j(\omega_1 t+\frac{\gamma_1}{2}t^2)}+A_2e^{j(\omega_2 t+\frac{\gamma_2}{2}t^2)}$$

we consider a particular case that $\gamma_1=\gamma_2=\gamma$ and give the WD of the signal for comparison

$$\begin{aligned}W_{x_2}(t,\omega)&=A_1^2 2\pi\delta(\omega-\omega_1-\gamma t)+A_2^2 2\pi\delta(\omega-\omega_2-\gamma t)\\&+4\pi A_1A_2\cos\left((\omega_2-\omega_1)t\right)\delta\left(\omega-\tfrac{1}{2}(\omega_1+\omega_2)-\gamma t\right)\end{aligned} \tag{4}$$

There are two auto terms centred at $\omega=\omega_1,\omega_2$ and a cross term whose peak locates on the straight line $\omega=\frac{1}{2}(\omega_1+\omega_2)-\gamma t$. The FSD of the same signal is given by

$$\begin{aligned}\Theta_{x_2}(\nu,q)&=\frac{A_1A_2\sqrt{\pi}}{|\sigma(q-\gamma)|}\left(e^{-\frac{(\nu-\omega_1)^2}{\sigma^2(q-\gamma)^2}}+e^{-\frac{(\nu-\omega_2)^2}{\sigma^2(q-\gamma)^2}}\right)\qquad \text{when } \sigma>>1\\&+\frac{2A_1A_2\sqrt{\pi}}{|\sigma(q-\gamma)|}e^{-\frac{(\nu-\omega_\Sigma)^2+\frac{1}{4}\omega_\Delta^2}{\sigma^2(q-\gamma)^2}}\cos\left(\frac{\omega_\Delta(\nu-\omega_\Sigma)}{q-\gamma}\right)\end{aligned} \tag{5}$$

where $\omega_\Sigma=\frac{1}{2}(\omega_1+\omega_2)$, $\omega_\Delta=\omega_2-\omega_1$. The cross term at $\nu=\omega_\Sigma$ reduces to $\Theta_{x_2}(\nu,q)=\frac{2A_1A_2\sqrt{\pi}}{|\sigma(q-\gamma)|}e^{-\frac{\omega_\Delta^2}{4\sigma^2(q-\gamma)^2}}$. The magnitude of the cross term is largely suppressed if compared to that of the WD.

The analytical results given in this paper have shown that the new frequency-shear distribution provides a more effective tool for analyzing multicomponent chirp signals than the generalized time-frequency distribution. It can suppress the cross terms and clearly locate the signal components onto the frequency-shear plane. In fact, this advantage is caused by introducing a chirplet function that corresponds to the structure of a chirp signal. Similarly, signal representation can be extended to scale-shear and shift-shear plane using so-called fan bases and chevron bases [2] whose elements scale or translate and shear in the time-frequency plane.

REFERENCES

[1]. L. Cohen, "Time-frequency distributions-a review", Proc. IEEE. vol. 77, pp.941-981, July 1989.

[2]. R.G. Baraniuk and D.L. Jones, "Shear madness: new orthonomal bases and frames using chirp functions", IEEE Trans. on Signal Processing, Vol. 41, No. 12, pp.3543-3549, 1993.

[1] This work was supported by the State Education Commission and the Climbing Programme-National Key Project for Fundamental Research in China, Grant NSC 92097.

A Paradigm for Distributed Detection Under Communication Constraints

Chao-Tang Yu and Pramod K. Varshney
Department of Electrical and Computer Engineering, Syracuse University, Syracuse, NY 13244

Abstract — **We present a new paradigm for the decentralized detection problem under communication constraints. In this problem, local sensors send a hard decision or the likelihood ratio itself to the fusion center based on the specified communication constraint. Optimal system is designed by minimizing the risk function. Also, a simpler system design procedure based on Ali-Silvey distances is presented.**

In this paper, we present a novel paradigm for the decentralized detection problem under communication constraints. The proposed approach is flexible and combines the features of both centralized and hard decision decentralized detection problems. Under specified constraints, we design the optimum decentralized detection scheme. The system can operate at the two extremes, i.e., it can be a centralized system or a hard decision decentralized detection system, or anywhere in-between. In this scheme, local sensors send a hard decision to the fusion center when the local sensors have a relatively high confidence in the decision, otherwise a perfect version of the LLR (in practice, a finely quantized version of the LLR) is sent. The degree of confidence at which this switch is made is determined by the specified communication constraint. The fusion center makes a final decision based on the received information from local sensors.

Observation samples at the local sensors are denoted by $\mathbf{r_i}$, $i = 1, \cdots, M$, and their joint conditional densities are assumed known. Based on its own observation $\mathbf{r_i}$, each local sensor makes a local decision $u_i \in \{0,\ 1,\ 2\}$, $i = 1, \cdots, M$, where $u_i = 0$ and $u_i = 1$ represent the fact that the i^{th} local sensor decides hypotheses H_0 and H_1 and correspondingly sends a zero and a one to the fusion center, $u_i = 2$ indicates that the i^{th} local sensor computes and sends its LLR L_i to the fusion center. Let u_{F_i} represent the output of the sensor i, i.e., $u_{F_i} = u_i$ when u_i=0 or 1; $u_{F_i} = L_i$ when u_i=2. Local sensor outputs are transmitted to the fusion center where a global decision is made based on the received data vector, $\mathbf{u_F}^T$=$[u_{F_1}\ u_{F_2}\ \cdots\ u_{F_M}]$.

The probability that $L_i(\mathbf{r_i})$ is transmitted from the sensor i is employed as a measure of the transmission rate on the channel i. We define

$$R_i = p(\text{send } L_i) = 1 - p(\text{send } H_0 \text{ or } H_1). \tag{1}$$

Note that R_i=1, $i = 1, \cdots, M$, represents the centralized case, and R_i=0, $i = 1, \cdots M$, represents the case that hard decisions are made at the local sensors[1, 2]. We are interested in examining the flexible hybrid decision scheme in the decentralized detection system with a lower average communication rate (as compared to the centralized detection problem) on the channels linking local sensors to the fusion center.

Design of a decentralized detection system involves specifying both the local decision rules and the global decision rule. By employing the person-by-person optimization methodology, the system is designed so as to minimize the risk function. The system is specified by

- Optimal local decision rule at sensor k, $k = 1, \cdots, m$:

$$u_k = \begin{cases} 0, & \frac{p(\mathbf{r_k}|H_1)}{p(\mathbf{r_k}|H_0)} < t_{20}^{(k)}, \\ 1, & \frac{p(\mathbf{r_k}|H_1)}{p(\mathbf{r_k}|H_0)} > t_{12}^{(k)}, \\ 2, & \text{otherwise.} \end{cases} \tag{2}$$

- Optimal fusion rule:

$$\frac{p(\mathbf{u_F^*}|H_1)}{p(\mathbf{u_F^*}|H_0)} \begin{matrix} u_0 = 1 \\ > \\ < \\ u_0 = 0 \end{matrix} \frac{C_f}{C_d}, \tag{3}$$

where u_F^* is the one of the 3^M possible combinations of u_F.

Motivated by the difficulty and excessive computational requirements of the above PBPO system design, a simplified design procedure based on the class of Ali-Silvey distance measures is also presented. Following the lead of [3, 4], we obtain local decision rules that maximize the Ali-Silvey distances between the conditional densities at the input of the fusion center.

It should be noted that both system designs are obtained under communication constraints given in Equation (1). An example is considered for this flexible hybrid decision scheme for the decentralized detection problem. Results show that the system performance of the proposed scheme with lower average communication rate is fairly close to the performance of the centralized system.

[0]Research sponsored by Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. F49620-94-1-0182.

References

[1] R.R. Tenney and N.R. Sandell,. "Detection with Distributed Sensors,". *IEEE Trans. Aerospace and Electronic Systems*, Vol. AES-17:pp. 501–510, July 1981.

[2] I. Hoballah and P.K. Varshney,. "Decentralized Bayesian Signal Detection,". *IEEE Trans. Information Theory*, Vol. IT-35:pp. 995–1000, Sep. 1989.

[3] H.V. Poor,. "Fine Quantization in Signal Detection and Estimation,". *IEEE Trans. Information Theory*, Vol. IT-34:pp. 960–972, Sep. 1988.

[4] M. Longo, T. Lookabaugh, and R.M. Gray,. "Quantization for Decentralized Hypothesis Testing Under Communication Constraints,". *IEEE Trans. Information Theory*, Vol. IT-36, No. 2:pp. 241–255, Mar. 1990.

Decentralized Quickest Change Detection

Venugopal Veeravalli

ECE Department, Rice University, 6100 S. Main Street, Houston, TX 77005, USA, *e-mail*: venu@rice.edu

Abstract — **The problem of change detection is considered in a decentralized setting. A Bayesian framework is introduced for this problem, and an optimal solution is obtained for the case when the information structure in the system is quasiclassical.**

I. Problem Formulation

The *centralized* version of the change detection problem—where all the information about the change is available at a single location—is well-understood and has been solved under a variety of criteria since the seminal work by Page [1]. However, there are situations where the information available for decision-making is decentralized, an example being link failure detection in a large communication networks. We focus on this decentralized setting.

Consider a system with N sensors $S_1, \ldots, S_N$. At time $k \in \{1, 2, \ldots\}$, sensor S_l observes a random variable $X_k^{(l)}$, and forms a message $U_k^{(l)}$ (belonging to a finite set) based on the information it has at time k. Assume that two-way communication is possible between the sensors and the fusion center. In particular, at time k the fusion center broadcasts to each sensor, all the sensor messages it received at time $k-1$. This means that at time k, each sensor has access to all its observations up to time k and all the messages of all the other sensors up to time $k-1$, and the fusion center has access to all the sensor messages up to time k. Based on the sequence of sensor messages, a decision about the abrupt change is made at the fusion center.

We take the approach of Shiryayev [2] and assume that the change time Γ is geometric distributed, i.e.,

$$\mathrm{P}(\Gamma = 0) = \nu \quad \text{and} \quad \mathrm{P}(\Gamma = i | \Gamma > 0) = \rho(1-\rho)^i$$

Further, we assume that observations at each sensor S_l are independent, have a common pdf $f_0^{(l)}$ before the disruption, and common pdf $f_1^{(l)}$ from the time of disruption. We also assume that the observations are independent from sensor to sensor.

As in [4], we restrict the local memory at sensor S_l to only past messages. The resulting information structure is said to be *quasi-classical* [3] and it makes the joint optimization problem tractable via DP arguments. At any time k, the one-step delayed information is the same for all members and is given by $I_{k-1} = \{U_{[1,k-1]}^{(1)}, U_{[1,k-1]}^{(2)}, \ldots, U_{[1,k-1]}^{(N)}\}$.

With this understanding, the sensor function at S_l at time k can be regarded as a *quantizer* of the observation $X_k^{(l)}$ that depends on I_{k-1}, i.e., $U_k^{(l)} = \phi_{k,I_{k-1}}^{(l)}(X_k^{(l)})$. The message $U_k^{(l)}$ is assumed to take some value (say, d_l) in the finite set $\{1, \ldots, D_l\}$. Further, we use the notation ϕ_k, d and U_k to denote the corresponding N-dimensional vectors.

The fusion center policy ψ consists of selecting a stopping time τ at which it is decided that the disruption has occurred. In a Bayesian formulation, the goal is to minimize a linear combination of the cost associated with incorrect decision ("*false alarm*") and the cost associated with the delay in detecting the disruption under the assumption that the "*alarm*" signal is correctly given. This leads to the following optimization problem.

Problem (P): Minimize $\mathrm{E}\left[\mathbf{1}_{\{\tau<\Gamma\}} + c(\tau - \Gamma)\mathbf{1}_{\{\tau \geq \Gamma\}}\right]$ over all admissible choices of ψ and $\phi_k^{(l)}, l = 1, \ldots, N, k = 1, 2, \ldots$, where the constant $c > 0$ is the cost of each unit of delay.

II. Results

The solution to (P) is obtained using dynamic programming (DP) arguments. A sufficient statistic at time k for the DP recursions is the posterior probability of the change having happened before time k given I_k, i.e., $p_k = \mathrm{P}(\Gamma \leq k | I_k)$. This one-dimensional sufficient statistic is all that the sensors and fusion center need to store at any given time k, and it can be easily updated using the recursion given below in (1). The complete solution to (P) is stated below.

Theorem 1 (i) The optimum fusion center policy is to stop and declare that a change has occurred at the first k, such that $p_k > a$, where a is the solution to $c + A_J(a) = 1 - a$. (ii) At each time k, it is optimum for the sensors to use monotone likelihood ratio quantizers [4] whose thresholds depend on p_k. Furthermore, a stationary set of sensor functions is optimal, and this set is given by

$$\phi_{p_k}^{\star} = \arg\min_{\phi} W_J(\phi; p_k)$$

where the function J is the unique solution to

$$J(p) = \min\{(1-p), c + A_J(p)\}, \quad \textit{for all } p \in [0,1],$$

and

$$A_J(p) = \min_{\phi} W_J(\phi; p),$$

$$W_J(\phi; p) = \sum_d J\left(\frac{g(d;\phi;p)}{f(d;\phi;p)}\right) f(d;\phi;p),$$

$$g(d;\phi;p) = [p + (1-p)\rho]\; q_{\phi^{(1)}}^1(d_1) \cdots q_{\phi^{(N)}}^1(d_N),$$

$$f(d;\phi;p) = g(d;\phi;p) + (1-\rho)(1-p)\; q_{\phi^{(1)}}^0(d_1) \cdots q_{\phi^{(N)}}^0(d_N),$$

and

$$q_{\phi^{(l)}}^j(d_l) = P_{f_j^{(l)}}\left(\phi^{(l)}(X^{(l)}) = d_l\right).$$

Finally, the recursion for p_k is given by

$$p_{k+1} = \frac{g(U_{k+1}; \phi_{p_k}^{\star}; p_k)}{f(U_{k+1}; \phi_{p_k}^{\star}; p_k)}, \qquad p_0 = \nu \tag{1}$$

References

[1] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.

[2] A. N. Shiryayev, *Optimal Stopping Rules*. New York: Springer-Verlag, 1978.

[3] T. Başar and J. B. Cruz, "Concepts and methods in multi-person coordination and control," in *Optimisation and Control of Dynamic Operational Research Models* (S. G. Tzafestas, Ed.). Amsterdam: North-Holland Publishing Company, 1982.

[4] V. V. Veeravalli, T. Başar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. IT-39, March 1993.

Performance Loss Computation in Distributed Detection

Hamid Amirmehrabi and R. Viswanathan*
Electrical Engineering Dept., Southern Illinois Univ.
Carbondale, IL 62901-6603

Abstract - The loss associated with a distributed signal detection system as compared to a centralized scheme is evaluated with respect to probability of error. Such a loss is numerically computed for several members of the exponential family.

I. INTRODUCTION

An important problem in a Distributed Signal Detection (DSD) scheme is the loss associated with the system. Hence, error analysis plays a significant role in the design of DSD processors. Here, we make an attempt to quantify the loss associated with a DSD system as compared to a centralized scheme by providing an easily computable probability of error expression.

Consider a network of n distributed sensor communicating with a fusion center. Let $\{U_1, U_2, \ldots, U_k\}$ represent the quantized data passed from the sensors numbered 1 through k to the fusion center. Let $\{X_{k+1}, X_{k+2}, \ldots, X_n\}$ represent the observations at the remaining sensors, which are passed directly on to the fusion center without any quantization. Let us assume that U_i's, $i = 1,2,\ldots,k$ are binary valued and that the problem is to decide between two hypotheses H_0 and H_1. Denoting the density of the ith sensor as $f(x_i|H_j)$, $j = 0,1$, and assuming that sensor observations given the hypothesis are independent and identical, we can formulate an optimum fusion center test based on a Likelihood Ratio Test (LRT) [1]. The LRT is given by the following

$$\Lambda_k = C_k . D_k \underset{H_0}{\overset{H_1}{\gtrless}} t_k \qquad (1)$$

where

$$C_k = \frac{f(x_{k+1},\ldots,x_n|H_1)}{f(x_{k+1},\ldots,x_n|H_0)}, \text{ and } D_k = \frac{P(U_1,\ldots,U_k|H_1)}{P(U_1,\ldots,U_k|H_0)} \qquad (2)$$

and t_k is an appropriate threshold.

II. AVERAGE PROBABILITY OF ERROR

The average probability of error corresponding to (1) can be written as

$$P_e(k) = P(H_0)P\left(C_k \geq \frac{t_k}{D_k}|H_0\right) + P(H_1)P\left(C_k < \frac{t_k}{D_k}|H_1\right) \qquad (3)$$

In many problems of practical interest, sufficient statistics of fixed low dimensions exist. Hence, the probability sets involving the C_k in (3) can be replaced by appropriate sets involving the sufficient statistic. Moreover, the D_k in (3) can only take discrete number of values, a maximum of $k+1$ different values. These possible values are $r^j s^{k-j}$, $j = 0,1,\ldots,k$, where

$$r = \frac{P(U_i = 1|H_1)}{P(U_i = 1|H_0)}, \text{ and } S = \frac{P(U_i = 0|H_1)}{P(U_i = 0|H_0)} \qquad (4)$$

Therefore, the probabilities of the type (3) can be very easily computed as a function of k. Such computations are carried out for the case when the density of observation belongs to an exponential family.

III. PERFORMANCE ANALYSIS

Closed form error expressions for gamma, exponential (for testing scale parameter) and normal (for testing location parameter) densities are derived. Table 1 shows the ratio of the error probabilities when $n = 5$ and Signal power to Noise power Ratio (SNR) is 10 dB. α is the shape parameter of the gamma density. As α increases the ratio of the error probabilities also increases.

	Exponential	Normal	Gamma , $\alpha = 3$
$\frac{P_e(2)}{P_e(1)}$	1.2	1.6	2.0
$\frac{P_e(4)}{P_e(1)}$	1.8	4.4	8.0

Table 1

Numerical results indicate that for normal and gamma (with large α) densities the loss due to quantization is more significant than for exponential density.

*This work was supported by BMDIO/IST and managed by the office of naval research under contract N00014-94-1-0736.

REFERENCE

[1] J.N. Tsitsiklis, "Decentralized Detection," in *Advances in Statistical Signal Processing* , Vol. 2., *Signal Detection*, H.V. Poor and J.B. Thomas, Eds., Greenwich, CT: JAI press, 1990.

HOS-based noise models for signal-detection optimization in non-Gaussian environments

A. Tesei, and C.S. Regazzoni
DIBE, University of Genoa, Genoa, Italy

Abstract - **Two pdf models suitable for describing non-Gaussian iid noise are introduced. The models are used in the design of a LOD test for detecting weak signals in real non-Gaussian noise. Results obtained in the context of an underwater acoustic application are encouraging.**

1. INTRODUCTION[1]

Conventional signal processing and detection criteria, optimised in presence of Gaussian noise, may decay in non-Gaussian environments. Higher Order Statistics (HOS) [1] is a powerful means to analyse non-Gaussian noise and build robust detectors.

This work focuses attention on the problem of *optimizing detection in presence of additive, iid, stationary, non-Gaussian noise under the conditions of weak signals* (i.e., for low Signal-to-Noise Ratio - SNR). In order to optimize the Probability of Detection P_{det} for low SNR values, the selected binary statistical test consists in a Locally Optimum Detector (LOD) [2], whose test rule is computed on the basis of new models of noise univariate probability density function (pdf) [3]. The investigated models are expressed in terms of the HOS parameters *skewness* (of the 3rd-order) which quantifies the deviation from shape symmetry, and *kurtosis* (of the 4th order) which quantifies the sharpness of a shape. The detector has been tested in the case of deterministic signals corrupted by real shipping-traffic noise, acquired during a sea campaign, in the context of CEC MAST-I SNECOW project (May 1993) [4].

2. DESCRIPTION OF THE APPROACH

The proposed method is based on the statistical analysis of channel noise. As LOD requires the *analytical model* of noise pdf, attention is focused on this aspect. The **first** model is a generic pdf introduced by Champernowne and used in [3]. It can be applied if the N noise components have an hyperbolic distribution of power. In this acoustic application, in which noise main components are the ship, from which the sensor was dropped (strong source), and the surrounding traffic ships (which can be considered equally distributed on the sea, and contribute weakly to noise), this pdf model is reasonable. It depends on β_2, the ratio between the 4th and the square of the 2nd moments [3]. A **second** new model is presented, the "asymmetric Gaussian" pdf, consisting of two Gaussian parts, and depending on *two second-order parameters* (deriving from the definition of variance), i.e., the "left and right variances", which together maintain the same information provided by the skewness. The non-linear function $g_{lo}(\cdot)$, in terms of which the likelihood-ratio of the LOD rule is expressed [2], is easily expressed in terms of these two models. Information added by HOS-based description is contained in simple parameters (β_2 or σ_l and σ_r), and no constraint has to be satisfied about signal characteristics.

3. EXPERIMENTAL RESULTS AND FUTURE WORK

An extensive test phase was carried out. Noise was acquired in a coastal shallow-water area. The presence of a lot of traffic and of reflection and refraction makes the detector work in critical conditions. The LOD performances are summarized in Fig. 1 in terms of P_{det} vs. SNR. A comparison among the results of the two proposed pdfs and the Gaussian model is presented.

The tests were carried out by fixing the Probability of False Alarm $P_{FA}=\alpha=5\%$. Non Gaussian real underwater acoustic ship-traffic noise was characterized by $\mu=0$, $\beta_2=2.84$, $\sigma_l= 1860$, $\sigma_r=1500$. The proposed models appear approximately equivalent, as noise presents deviation from both Gaussian sharpness and symmetry. The next investigation step, concerning the model of propagation through a real shallow-water channel, is going to be carried out.

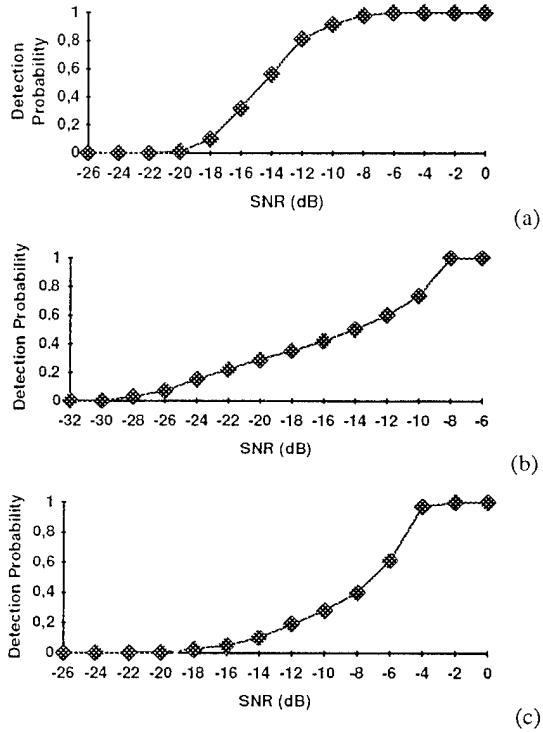


Fig. 1 Results of the LO detector under the *Champernowne* (a), the *asymmetric-Gaussian (b)* and the *Gaussian* (c) *hypotheses.*

REFERENCES

[1] C. Nikias, J. Mendel, "Signal Processing with Higher-Order Spectra", *IEEE SP Mag.*, pp. 10-37, July 1993.

[2] S.A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer Verlag, Berlin, 1988.

[3] R.J. Webster, "Ambient Noise Statistics", *IEEE Trans. SP*, Vol. 41 (6), pp. 2249-53, 1993.

[4] DIBE, *SNECOW Project-MAST 0029-C(A) Final Report-Task 5: Ship traffic noise statistical evaluation*, Dec. 1993.

[1] This work was partially supported by the Commission of European Community in the context of MAST-I SNECOW Project

Optimum Detection of Gaussian Signals in non-Gaussian Noise

S. Buzzi, E. Conte, M. Lops
Dipartimento di Ingegneria Elettronica
Via Claudio 21, 80125 Napoli, Italy

Abstract —

This paper handles the detection of Gaussian signals in compound-Gaussian noise. We show that the optimum detector is the conventional one plus an estimator of the short-time noise power spectral density.

I. Introduction

In this paper we consider the problem of detecting one out of M Gaussian processes with known autocorrelation functions (acf's) in the presence of non-Gaussian noise: such a problem is commonly encountered in radio communications over fading dispersive channels subject to atmospheric noise. Denoting by $a_1(t), a_2(t), \ldots, a_M(t)$ M complex Gaussian random processes with given acf's, the detection problem under study amounts to the following M-ary hypothesis test:

$$H_i: \quad r(t) = a_i(t) + c(t) \tag{1}$$

wherein $r(t)$ and $c(t)$ denote the complex envelopes of the received signal and of the impinging noise, respectively. Such a noise is modeled as a compound-Gaussian process, namely as the product of a real, non-negative component, $s(t)$ say, times an independent, Gaussian, possibly complex process, $g(t)$. Theoretical considerations, supported by experimental results, show that, if the correlation time of $s(t)$ is much smaller than that of $g(t)$, then the model represents a faithful description of some important noise sources, such as atmospheric noise and scattering from composite surfaces (see [1] and references thereof). Since the signalling interval is typically much smaller than the average decorrelation time of $s(t)$, the modulating process degenerates into a random constant and the noise process reduces to a Spherically Invariant Random Process (SIRP).

II. Receiver Design

We focus on the case of uncorrelated noise observations with Power Spectral Density (PSD) $2\mathcal{N}_0 \mathrm{E}[s^2]$ (where $2\mathcal{N}_0$ is the PSD of the Gaussian component and $\mathrm{E}[\cdot]$ denotes statistical expectation), since, due to the closure of both Gaussian processes and SIRP's with respect to linear transformations, the case of correlated noise can be easily handled via whitening approach.

Denoting by $\Lambda_g[r(t); 2\mathcal{N}_0 | H_i]$ the likelihood functional under hypothesis H_i for complex, uncorrelated Gaussian noise with PSD $2\mathcal{N}_0$, the likelihood functionals in the presence of SIRP can be shown to assume the form

$$\Lambda[r(t)|H_i] = \Lambda_g[r(t); 2\mathcal{N}_0 \lim_{N\to\infty} \widehat{s_N^2} | H_i] \tag{2}$$

wherein $\widehat{s_N^2}$ represents a consistent estimator of the random variable s^2 and can be computed by properly processing the observables. Otherwise stated, since the noise process, as observed in sufficiently short time intervals, is a conditionally Gaussian random process, then the likelihood functionals coincide with those for Gaussian noise, provided that the noise PSD is substituted by an estimate of the short-term PSD (i.e., of the conditional noise PSD given s). This fact does not entail that the conventional detector is optimum under SIRP disturbance, since the estimator-correlator is to be keyed to the estimated short-term noise PSD. In any case, we stress here that the receiver is canonical, in the sense that its structure is one and the same, independent of the probability density function (pdf) of the modulating process and, hence, of the statistics of the noise process.

So far, the structure of the estimator $\widehat{s_N^2}$ has been left aside: interestingly, it can be shown to coincide with the average of the square modula of the projections of the first N versors of the received signal along any orthonormal basis of the space $\mathrm{L}^2(0,T)$; as $N \to \infty$, $\widehat{s_N^2}$ can be shown to converge in the mean square sense to the random variable s^2.

Choosing the complex exponentials of period T as a basis yields

$$\widehat{s_N^2} = \frac{1}{2\mathcal{N}_0} \frac{1}{NT} \sum_{k=1}^{N} \left| R_T\left(\frac{k}{T}\right) \right|^2 \tag{3}$$

where $R_T(f)$ is the Fourier Transform of the received signal, as observed in the interval $(0,T)$: thus, $\widehat{s_N^2}$ is an average of the sampled periodogram of the received signal.

Summing up, the minimum error-probability decision rule for equally likely signals is written as

$$\text{decide } H_i: \int_0^T r(t)\widehat{a}_i^*(t)dt - b_i > \int_0^T r(t)\widehat{a}_k^*(t)dt - b_k \ \forall k \neq i \tag{4}$$

wherein $\widehat{a}_k(t)$ is the *linear* minimum mean-square estimation of the $k-$th signal in Gaussian noise with PSD $2\mathcal{N}_0 s^2$ and $b_i = b_i(s^2)$ are proper bias terms, depending on the value of the noise short-term PSD.

III. Performance analysis

As to the performance of this detector, the analysis of On-Off Keying (OOK) signalling with exponential correlation subject to noise with Laplacian pdf demonstrates that the error probability depends on two parameters, the ratio of the received energy to the noise long-term PSD and the time-bandwidth product of the signal, namely the product of the correlation length times the spectral width of the Gaussian random process. Interestingly, as for the case of Gaussian noise, the larger such a product, the better the performance. Additionally, the noise spikyness seems not to dramatically affect the performance, even though, as for the case of non-dispersive channels, increased noise spykiness results in worse and worse performance, especially in the interest region of extremely low error probabilities.

References

[1] E. Conte, M. Di Bisceglie, M. Lops, "Optimum Detection of Fading Signals in Impulsive Noise", *IEEE Trans. on Communications*, April 1995.

Asymptotically Robust Detection Using Statistical Moments

Kevin R. Kolodziejski* and John W. Betz†

*Center for Communications and Digital Signal Processing (CDSP)
Dept. of Electrical and Computer Eng., Northeastern University, Boston, MA 02115

†The MITRE Corporation, Bedford, MA 01730

Abstract — While locally optimum detection requires complete knowledge of the noise density, we use only the first few absolute moments of the independent, identically distributed (iid) noise to obtain a robust detector that is locally optimum for the least favorable noise satisfying these moments. This robust detector's efficacy approaches that of the asymptotically optimum detector, while requiring limited knowledge of the noise statistics.

I. SIGNAL AND NOISE MODEL

The problem is modeled as deciding between the null hypothesis $\mathbf{X}=\mathbf{W}$ and the alternative hypothesis $\mathbf{X}=\theta\mathbf{s}+\mathbf{W}$ where $\mathbf{X}$ is an n-element observation vector, $\mathbf{W}$ is a vector of zero-mean iid noise random variables with univariate density f, $\mathbf{s}$ is a vector of known signal samples with nonzero, finite asymptotic average power, and $\theta=K/\sqrt{n}$, for some unknown $K>0$.

II. COMPLETELY KNOWN NOISE STATISTICS

The locally optimum (LO) detector of a known signal in iid noise is a memoryless nonlinearity followed by a correlator [1]. The memoryless nonlinearity g depends on the noise density by $g(x)=-f'(x)/f(x)$, where $f'(x)=df(x)/dx$. When the noise is zero-mean Gaussian with unit variance, $g(x)=x$ and the LO detector is a linear correlator. A generalization of the LO detector is a nonlinear correlator where g is any function satisfying mild regularity conditions. A common example is the sign correlator, whose nonlinearity is the signum function.

Efficacy $\eta(g,f)$ is an asymptotic measure for predicting detection performance. In the asymptotic case, $n\to\infty$ which implies $\theta\to 0$. The asymptotic LO detector is equivalent to the asymptotically optimum (AO) Neyman-Pearson detector, and its efficacy is equal to Fisher information $I(f)$. $\eta(g,f)$ is concave in g and convex in f and satisfies the saddle point inequalities $\eta(g,f_0)\le\eta(g_0,f_0)=I(f_0)\le\eta(g_0,f)$, where $g_0=-f_0'/f_0$ for some density f_0. At the saddle point, efficacy is equal to Fisher information.

III. PARTIALLY KNOWN NOISE STATISTICS

Only the first few absolute moments of the noise are assumed known. The admissible set of absolutely continuous densities is $\mathcal{F}=\{f\,|\int|x|^j f(x)dx=v_j,\ \ j=1,2,\ldots,J\}$, where J is typically 2 or 3. It can be shown that there exists a least favorable density $f_{\mathrm{LF}}\in\mathcal{F}$ such that $I(f_{\mathrm{LF}})=\inf I(f)$ for all $f\in\mathcal{F}$. Since it is difficult to analytically determine f_{LF}, a Gram-Charlier series approximation [2] is used to model the noise densities in the admissible set. Many terms are used to develop good Gram-Charlier series approximations for the $f\in\mathcal{F}$. Constrained numerical optimization is used to find the series coefficients that determine the least favorable density. The robust detector is a nonlinear correlator with nonlinearity $g_{\mathrm{LF}}=-f_{\mathrm{LF}}'/f_{\mathrm{LF}}$ that is LO for this least favorable noise density.

IV. NUMERICAL RESULTS

Our results demonstrate that only the first few absolute moments of the noise are needed to approach the performance of the AO detector derived with full knowledge of the noise density. Apparently, these low-order absolute moments are most influential in determining the shape of the density about the mode, and therefore in shaping the nonlinearity at values most often occupied by the noise. We have used the first two and three absolute moments of Gaussian-Gaussian mixture (GGM) and Johnson distributions to derive the LO nonlinearity (NL) g_{LF} for the least favorable noise density. The efficacy of the resulting asymptotically robust detector $\eta(g_{\mathrm{LF}},f)$ is only slightly less than that of the AO detector and is significantly greater than the asymptotic linear and sign correlators. Fig. 1 shows efficacy results for detectors in one example: the density $f=f_{\mathrm{GGM}}$ is from a unit-variance Gaussian-Gaussian mixture class with a contamination parameter of 0.05, and the first two absolute moments of f_{GGM} are used to obtain f_{LF} and hence g_{LF}. Fifty terms were used in the Gram-Charlier series. The abscissa in Fig. 1 is the ratio of the contamination variance to the nominal variance of the two Gaussian distributions comprising the Gaussian-Gaussian mixture. When this ratio is one, the noise is Gaussian. The results are computed for correlators preceded by a linearity, sign NL, robust NL, and GGM LO NL using GGM noise. The robust NL's performance is also shown for the least favorable noise for which the robust NL is LO; while this noise satisfies the moment constraints, it is not GGM. Clearly, the performance of the robust detector approximates that of the AO detector, far exceeding that of a linear correlator or a sign correlator.

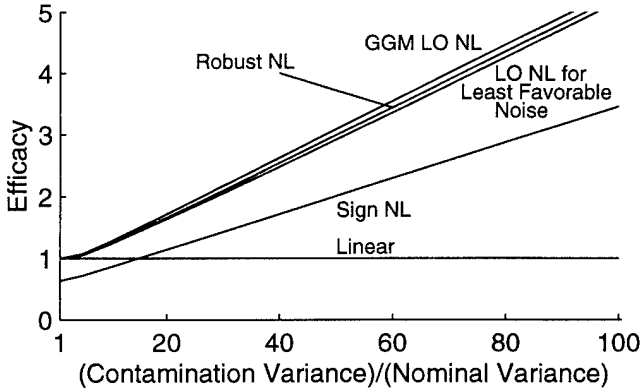


Fig. 1. Comparison of Detector Efficacy for Example

REFERENCES

[1] S. A Kassam, *Signal Detection in Non-Gaussian Noise,* New York, NY: Springer-Verlag, 1987.

[2] A. D. Whalen, *Detection of Signals in Noise*, New York, NY: Academic Press, 1971.

Any opinions expressed in this paper do not necessarily represent those of The MITRE Corporation.

Conditional Testing In Two-Input Detectors With Single Input Conditioning

Babak Seyfe and Masoud Kahrizi[1]
Dept. of Elect. Eng. ,Tarbiat Modarres University, Tehran, Iran

Abstract- **In this paper we examine some new methods in conditional testing in two-input signal detection with condition on one of the inputs. In this method ,the number of samples has been considerably reduced.**

I. INTRODUCTION

In a conditional test the threshold and randomization probability of a threshold test are not taken to be fixed parameters independent of the data but are directly dependent on the specific data set being analyzed for a test of hypotheses. In our paper, conditional testing in two-input detectors is performed by using only one of the inputs. Asymptotic Relative Efficiency (ARE) has been computed with respect to Generalized Cross Correlation (GCC) ,e.g.[1]. Also this method is performed on two-input optimum Three Level Coincidence (TLC) correlator , e.g.[2].

II . DETECTOR STRUCTURE

Consider a binary problem with a null hypotheses H and an alternative hypotheses K and let $\overline{X}_n = (x_1, x_2, \ldots, x_n)$ and $\overline{Y}_n = (y_1, y_2, \ldots, y_n)$ denote the n-component random observation vectors. A fixed threshold test for H against K is compared with test function $T(\overline{X}_n, \overline{Y}_n)$. Block diagram of our detector is shown in Fig.1

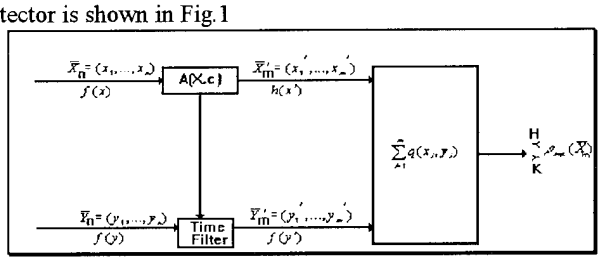


Fig.1

$A(\overline{X}, c)$ is a function of $\overline{X}_n$ and a parameter c . f and h are pdf of noise inputs. The threshold $\rho_{n,\alpha}$ is a function of $\overline{X}_n$ *or* $\overline{Y}_n$. When $\overline{X}_n$ is passed through the block $A(\overline{X}, c)$, a subvector $\overline{X}'_m$ has been formed from $\overline{X}_n$ comes to detector , where $m \le n$ and in general case m is a random variable.

III . ASYMPTOTIC PERFORMANCE

If the functions f and h are even functions and components of $\overline{X}_n$ and $\overline{Y}_n$ are iid and $\lim_{n\to+\infty}(m/n) = k$,Then the efficacy of two input conditional testing with single input conditioning according Fig .1 is

$$\varepsilon_{cond} = \frac{k\left[\int_{-\infty-\infty}^{+\infty+\infty} q(x,y)\left[h''(x)f(y)+2h'(x)f'(y)+h(x)f''(y)+\right]dxdy\right]^2}{4\int_{-\infty-\infty}^{+\infty+\infty} q^2(x,y)h(x)f(y)dxdy}, \quad (1)$$

IV. SPECIAL CASES

Let $A(\overline{X}, c)$ is a function as in Fig.2

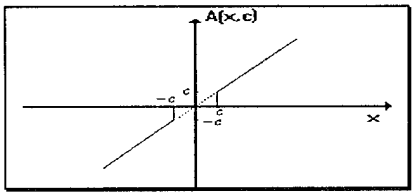


Fig.2

Then Fig . 3 shows the $ARE_{cond.GCC,GCC}$ and $\overline{m}$ for Gaussian noise .

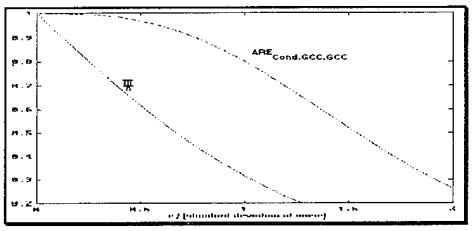


Fig.3

The ARE of conditional TLC detector with respect to TLC ,e.g. [2], in Gaussian noise has been shown in Fig. 4 .

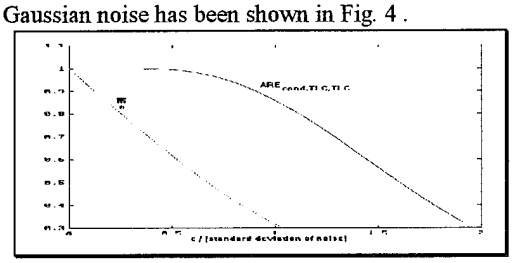


Fig. 4

V. CONCLUSION

For single input conditional testing in two inputs detector when the noise is Gaussian or nearly Gaussian we have an appropriate method for detection . Also the number of observation for process is considerably reduced . The percentage of this reduction depends on the number of input samples, n as shown in Fig.3 and Fig.4 . However , the percentage of time processing reduction is much higher than that for the samples. The detail of the method is presented in [3].

REFERENCES

[1] Kassam , S.A, ***Signal Detection in Non Gaussian Noise,*** Springer Verlag ,1988.

[2] Shin ,J.G. and Kassam ,S.A,"Three Level Coincidence - Correlator Detectors, "J. Acoust. Soc. Amer. ,vol. 63, pp: 1389-1395,1978.

[3] Seyfe , B.,"Non-parametric Detection With Emphasis on RADAR Pulses," MS Thesis , Dept. Elect. Eng., Tarbiat Modarres University, Tehran,Iran,1995 (In Persian).

[1]This work is supported by ***Institute of Science and Technology*** , Tehran , Iran .

Signal Detection in Continuous-Time White Gaussian Channel

Shunsuke Ihara
School of Informatics and Sciences
Nagoya Univ., Nagoya, 464-01 Japan

Yusuke Sakuma
Graduate School of Human Informatics
Nagoya Univ., Nagoya, 464-01 Japan

Abstract — **We consider a signal detection problem in a continuous-time white Gaussian channel. The signal is assumed to be a stationary Gaussian process. We prove that the error probability in the signal detection tends to zero exponentially fast, as the observation time goes to infinity.**

Summary

The aim is to study the exponential-type asymptotic behavior of the error probabilities of the signal detection in a continuous-time white Gaussian channel (WGC). The model of a WGC is presented by

$$Y(t) = \int_0^t X(s)ds + B(t), \quad t \in [0,T],$$

or

$$\dot{Y}(t) = X(t) + \dot{B}(t),$$

where $\{\dot{B}(t)\}$ is a Gaussian white noise, $X(t)$ and $\dot{Y}(t)$ are a channel input and the corresponding output, respectively. The signal detection problem consists of deciding, based on the observation of the output $\{Y(t)\}$, whether the signal $\{X(t)\}$ is sent or not. In other words, we consider testing problem of two hypotheses

$$H_0: \quad Y(t) = B(t), \qquad t \in [0,T],$$
$$H_1: \quad Y(t) = \int_0^t X(s)ds + B(t), \quad t \in [0,T].$$

Two probabilities of error are defined by

$$e_0(T) = \Pr(\{Y(t)\} \notin \mathcal{S} | H_0 \text{ is true}),$$
$$e_1(T) = \Pr(\{Y(t)\} \in \mathcal{S} | H_1 \text{ is true}),$$

with a decision region $\mathcal{S} \subset \mathbf{R}^{[0,T]}$. A Neyman-Pearson test is a test given by a decision region of the form

$$\mathcal{S}_T(u) = \left\{ y \in \mathbf{R}^{[0,T]}; \frac{1}{T} \log \frac{d\mu_0^T}{d\mu_1^T}(y) < u \right\},$$

where μ_i^T is the probability distribution of $\{Y(t)\}$ under the hypothesis H_i. It is well known that Neyman-Pearson tests are optimal to minimize $e_1(T)$, where u is chosen so that $e_0(T) = \mu_0^T(\mathcal{S}_T(u)^c)$. In this casse, $e_1(T) = \mu_1^T(\mathcal{S}_T(u))$.

We assume that the signal $\{X(t)\}$ is a regular stationary Gaussian process with spectral density function (SDF) f. Note that $\{\dot{B}(t)\}$ is a generalized stationary process with SDF $f_0(\lambda) = 1/(2\pi)$ and that, under H_1, $\{\dot{Y}(t)\}$ is a stationary process with SDF $f_1(\lambda) = f(\lambda) + 1/(2\pi)$. To state the result we define a SDF f_θ by

$$1/f_\theta(\lambda) = (1-\theta)/f_0(\lambda) + \theta/f_1(\lambda).$$

We define $\overline{H}(f;g)$, for each SDF's f and g, by

$$\overline{H}(f;g) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \left(\frac{f(\lambda)}{g(\lambda)} - 1 - \log \frac{f(\lambda)}{g(\lambda)} \right) d\lambda.$$

We can show that $\overline{H}(f_\theta; f_0)$ is the relative entropy (or information divergence) of a stationary Gaussian process with SDF f_θ with respect to the white noise $\{\dot{B}(t)\}$.

Concerning the exponential-type asymptotic behavior of the error probabilities, we can prove the following theorem.

Theorem 1 Assume that the SDF f is continuous. Then, for any $\alpha > 0$, there exists a constant u_α such that

$$\lim_{T\to\infty} T^{-1} \log \mu_0^T(\mathcal{S}_T(u_\alpha)) = -\alpha.$$

If $0 < \alpha = \overline{H}(f_\theta; f_0) < \overline{H}(f_1; f_0)$ $(0 < \theta < 1)$, then

$$\lim_{T\to\infty} T^{-1} \log \mu_1^T(\mathcal{S}_T(u_\alpha)^c) = -\overline{H}(f_\theta; f_1).$$

If $\alpha = \overline{H}(f_\theta; f_0) > \overline{H}(f_1; f_0)$ $(\theta > 1)$, then

$$\lim_{T\to\infty} T^{-1} \log\{1 - \mu_1^T(\mathcal{S}_T(u_\alpha)^c)\} = -\overline{H}(f_\theta; f_1).$$

The proof is based on a large deviation theorem.

In discrete-time cases, the asymptotic behavior of error probabilities in hypothesis testing has been studied [1, 2].

References

[1] R.E. Blahut, "Hypothesis testing and information theory," IEEE Trans. Inform. Theory, IT-20, pp. 405–415, 1974.

[2] T.S. Han and K. Kobayashi, "The strong converse theorem in hypothesis testing," IEEE Trans. Inform. Theory, IT-35, pp. 178–180, 1989.

A Recursive Formulation for Quadratic Detection on Rayleigh Fading Channels

Piero Castoldi and Riccardo Raheli

Dipartimento di Ingegneria dell'Informazione, Università di Parma, 43100 Parma, Italy

I. Introduction

We address the problem of optimal detection of a random signal transmitted over a time-varying frequency-selective correlated Rayleigh fading channel. We present a general recursive solution which may be operated at full complexity to provide optimal detection or at reduced complexity, using Per-Survivor Processing (PSP) techniques [1], to yield a suboptimal receiver. An alternative full-complexity solution based on the innovations approach may be found in [2].

II. Proposed Recursive Receiver

To derive the optimal receiver structure, we adopt a discrete-time representation of the received signal obtained by Nyquist sampling (let β be the resulting oversampling factor). We denote with $\{r_n\}_{n=1}^{\infty}$ the samples of the received signal (sufficient statistics), $\{a_k\}_{k=1}^{\infty}$ the information sequence and L_{ISI} the intersymbol interference length in symbol periods. We also denote with L_c the channel coherence time expressed in symbol periods and assume it finite. Let $L = L_{ISI} + L_c$ and define the following vectors: $\mathbf{r}_{i\beta} = (r_1, r_2, \ldots, r_{i\beta})$, $\mathbf{a}_i = (a_{1-L}, a_{2-L}, \ldots, a_0, a_1, \ldots, a_i)$, $\mathbf{r}'_{i\beta} = (r_{(i-L_c)\beta+1}, r_{(i-L_c)\beta+2}, \ldots, r_{i\beta})$, $\mathbf{a}'_i = (a_{i+1-L}, \ldots, a_i)$, $\mathbf{r}''_{i\beta} = (\mathbf{r}'_{(i-1)\beta} \vdots r_{(i-1)\beta+1}, r_{(i-1)\beta+2}, \ldots, r_{i\beta})$, $\mathbf{a}''_i = (\mathbf{a}'_{i-1} \vdots a_i)$, where $(\bullet \vdots \bullet)$ denotes vector concatenation.

Optimal detection requires to perform the maximization $\hat{\mathbf{a}}_i = \arg\max_{\mathbf{a}_i} p(\mathbf{r}_{i\beta}|\mathbf{a}_i)$ where $p(\mathbf{r}_{i\beta}|\mathbf{a}_i)$ is the conditional Probability Density Function (PDF) of $\mathbf{r}_{i\beta}$ given $\mathbf{a}_i$. Because of the assumed channel model, this PDF is multivariate zero-mean Gaussian. Due to the limited channel coherence time L_c, it is possible to factorize this PDF. By the second Bayes theorem, each factor may be expressed as a ratio of the PDFs of the partial observation vectors $\mathbf{r}'_{(k-1)\beta}$ and $\mathbf{r}''_{k\beta}$, which do not depend on the complete data sequence but only on $\mathbf{a}'_{k-1}$ and $\mathbf{a}''_k$, respectively. The parameter L, which is the length of the vector $\mathbf{a}'_{k-1}$, plays the role of an overall channel memory, as pointed out also in [2].

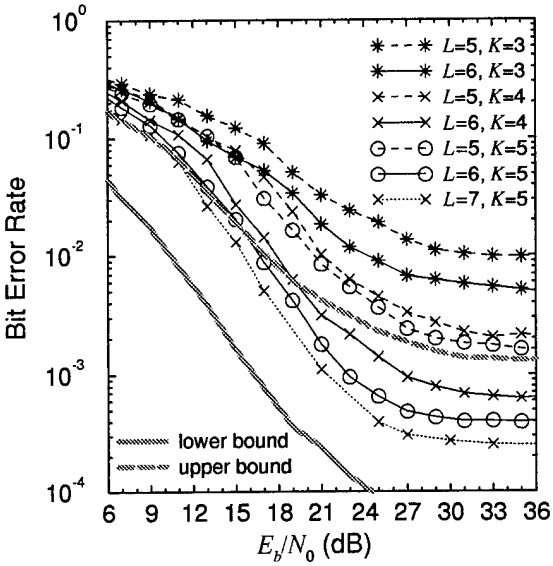
Making use of the correlation matrices $\mathbf{R}_{\mathbf{r}'_{k\beta}}(\mathbf{a}'_k) = E[\mathbf{r}'^{\mathrm{H}}_{k\beta}\mathbf{r}'_{k\beta} \,|\, \mathbf{a}'_k]$, $\mathbf{R}_{\mathbf{r}''_{k\beta}}(\mathbf{a}''_k) = E[\mathbf{r}''^{\mathrm{H}}_{k\beta}\mathbf{r}''_{k\beta} \,|\, \mathbf{a}''_k]$, we can express the likelihood function (path metric) to be minimized as:

$$\Lambda_i(\mathbf{a}_i) = \sum_{k=1}^{i} \log\left(\frac{\det \mathbf{R}_{\mathbf{r}''_{k\beta}}(\mathbf{a}'_{k-1} \vdots a_k)}{\det \mathbf{R}_{\mathbf{r}'_{(k-1)\beta}}(\mathbf{a}'_{k-1})}\right) + \quad (1)$$

$$+ \mathbf{r}''^{\mathrm{H}}_{k\beta}\mathbf{R}^{-1}_{\mathbf{r}''_{k\beta}}(\mathbf{a}'_{k-1} \vdots a_k)\,\mathbf{r}''_{k\beta} - \mathbf{r}'^{\mathrm{H}}_{(k-1)\beta}\mathbf{R}^{-1}_{\mathbf{r}'_{(k-1)\beta}}(\mathbf{a}'_{k-1})\,\mathbf{r}'_{(k-1)\beta}$$

where $\det(\cdot)$ denotes the determinant of a matrix and $[\cdot]^{\mathrm{H}}$ is the Hermitian operator. The above minimization can be performed by searching the optimum path in a trellis diagram whose state is defined as $\mu_k = \mathbf{a}'_k$. This search may become prohibitive for highly correlated channels, since the number of trellis states might be very high (M^L, if M is the number of constellation symbols).

An alternative suboptimal solution is offered by well-known PSP techniques. We define a reduced state $\tilde{\mu}_k = \tilde{\mathbf{a}}_k = (a_{k-K+1}, a_{k-K+2}, \ldots, a_k)$ where K ($1 \le K \le L$) is an integer which controls the degree of desired complexity reduction and a vector $\tilde{\mathbf{a}}'_k = (\hat{a}_{k+1-L}, \ldots, \hat{a}_{k-K} \vdots \tilde{\mathbf{a}}_k)$, in which $\{\hat{a}_i; i = k-K, \ldots, k-L\}$ denote information symbols associated with the survivor of state $\tilde{\mu}_k$. The resulting path metric is formally identical to (1) after substituting $\mathbf{a}'_{k-1}$ with $\tilde{\mathbf{a}}'_{k-1}$.



III. Numerical Results and Conclusions

The performance of the proposed receivers is assessed in terms of Bit Error Rate (BER) versus E_b/N_0 (E_b is the bit energy averaged over channel and data statistics). The overall channel is a symbol-spaced ($\beta = 1$) finite impulse response filter with three independent taps, modeled as first order autoregressive (the forgetting factor is 0.998). For a QPSK modulation format, blocked transmission with blocks of 60 symbols is assumed. A preamble and tail both of 2 symbols are used.

In the figure, BER of the proposed detectors is compared to lower and upper bounds derived as in [3]. Complexity savings (M^K instead of M^L trellis states) may be achieved with the proposed suboptimal algorithms based on PSP at the expense of a moderate performance loss (compare the performance when $L = 5$ for $K = 5, 4$). Furthermore for an equal number of trellis states ($K = 5$), PSP allows to improve significantly the performance by increasing the assumed channel memory from $L = 5$ to 6 and 7. In three cases the proposed receiver performance lies between the lower and upper bounds.

References

[1] R. Raheli, A. Polydoros, C.-K. Tzou, "Per-survivor processing: a general approach to MLSE in uncertain environments," *IEEE Trans. on Commun.*, Feb.-Apr. 1995.

[2] X. Yu, S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *Proc. of Pacific Rim Conf. on Commun., Computers and Signal Proc.*, Victoria, B.C., Canada, May 1993.

[3] P. Castoldi, R. Raheli, "Optimal versus PSP-based sequence estimation for Rayleigh fading channels," *Proc. of Intern. Conf. on Personal Wireless Commun.*, Sidney, B.C., Canada, June 1995.

Robust Detection of Impulse Signals in Random Impulse Interferences

Yuri P. Grishin, Alexej I. Sokolov, Yuri S. Yurchenko

Dept. Radio Eng. , St.-Petersburg State Electrical Engineering University
Prof. Popova st. 5, 297376, St.-Petersburg, Russia

Abstract - **New robust detection algorithms have been developed for detection of pulse signals in the presence of a random noise and random pulse interferences. The algorithms are designed in assumption that a priori information on compactness of the useful pulse signals is known. It is shown that estimation of the noninformative signal parameters decreases the detection quality. The numerical simulation of the proposed algorithm is carried out.**

I. Introduction

The problem of detecting pulse signals in the presence of a random noise and random pulse interferences is of great significance for time synchronization channels of TDMA systems, for radionavigation (VOR/DME) and radars [1,2,3]. The presence of pulse interferences can lead to appearance of outliers at the input of a signal detector. The outliers considerably complicate the solution of the problem [4].

As the input the detector is assumed to use time delay of the received signal which can be written and stored in a detector memory. Thus in the time domain the problem of robust signal detection can be formulated in the following way:

$$\begin{aligned} H_0 &: x_i = \varepsilon_i , \\ H_1 &: x_i = \gamma_i (\theta_i + \upsilon_i) + (1-\gamma_i)\varepsilon_i , \end{aligned} \qquad (1)$$

where x_i is observation (time delay), ε_i - time delay of the interference impulse, γ_i -a random sequence with a value 1 when x_i belongs to an informative (signal) set and zero when x_i belongs to a noninformative (interference) set, ν_i - time delay estimation error due to a random noise. Conditional probability density functions (pdf) $f_1(x_i/\Theta_i,\gamma_i=1)$ and $f_2(x_i/\gamma_i=0)$ are assumed to be either normal or exponential.

A dynamics of time delay of the received signal can be described by a difference equation $y_k = \phi\, y_{k-1} + \omega_k$, $\theta_k = H_k\, y_k$, where all symbols are commonly used.

II. Robust Detection Algorithms

For solving the problem it is necessary to calculate the generalized likelihood ratio (GLR) $l(x_n/\Theta,\Gamma) = f(x_n/\Theta,\Gamma,H_1)/f(x_n/H_0)$, where $\Theta=[\Theta_1,\ldots,\Theta_n]^T$ is the vector of informative parameters and $\Gamma=[\gamma_1,\ldots,\gamma_n]^T$ is the vector of noninformative parameters.

The detection statistics can be obtained either by averaging the GLR by all possible values of noninformative parameters or by estimation of them (the case of classification of the received signal). In both cases first of all it is necessary to estimate the informative parameter vector that is to estimate time delay of the received signals. The problem is complicated by the presence of outliers in the observations. In this paper we developed a fixed-interval smoothing algorithm on the basis of the invariant embedding method [5]. This algorithm showed a high accuracy of the estimates and it consists of two nonlinear Kalman filters of which one is a backward filter. A matrix gain of the filter depends on a posteriori probabilities $P(\gamma_i = j/x_i), j=0,1$.

As was mentioned above one way of developing an optimal detection statistics is averaging the GLR by all noninformative parameters. It is easy to show that in this approach the likelihood ratio logarithm can be written as:

$$l_1(X_n) = \sum_{i=1}^{n} ln\left[1 - P(\gamma_i = 1/x_i,\hat{\Theta})\right]. \qquad (2)$$

The statistics (2) is optimal for given vector $\hat{\Theta}$. The other way is to estimate the noninformative parameters of the received signal. Two possible situations which can be encountered in practice have been considered:

1). If the number of the signal samples q is known, we can classify as the signals those of them which has maximum value of $P(\gamma_i = 1/x_i,\hat{\Theta})$. Then it follows that

$$l_2(X_n) = \sum_{j=1}^{q} ln\left[f_1(x_j/\gamma_j = j,\hat{\Theta})/f_2(x_j/\gamma_j = 0)\right]; \quad (3)$$

2). For unknown number of signal samples q the estimate of the vector Γ can be found as a maximum of the a posteriori probability $P(\Gamma/x,\Theta)$. In this case the expression (3) can be written in approximate form

$$l_3(X_n) = \sum_{i=1}^{n} \hat{\gamma}_i . \qquad (4)$$

III. Conclusion

The computer modelling of proposed algorithms for a radionavigational system was carried out for Gaussian and Laplace pdf of contaminated observations. The best results were obtained for Laplace pdf because of the great contrast between pdf of normal measurements and outliers. The algorithms (2) and (3) showed a higher efficiency in comparison with a nonparametric (median) signal detector which is usually used in such a situation. The algorithm (4) had practically the same characteristic as the median detector. It should be noted that all proposed algorithms are sensitive to a priori information on probability p.

References

[1] Y. Bar-Shalom, *Tracking and Data Association*, Academic Press, 1988.

[2] C.E.Cook, M.Bernfeld, *Radar Signals: An Introduction to Theory and Application*, Artech House 1993.

[3] Yu.P. Grishin, Yu.M. Kazarinov, *Fault - Tolerant Dynamic Systems*, (in Russian). Moscow, Radio i Svjaz, 1985.

[4] S.A Kassam, H.V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol.73, no.3, pp.433 - 481, 1985.

[5] S.B. Dobrohodov, A.I. Sokolov, Yu.S. Yurchenko, "Fix - interval smoothing algorithms in the presence of noninformative measurements," *Radioelectronics and Communications Systems*, Allergton Press, vol.35, no.7, 1992.

Rotating Group Codes for the ISI Channel

Peter Massey and Peter Mathys
Dept. of Elect. and Comp. Eng.
University of Colorado
Boulder, CO 80309-0425
massey@prony.colorado.edu

Abstract — **Time-varying mappings are used in place of a stationary mapping to improve the performance of Euclidean-space codes on ISI channels.**

I. Introduction

A Euclidean-space (ES) code that utilizes a group code designed over the group G is mapped to a signal set S of QAM modulation waveforms by a stationary mapping $\mu : G \rightarrow S$. The underlying group code is described by an encoder of finite-length generator sequences [1]. The QAM system used on a channel with ISI can be equivalently represented as a discrete-time (DT) ISI channel with AWGN [2]. The combination of an ES code with a DT ISI channel can be combined into a more complex composite ES code. A Viterbi decoder is the nearly-optimal ML decoder. Typically, there is a significant reduction in the free distance for the resulting composite ES code compared to the d^2_{free} for the memoryless channel. A technique known as TH-precoding has been used to regain some of the loss by performing the inverse of the DT ISI channel in the transmitter along with a modulo power constraint [3]. This technique requires that the transmitter has exact knowledge of the ISI channel through a feedback channel from the receiver.

II. Time-Varying Mappings

An alternative proposed method for coding on a k^{th}-order DT ISI channel is to use an ES code that has time-varying mappings $\mu_i : G \rightarrow R_i(S)$ where $\mu_i = R_i \circ \mu$. These codes will be called TVMES codes. For a specific channel and stationary ES code, there typically exists an ordered collection of mappings R_i that regains some of the loss in d^2_{free}. In many cases, the performance is better than that of the TH-precoding technique, but at the cost of an exponentially more complex Viterbi decoder which requires synchronization. The transmitter does not require exact knowledge of the ISI channel, so a more robust code can be designed over a range of possible channels. Implicit knowledge of the range is necessary to find the best combination of group code and mappings for the range.

III. Restrictions on the Time-Varying Mappings

Just as an exhaustive search is required to find the best ES code on the memoryless channel, the TVMES codes require an additional search over all ordered collections of mappings for each ES code. To make this search managable, restrictions are necessary for the type of mapping R_i that is permitted, and restrictions are necessary for the form of the ordered collection of mappings. This is a current area of research. The most severe restrictions are that the collection be of the form of incremental powers of a single unitary transformation $\mu_i = R^i\mu$. This will be called a rotating (or reflecting) ES code (RESC). RESC codes have shift-invariant distances on the k^{th}-order DT ISI channel, but more importantly, the problem of finding the best unitary transformation can be set up as an unconstrained optimization problem. A Newton-Raphson type algorithm can be used to solve for the pseudo-globally best unitary transformation for a given DT ISI channel and a given ES code. This severe restriction on the time-varying mappings actually includes many other types of collections because TVMES codes do not have a unique representation. Many good codes have been seen for small order DT ISI channels. Several specialized techniques for designing a code for the ISI channel can be generalized as TVMES codes.

References

[1] G.D. Forney Jr and M.D. Trott, "The dynamics of group codes: state spaces, trellis diagrams, and canonical encoders," *IEEE Trans. Inform. Theory*,vol. 39, pp. 1491–1513, Sept. 1993.

[2] S. Benedetto, E. Biglieri, V. Castellani, *Digital Transmission Theory*, N.J.: Prentice-Hall, 1987.

[3] M.V. Eyuboglu and G.D. Forney Jr, "Trellis precoding: combined coding, precoding, and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*,vol. 38, pp. 301–314, March. 1992.

On the Trellis of Convolutional Codes over Groups

Hans-Andrea Loeliger
ISY, Linköping University, S-58183 Linköping, Sweden

Abstract — **An algebraic characterization is given for the groups that can appear as the branch group ($\approx$ trellis section) of some convolutional code over a group (ring, field).**

A *convolutional code over a group* G is basically a shift-invariant subgroup of $G^{\mathbb{Z}}$ (subject, perhaps, to some further conditions such as controllability and observability, or completeness) [1] [2]. The *trellis* of such a code consists of identical sections, each of which is a triple (G, S, B), where S is the *state space* (or *state group*) and the *branches* B are a subgroup of $S \times G \times S$.

The standard "algorithm" to construct convolutional codes over a *field* goes as follows:

i. Choose a configuration of shift registers (cf. Fig. 1).

ii. Choose a linear mapping from the shift registers into F_2^n.

Note that the configuration of shift registers corresponds to the projection of the above group B onto $S \times S$.

The attempt to generalize this "algorithm" to codes over general groups leads to the problem of characterizing those groups that can appear as (the projection onto $S \times S$ of) B. Our main result is the following characterization of such groups.

Definition: A *shift structure* $(H_0, H_1, \ldots H_\ell; \varphi)$ for a group (module, vector space) H consists of a collection H_0, H_1, $\ldots H_\ell$ of normal subgroups (submodules, subspaces) of H (that need not be disjoint) together with an isomorphism φ from H/H_ℓ onto H/H_0 such that

i. $H_0 * H_1 * \ldots * H_\ell = H$;

ii. $(H_0 * H_1 * \ldots * H_j) \cap (H_j * H_{j+1} * \ldots * H_\ell) = H_j$ for $0 < j < \ell$;

iii. $\varphi(H_j * H_\ell) = H_{j+1} * H_0$ for $0 \leq j < \ell$.

Main Theorem: *Every strongly controllable, shift-invariant group code over any group (module, vector space) G can be found by the following "algorithm":*

i. Choose a group H with a shift structure $(H_0, H_1, \ldots H_\ell; \varphi)$.

ii. Choose a homomorphism $\omega : H \to G$.

iii. Construct the trellis (G, S, B) with states $S \triangleq H/H_0$ and branches

$$B \triangleq \{(h * H_0, \omega(h), \varphi(h * H_\ell)) : h \in H\}.$$

For Euclidean-space codes, G need not be specified a priori. In this case, step (ii) may be replaced by

ii. Choose a homomorphism ω from H into the isometry group of $\mathcal{R}^N$.

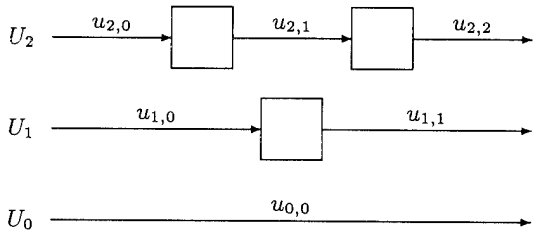


Figure 1: shift register.

The simplest example of a group (module, vector space) with a shift structure is a direct product

$$\mathcal{U} \triangleq U_0 \times U_1^2 \times U_\ell^{\ell+1}, \tag{1}$$

where U_0, U_1, $\ldots U_\ell$ are groups (modules, vector spaces) and where the terms U_i^{i+1} are themselves direct products. Such a group may be represented as a collection of delay lines as in Fig. 1, where the mapping φ may be interpreted as the shift operator. Note that the corresponding class of group codes includes all (strongly controllable) convolutional codes over any *field*.

If H is an arbitrary group with a shift structure, it can be shown that a (set-theoretic) one-to-one correspondence exists between H and a group of the type (1) such that φ corresponds to the shift operator in Fig. 1. In general, however, this one-to-one correspondence is not an algebraic isomorphism; in other words, the shift register of Fig. 1 is equipped with an algebraic structure other than the "natural" direct product.

So far, we have found just one class of groups with a non-standard (i.e., not the direct-product) shift structure: (multiplicative) groups of matrices with ones in the main diagonal and zeros above the main diagonal. By the main theorem, these groups give rise to a whole new class of noncommutative group codes. (Some such codes seem closely related to certain codes from [3].)

References

[1] G. D. Forney, Jr., and M. D. Trott, 'The dynamics of group codes: state spaces, trellis diagrams and canonical encoders', *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, Sept. 1993.

[2] H.-A. Loeliger, G. D. Forney, Jr., T. Mittelholzer, and M. D. Trott, 'Minimality and observability of group systems', *Linear Algebra & Appl.*, vol. 205–206, pp. 937–963, July 1994.

[3] E. J. Rossin, N. T. Sindhushayana, and C. Heegard, 'Trellis group codes for the Gaussian channel', submitted to *IEEE Trans. Inform. Theory*.

Minimality Tests for Rational Encoders over Rings

Thomas Mittelholzer

Signal and Information Processing Laboratory, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland

Given an encoding matrix over some field, various criteria are known to check minimality (cf. [1], [2], [3]). Most of these criteria apply to encoders of a particular class, e.g., basic encoders or systematic encoders, and only a few criteria are general in the sense that they apply to arbitrary rational encoding matrices. In this paper, causal rational encoders over commutative rings are considered and a general criterion of Johannesson and Wan [2] is generalized to rings, which satisfy the descending chain condition. Moreover, a new simple test is presented that reduces the minimality question from the ring to the field case. The basis for these new minimality tests are the concept of minimality of group systems and convolutional codes as presented in [4] and [5].

Let R be a commutative ring and let $R[D]$ denote the ring of polynomials over R. The *ring of rational functions* over R is defined by

$$R(D) = \left\{ \frac{f(D)}{D^m s(D)} \mid f(D), s(D) \in R[D], s(0) = 1 \text{ and } m \in \mathbb{Z} \right\}.$$

A $k \times n$-matrix $G(D)$ over $R(D)$ is called a *rational (n,k) encoding matrix over R* if it has k linearly independent rows over $R(D)$, or equivalently, if its kernel is zero. The matrix $G(D)$ is called *causal (or realizable)*, if all its components are causal rational functions, i.e., they have an expansion as formal power series in D. Every rational (n,k) encoding matrix $G(D)$ gives rise to an (n,k) convolutional code over R, which is defined by $C = \{\mathbf{u}(D)G(D) : \mathbf{u}(D) \in R(D)^k\}$.

To every convolutional code C, one can associate a canonical state space S_C that depends only on the code and not on a particular encoding matrix for C (cf.[4], [5]). A causal encoding matrix $G(D)$ is said to be *minimal*, if the abstract state space of $G(D)$ is isomorphic to the canonical state space S_C of the code C, which is generated by $G(D)$. In case of a finite alphabet, this definition is equivalent to the usual notion of minimality, which states that the encoder $G(D)$ requires the least number of states among all encoders that generate the code C.

Johannesson and Wan have presented the following general minimality criterion for the field case [2]. A causal encoding matrix $G(D)$ is minimal if and only if $G(D)$ has a polynomial right inverse in D and a polynomial right inverse in D^{-1}. This criterion cannot be generalized to arbitrary commutative rings because one can show that it does not hold over the ring of integers. However, there is a suitable class of rings to which the criterion can be extended, namely, the class of commutative rings satisfying the descending chain condition (DCC). The DCC is a rather weak restriction for practical purposes because most encoding alphabets are finite and every finite ring satisfies the DCC. There exists an important structure theorem for commutative rings satisfying the DCC, which can be viewed as an extension of the Chinese Remainder Theorem (cf. Chap. 7.10 of [6]). Such a ring decomposes into

$$R = R_1 \oplus R_2 \oplus \ldots \oplus R_s, \tag{1}$$

where the R_i's are local rings satisfying the DCC. In particular, it follows from (1) that R has only a finite number of maximal ideals $I_1, I_2, \ldots, I_s$.

Theorem 1 *Let R be a commutative ring satisfying the DCC and let the maximal ideals be denoted by $I_1, I_2, \ldots, I_s$. Let $G(D) \in R(D)^{k \times n}$ be a causal encoding matrix. Then the following statements are equivalent:*

(i) *$G(D)$ is minimal;*

(ii) *$G(D)$ has a polynomial right inverse in D and a polynomial right inverse in D^{-1};*

(iii) *for all $i = 1, \ldots, s$, the reduction of $G(D)$ modulo I_i is minimal over the field R/I_i.*

Condition (ii) of this theorem extends the Johannesson/Wan criterion to the ring case. Condition (iii) gives a new minimality test that reduces the question of minimality from the ring to the field case. It is illustrated by the following example.

Example 1 Consider the following encoding matrix over the ring of intergers modulo 4, $\mathbb{Z}_4$, given by

$$G(D) = \frac{1}{1+3D} \cdot \begin{bmatrix} 1+D & 1+2D+3D^2 \end{bmatrix}.$$

Reducing $G(D)$ modulo the only maximal ideal $(2) \subset \mathbb{Z}_4$, one obtains the binary encoding matrix

$$\bar{G}(D) = \begin{bmatrix} 1 & 1+D \end{bmatrix},$$

which is minimal over the binary field $\mathbb{Z}_4/(2)$. Hence, condition (iii) of the theorem holds and, therefore, $G(D)$ is minimal.

References

[1] G. D. Forney, Jr., 'Convolutional codes I: algebraic structure', *IEEE Trans. Inform. Theory*, vol. 16, pp. 720–738, Nov. 1970.

[2] R. Johannesson and Z. Wan, 'A linear algebra approach to minimal convolutional encoders', *IEEE Trans. Inform. Theory*, vol. 39, pp. 1219–1233, July 1993.

[3] G. D. Forney, Jr., 'Algebraic structure of convolutional codes, and algebraic system theory,' in *Mathematical System Theory*, A. C. Antoulas, Ed., pp. 527 –558. Springer 1991.

[4] H.-A. Loeliger, G. D. Forney, Jr., T. Mittelholzer, and M. D. Trott, 'Minimality and observability of group systems', *Linear Algebra & Appl.*, vol. 205–206, pp. 937–963, July 1994.

[5] T. Mittelholzer, "Minimal encoders for convolutional codes over rings," in *Communications Theory and Applications*, Eds. B. Honary, M. Darnell and P.G. Farrell, pp. 30 – 36, HW Comm. Ldt., 1993.

[6] N. Jacobson, *Basic Algebra II*, Freeman & Co., San Francisco 1974.

Permutation decoding of group codes

Ezio Biglieri[1]

Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy)
fax: +39 11 5644099 • e-mail: biglieri@polito.it •

Abstract — **We consider suboptimum decoding of group codes, represented in the form of a set of n-vectors whose components are obtained by permuting the components of an initial vector according to a certain group $\mathcal{G}$ of permutations. Permutation decoding consists of the following two steps. First, we decode the received vector by searching for the most likely permutation in the symmetric group $\mathcal{S}_n$, next we select the element in $\mathcal{G}$ closest to the permutation found. Here we focus on the first step. In particular, we show how any group code can be represented as a permutation code, and we determine the minimum value of n.**

I. Introduction

Consider, for motivation's sake, decoding of a binary (n, k) block code transmitted over the additive white Gaussian noise channel with the standard mapping m :GF(2)$\rightarrow \mathbf{R}$ defined by $0 \rightarrow -1$, $1 \rightarrow +1$. Soft decoding is performed by picking the code word closest to the received vector $\mathbf{r}$, while hard decoding can be viewed as an approximation of maximum-likelihood decoding performed in two steps. First, one uses preliminary decision regions formed by the orthants of $\mathbf{R}^n$, thus obtaining an element $\mathbf{y} \in m^{-1}\{\pm 1\}^n$. Next, algebraic decoding transforms the resulting n-tuple $\mathbf{y}$ into a code word. The whole procedure may be viewed as an approximation of the Voronoi regions of the code by a union of orthants of $\mathbf{R}^n$.

This procedure works because, while the determination of the one among the Voronoi regions in which $\mathbf{r}$ is falling is a complex task, we can make it easier by approximating them by a union of regions such that it is easy to determine in which one the received vector is falling. Here we apply this idea to group codes: their Voronoi regions are approximated by a union of smaller regions with the property that determining the position of $\mathbf{r}$ with respect to them is an easy task.

II. Group codes

Group codes are generated as follows. Consider a group $\mathbf{G}$ of $N \times N$ orthogonal matrices which forms a faithful representation of an abstract group $\mathcal{G}$ with M elements, and an "initial vector" $\mathbf{x} \in \mathbf{R}^N$, $\mathbf{R}^N$ the Euclidean N-dimensional space. A group code $\mathcal{X}$ is the orbit of $\mathbf{x}$ under $\mathcal{G}$, i.e., the set of vectors $\mathbf{Gx}$. By assuming that the only solution of the equation $G\mathbf{x} = \mathbf{x}$, $G \in \mathbf{G}$, is $G = I$ (the identity matrix), the code $\mathcal{X}$ has M elements. We may thus denote $\mathbf{x}_g$ the code vector associated with $g \in \mathcal{G}$.

With the vectors of $\mathcal{X}$ transmitted over the additive white Gaussian noise channel, the optimum (i.e., maximum-likelihood) decoder, upon reception of the noisy vector $\mathbf{r} = \mathbf{x}_g + \mathbf{n}$, chooses as the most likely transmitted vector the one that yields

$$\min_{g \in \mathcal{G}} \| \mathbf{r} - \mathbf{x}_g \|^2 . \tag{1}$$

If $\mathcal{G}$ is not endowed with any special structure, decoding (i.e., the solution of (1)) is obtained by exhaustive search among all the candidate $g \in \mathcal{G}$. This requires a number of calculations $\nu_C = NM$ (in fact, M scalar products of N terms each must be computed) and a storage of $\nu_S = NM$ real numbers (M vectors of N components each). In addition to this, the minimum has to be found, which requires ν_M operations.

III. Permutation decoding

We call Permutation Signal Set (PSS) a set of vectors that are obtained by applying a group $\mathcal{G}$ of permutations π to an initial vector $\mathbf{x}$. If the vectors have n components, application of the symmetric group $\mathcal{S}_n$ of all the permutations of n letters to an initial n-vector gives a class of codes known as "permutation modulation".

The latter codes admit an especially simple maximum-likelihood (ML) decoding algorithm. Assume that vector $\mathbf{r}$ was received. The ML decoder must seek the vector $\pi\mathbf{x}$ which maximizes the scalar product

$$\sum_{\ell=1}^{n} r_\ell (\pi\mathbf{x})_\ell .$$

This maximum is achieved when the largest component of $\pi\mathbf{x}$ is paired with the largest component of $\mathbf{r}$, the second largest component of $\pi\mathbf{x}$ is paired with the second largest component of $\mathbf{r}$, etc. This algorithm is algebraic in nature, and does not require the receiver to store all the code words.

Now, if the PSS is generated by a subgroup $\mathcal{G}$ of $\mathcal{S}_n$, we may use the same basic decoding idea in two steps:

1. We first decode $\mathbf{r}$ as if $\mathcal{G} = \mathcal{S}_n$, obtaining as a result a permutation π of n letters. This may not belong to $\mathcal{G}$.
2. Next we "algebraically decode" π into an element of $\mathcal{G}$.

Here we focus on the first decoding step. In particular, it can be proved that

- Every group code can be represented in the form of a permutation signal set acting on an initial vector $\mathbf{x}$ with n components.
- The minimum value of n is obtained as follows. If $|\mathcal{H}'|$ denotes the largest non-normal subgroup of $\mathcal{G}$ that does not include normal subgroups of $\mathcal{G}$ other than the identity, then n is given by the ratio

$$n = \frac{|\mathcal{G}|}{|\mathcal{H}'|}.$$

[1]This research was sponsored by the Italian National Research Council (CNR) under "Progetto Finalizzato Trasporti."

On the Algebraic Fundamentals of Convolutional Encoders Over Groups

Jorge Pedraza Arpasi, and Reginaldo Palazzo Jr.[1]

Dept. of Communications, State University of Campinas - UNICAMP; P.O.Box 6101, 13081-970 Campinas, SP, Brazil
email:arpasi@decom.fee.unicamp.br, email:palazzo@decom.fee.unicamp.br

Abstract — **Algebraic fundamentals of convolutional encoders are given by using the Schreier product and the Theory of Machines.**

I. Introduction

The majority of the convolutional encoders known in the technical literature are over algebraic fields. Recently, [1], and [2], have shown that these encoders essentially make use of the additive group of those fields. We take this approach and define the elementary convolutional encoder (ECE) over abelian groups and we point out their main properties that will serve as a reference to the definition of general machines. By use of the Schreier product, the general convolutional encoder (GCE) is defined. As a consequence, the ECE is a particular case of the GCE. The Schreier product can be properly exploited in the design of the encoder. As an example of this fact, we provide two results about the machine only by looking at the properties of this product.

II. Machines

Definition 1 *A* **machine** *is a quintuple $M = (X, Y, Q, \delta, \beta)$; where X is a finite set of inputs; Y is a finite set of outputs; Q is a (not necessarily finite) set(space) of states; $\delta : X \times Q \to Q$ is the next-state application; $\beta : X \times Q \to Y$ is the output application.*$\diamond$

Let x^* be a finite **string** of elements of X. We say that the machine $M = (X, Y, Q, \delta, \beta)$ is controllable if for all q and q' $\in Q$; there is a string x^* such that $q' = \delta^*(x^*, q)$. Where δ^* is the natural recursive extension map of δ. Given $j \in N$; if $\forall q,\ q' \in Q\ \exists x^* \in X^*$ with $|x^*| \leq j$ such that $q' = \delta^*(x^*, q)$; then we will say that the machine is j-controllable. Clearly, if M is j-controllable then it is $(j+1)$-controllable. The number $\nu = min\{j \mid M \text{ is } j - controllable\}$ is the **control index** of M.

III. Elementary Convolutional Encoder

Definition 2 *Let n, k, and m be natural numbers such that $n > k \geq 1$, and $m \geq 1$. Consider the matrices $T^0, T^1, ..., T^m$, with $T^i = \left(t^i_{rs}\right)$, where $t^i_{rs} \in \mathbf{Z}$, $1 \leq r \leq k$, $1 \leq s \leq n$, and $i = 0, 1, ..., m$. We define an* **elementary convolutional encoder** *with parameters (n, k, m) over a finite abelian group G as a machine $M \doteq (X, Y, Q, \delta, \beta)$ where:*
$X \subset G^k$ is the finite set of the input alphabet;
$Y \subset G^n$ is the set of the output alphabet;
$Q = \left\{q = (x^1, x^2, ..., x^m) \mid x^i \in X\right\} \subset \left(G^k\right)^m \approx G^{km}$, is the set (or space) of the machine states;
$\delta : X \times Q \to Q$, is given by $\delta(x^0, q) = (x^0, x^1, x^2, ..., x^{m-1})$ (the next state map);
$\beta : X \times Q \to Y$, is given by $\beta(x^0, q) = \beta(x^0, x^1, ..., x^m) = x^0T^0 + x^1T^1 + ... + x^mT^m$; (machine's outputs). $\diamond$

From this definition we can show the following properties of the ECE:

Proposition 1 *If X is a group, then:* **i)** *Q and $\beta(X, Q)$ are also groups.* **ii)** *The Cartesian product $X \times Q$ becomes a direct product of groups and the mappings δ and β are group homomorphisms, with δ being surjective.* **iii)** *The sets $Y_0 = \{\beta(x, e_Q)\}_{x \in X}$ and $Y_1 = \{\beta(x, q) \mid \delta(x, q) = e_Q\}$ are normal subgroups of $\beta(X, Q)$ and $\frac{\beta(X,Q)}{Y_0} \approx \frac{\beta(X,Q)}{Y_1} \approx Q$.* **iv)** *The ECE is a controllable machine, with control index $\nu \leq m$.*

IV. General Convolutional Encoder

Definition 3 *Let X and Q be two finite groups. Let $\sigma : Q \to Aut(X)$ and $\mu : Q \times Q \to X$ be mappings such that for any $q_1, q_2, q_3 \in Q$, and $x \in X$ both satisfying the following conditions:* **1)** $\sigma(q_1)(\mu(q_2, q_3)).\mu(q_1, q_2q_3) = \mu(q_1, q_2).\mu(q_1q_2, q_3)$ *and* **2)** $\sigma(q_1)(\sigma(q_2)(x)) = \mu(q_1, q_2).\sigma(q_1q_2)(x).\mu(q_1, q_2)^{-1}$. *Then, we define the* **Schreier product** *$X_{\sigma,\mu}Q$, of X and Q as the ordered pair of the elements of the respective groups (h, k) satisfying the following operation:*

$$(x, q) * (x', q') \doteq \left(x.\sigma(q)(x').\mu(q, q'), qq'\right).\diamond$$

This Schreier product is a group with identity element $(\mu(e_Q, e_Q)^{-1}, e_Q)$, where e_Q is the identity element of Q.

Definition 4 *A general convolutional encoder, with parameter ν, is a ν-controllable Schreier machine $M_{\sigma,\mu} = (X, Y, Q, \delta, \beta)$ such that the application $\Psi : X_{\sigma,\mu}Q \to Q \times Y \times Q$ given by $\Psi(x, q) \doteq (q, \beta(x, q), \delta(x, q))$ is injective.* $\diamond$

Assuming the set X is a group, and since the direct product is a particular case of the Schreier product, we have that the ECE is a particular case of GCE. Let $T = Im(\Psi)$ be the **edges** of the **trellis** of $M_{\sigma,\mu}$. T is a group isomorphic to $X_{\sigma,\mu}Q$. Moreover the sets $T_0 = \{\Psi(x, e_Q)\}_{x \in X}$ and $T_1 = \{\Psi(x, q) \mid \delta(x, q) = e_Q\}$ are normal subgroups of T and $\frac{T}{T_0} \approx \frac{T}{T_1} \approx Q$. On the other hand, $T_0 \approx X$. Hence, if $T_0 = T_1$; then, given $q \neq e_Q$, there is no x^* such that $\delta^*(x^*, e_Q) = q$. Thus, we have:

Theorem 1 *If the class $\chi = \{H \subset X_{\sigma,\mu}Q \ / \ H \text{ is a normal subgroup with } |H| = |X|\}$, has no more than one element, then the machine is non-controllable.*

[1]This work was supported in part by FAPESP under grant 92/4845-7, and it has been supported by CNPq under grant 301416/85-0, Brazil.

References

[1] M.D.Trott, *The Algebraic Structure of Trellis Codes*, Ph.D. Dissertation, Dept. of Elect. Eng., Stanford University, Stanford, CA, Aug. 1992.

[2] H.A.Loeliger, *On Euclidean-Space Group Codes*, Ph.D. Dissertation, Swiss Federal Institute of Technology, Zurich, 1992.

Useful Groups for Trellis Codes

James P. Sarvis and Mitchell D. Trott[1]
Dept. of Elec. Engineering & Computer Science, MIT, 77 Massachusetts Ave., Cambridge, MA 02139

Abstract — **We show how to construct and classify inequivalent homogeneous rate-$k/k+1$ trellis codes using principles of computational group theory. Given a complete classification of useful trellis structures, trellis codes based on groups are no more difficult to construct than trellis codes based on binary fields.**

I. Motivation

A homogeneous code $\mathcal{C}$ [1, 2] is the orbit $C\boldsymbol{x}$ of a group code C [3, 4] acting on a constant sequence $\boldsymbol{x}$. The class of homogeneous codes is larger than the class of binary linear convolutional codes, and may therefore be expected to contain new useful trellis codes. While linear codes, which are always homogeneous, can be found by enumerating parity check equations, codes constructed from non-abelian groups cannot. We are therefore forced to use more complex methods from group theory.

II. Methods

We can separate the problem of finding homogeneous codes into three parts: choosing a group structure for the trellis, assigning labels to trellis branches, and testing for pathological behavior. Like the enumeration methods used for convolutional codes, the partitioning and labeling of the signal set is essentially independent of the code search.

A suitable definition of equivalence for homogeneous codes greatly reduces the number of distinct structures that must be examined at each step. Two homogeneous codes are equivalent if there is a bi-infinite sequence of label permutations that maps one to the other. It can be shown that equivalent codes are always related by a constant sequence of permutations. Thus, code equivalence is simply trellis equivalence, where two labeled trellises are equivalent if there is a permutation of states and labels that takes one trellis to the other.

Group trellis structures are enumerated using derivative codes and group extensions. The size, rate, and controllability properties of the trellis are selected in advance; this fixes the locations of the trellis branches. A given trellis admits only one binary linear algebraic structure. But it may have several different group structures. Fortunately, despite the enormous number of nonisomorphic groups of even small order, only a handful appear as the algebraic structure of a trellis.

These groups are found by enumerating group extensions. If C is a group code, its *derivative code* C' is formed by taking the set of state sequences traversed by the sequences of C. Iterated derivatives terminate at the trivial code. If C has no parallel transitions then C and C' are isomorphic. Hence (unlabeled) trellises can be enumerated up to equivalence by enumerating derivatives up to isomorphism.

The derivative operation strips away any parallel transitions in the trellis of C. Reversing the derivative in such cases requires a group extension of the trellis by its parallel branch group. Group extensions of 2-ary groups by 2-ary groups, which are the only type that arise for rate-$k/k+1$ codes, are straightforward to enumerate for moderately sized groups.

[1]This work was supported by NSF Grant NCR-9457509

Given an unlabeled group trellis, the next step is to assign labels to branches. It suffices to assign only the zero-labeled branches in the trellis because, for a homogeneous code, the zero-labeled branches form a subgroup of the trellis, and each right coset of this subgroup is distinctly labeled.

Zero labeling proceeds as follows. The states which have exiting zero branches are a subgroup of the state group, as are the states with entering zero branches. In fact, these two subgroups must be isomorphic and, for rate-$k/k+1$ codes, must be half the size of the state group. Choosing the left and right zero-labeled state groups therefore amounts to enumerating a restricted class of subgroups of index 2. The zero-labeled branches define an isomorphism between the left and right zero-labeled state groups; assigning zero branches is tantamount to enumerating isomorphisms from one subgroup to another. Recent advances in computational group theory have solved this problem for 2-ary groups.

The last step in the construction of useful trellis group structures is to test the trellis for catastrophic behavior. For group codes, this test is performed by checking if the zero-labeled branch group admits a periodic path through the trellis. Interestingly, this final test eliminates many nonabelian state groups for which no noncatastrophic labeling exists.

The final step of mapping branch labels to elements of a partitioned signal set can proceed as with the standard binary linear case.

III. Results

The methodology developed above reduces the problem of enumerating useful groups to an essentially mechanical process. Preliminary results for small codes are tabulated below. The results for binary linear codes were computed primarily for verification; they can also be found by counting parity check equations. Note that nonabelian codes become more plentiful beyond 16 states.

states	rate	state group	number
4	1/2	$\boldsymbol{Z}_2 \times \boldsymbol{Z}_2$	4
8	1/2	$\boldsymbol{Z}_2 \times \boldsymbol{Z}_2 \times \boldsymbol{Z}_2$	16
	2/3	$\boldsymbol{Z}_2 \times \boldsymbol{Z}_2 \times \boldsymbol{Z}_2$	12
	2/3	D_8	1
16	1/2	$(\boldsymbol{Z}_2)^4$	64
	2/3	$(\boldsymbol{Z}_2)^4$	48

References

[1] M. D. Trott and J. P. Sarvis, "Homogeneous Trellis Codes," in *Proceedings of the 32nd Annual Allerton Conference on Communication, Control, and Computing*, pp. 210–219, September 1994.

[2] N. T. Sindhushayana, B. Marcus, and M. D. Trott, "Homogeneous Shifts," submitted to *IMA Journal on Mathematical Control and Information*, March 1995.

[3] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, Sept. 1993.

[4] H.-A. Loeliger and T. Mittelholzer, "Convolutional codes over groups," *IEEE Trans. Inform. Theory*, to appear.

Codes from Iterated Maps

Håkan Andersson and Hans-Andrea Loeliger
ISY, Linköping University, S-58183 Linköping, Sweden

Keywords: **symbolic dynamics, chaos, group codes.**

We consider codes of the following type. Let S (the signal set) be a subset of n-dimensional Euclidean space $\mathcal{R}^n$. Let $f : S \to S$ be a continuous mapping. The code $C(S, f)$ consists of those bi-infinite sequences $x = \ldots x_{-1}, x_0, x_1, x_2, \ldots \in S^{\mathcal{Z}}$ that satisfy

$$x_t = f(x_{t-1})$$

for all $t \in \mathcal{Z}$. Note that the "future" of each codeword is completely determined by its "past."

At first sight, it might seem that the information rate (i.e., the number of information bits per code symbol) of any such code must be zero. However, as the example below shows, this need not be so if S is an infinite set.

Throughout the paper, A will denote some finite alphabet and B will denote a subshift (i.e., a closed, shift-invariant subset) of $A^{\mathcal{Z}}$. Let $\sigma : A^{\mathcal{Z}} \to A^{\mathcal{Z}}$ be the left shift operator.

Example: Let $A \triangleq \{0, 1\}$, let B be any subshift of $A^{\mathcal{Z}}$, and let $\rho : B \to [0, 1]$ be the mapping

$$\ldots b_{-1}, b_0, b_1, b_2, \ldots \mapsto \sum_{t=0}^{\infty} b_t / 2^{t+1}.$$

We then define a code C as the image of the encoding rule $B \to C : b \mapsto x$ with

$$x_t = e^{i2\pi\rho(\sigma^t(b))}.$$

Clearly, the information rate of C is one bit per symbol. The signal set S is some subset of the unit circle. But

$$\begin{aligned} x_t &= e^{i2\pi \cdot 2\rho(\sigma^{t-1}(b))} \\ &= x_{t-1}^2, \end{aligned}$$

which shows that $C = C(S, f)$ for $f : x \mapsto x^2$.

For $B = A^{\mathcal{Z}}$, this example was first presented in [1], where it was shown that the code is a group code (or geometrically uniform [2]) and has a well defined minimum distance. It then turned out that this code is actually a standard example of a chaotic dynamical system [3]. The related idea of using chaotic systems to produce waveforms for communications had earlier been proposed in [4].

The choice of $B = A^{\mathcal{Z}}$ in the example causes the following problem: the all-ones information sequence and the all-zeros information sequence are mapped to the same codeword. (One can prove that some problem of this type always occurs if S is connected.) The remedy is to restrict B to a subshift of $A^{\mathcal{Z}}$ that forbids too many consecutive zeros (or ones, or both zeros and ones). The resulting effective signal set S is a fractal and totally disconnected (like the Cantor set). While this seems odd at first sight, the resulting codes are well-behaved in every respect; in particular, they can be encoded and decoded with finite memory and finite-precision arithmetic.

It can also be shown that codes of this type can have an arbitrarily large minimum distance, which dispells any lingering suspicion that such codes are somehow inherently "bad."

References

[1] H.-A. Loeliger, 'Abelian-group convolutional codes for PSK need not be ring codes', Proc. 6th Joint Swedish-Russian Int. Workshop on Information Theory, Mölle, Sweden, Aug. 22–27, 1993, pp. 21–22.

[2] G. D. Forney, Jr., 'Geometrically uniform codes', *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241–1260, Sept. 1991.

[3] R. L. Devaney, *Chaotic Dynamical Systems,* Addison-Wesley, 1989.

[4] S. Hayes, C. Grebogi, and E. Ott, 'Communicating with chaos', *Phys. Rev. Let.*, vol. 70, no. 20, 17 May 1993, pp. 3031–3034.

On Binary-to-q-ary Codes over Groups

Leif Wilhelmsson
Department of Telecommunication Theory, Lunds University, Box 118, S-221 00 Lund, Sweden

Abstract — **A class of binary-to-q-ary convolutional codes is studied where the operation performed in the encoder is addition modulo q. For rate 1 codes an extended spectrum for the codes is defined and a necessary and sufficient condition for the encoder to be catastrophic is given. Optimal codes, for relatively small alphabet size q and memory size m, found by computer search are reported.**

I. Summary

Consider a rate 1 memory m binary-to-q-ary encoder where the operations in the encoder are performed over Z_q (the ring of integers mod q). The encoder input sequence is binary and the encoder output sequence as well as the encoder generator coefficients are q-ary, i.e.

$$v(D) = u(D)g(D) \bmod q,$$

where

$$\begin{aligned} v(D) &= v_0 + v_1 D + v_2 D^2 + \ldots & v_i &\in \{0, 1, \ldots, q-1\} \\ u(D) &= u_0 + u_1 D + u_2 D^2 + \ldots & u_i &\in \{0, 1\} \\ g(D) &= g_0 + g_1 D + \ldots + g_m D^m & g_i &\in \{0, 1, \ldots, q-1\}. \end{aligned}$$

If the input bits are viewed as indicator functions, the only operation performed in the encoder is addition modulo q. In comparison to the encoders reported in [1], the choice of output alphabet is therefore less restrictive. Since the input alphabet is not a subfield of the output alphabet a different approach must be taken regarding free distance and distance spectrum for the code. Furthermore, it is possible that a rate 1 binary-to-q-ary encoder is catastrophic although no input sequence of infinite Hamming weight results in an output sequence with only a finite number of nonzero symbols. An example is the rate 1 binary-to-6-ary encoder $g(D) = 2 + 2D + 4D^2$, shown in Fig. 1.

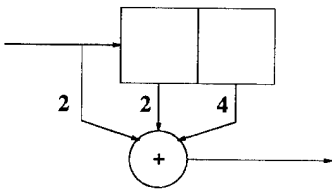


Figure 1: Rate 1 binary-to-6-ary encoder.

For this encoder it is easily seen that no input sequence with infinite Hamming weight gives an output sequence with finite Hamming weight. However, the encoder is catastrophic since the two (infinite) input sequences $u = 10010111\ \ 1001\ldots$ and $u' = 01111001\ \ 0111\ldots$ result in output sequences that only differ in the first position. Moreover, the distance spectrum for a binary-to-q-ary code can not be defined in an appropriate way, because different output sequences may have different distance spectra.

In order to "circumvent" these difficulties we observe that the difference between two input sequences, u and u', is a vector with elements $\in \{-1, 0, 1\}$. The properties of an encoder are for this reason evaluated by use of an extended input alphabet with elements from $\{-1, 0, 1\}$, i.e the input sequence is $u_{\text{ext}}(D) = u_0 + u_1 D + u_2 D^2 + \ldots$, where $u_i \in \{-1, 0, 1\}$. The corresponding weight distribution of the output sequence is then evaluated. For the encoder above we find that it is catastrophic since the (infinite) input sequence $1\ -1\ -1\ 0\ -1\ 1\ 1\ 0\ 1\ -1\ -1\ 0\ldots$ results in the output sequence $2\ 0\ 0\ 0\ \ldots$. The conditions for a rate 1 binary-to-q-ary encoder $g(D)$ to be catastrophic can be summerized in

Theorem 1 *A rate 1 binary-to-q-ary encoder is catastrophic if and only if there exist an integer N and a sequence $u_N(D)$ such that $(1 - D^N) | u_N(D)g(D) \bmod q$, where $u_N(D) = u_0 + u_1 D + \ldots + u_{N-1} D^{N-1}$, $u_n \in \{-1, 0, 1\}$.*

Applying the theorem to the encoder in Fig 1, we find that it is catastrophic since for $N = 8$ and $u_N(D) = 1 - D - D^2 - D^4 + D^5 + D^6$ we have $u_N(D)g(D) = 2 + 4D^8 = 2(1 - D^8) \bmod q$.

To find an optimum code an "extended distance spectrum" corresponding to input symbols from the extended input alphabet was calculated according to the idea described in [2]. If we let $n(d_{free} + i)$ denote the $(i+1)$th spectral component, then the codes found by computer search are optimal in the sense that the free distance is maximal, i.e. $d_{free} = m + 1$, and no code exists such that, for any $l = 0, 1, 2, \ldots$,

$$\begin{aligned} n(d_{free} + i) &= n_{opt}(d_{free} + i) & i &= 0, 1, \ldots l-1 \\ n(d_{free} + i) &< n_{opt}(d_{free} + i) & i &= l. \end{aligned}$$

In Table 1 the first three spectral components for the found optimal codes, corresponding to distances $m+1$, $m+2$ and $m+3$, are given.

	memorysize			
	m = 1	m = 3	m = 5	m = 7
q = 4	(2,4,8)	(6,20,92)	—	—
q = 5	(2,4,8)	(4,16,100)	(8,70,364)	—
q = 6	(2,4,8)	(2,12,62)	(4,16,126)	(10,42,224)
q = 7	(2,4,8)	(2,8,32)	(2,24,64)	(4,48,184)
q = 8	(2,4,8)	(2,4,16)	(2,6,42)	(2,14,70)

Table 1: Best distance spectrum for some values of memory size m and alphabet size q. That no code with $d_{\text{free}} = m + 1$ was found is indicated by "—".

References

[1] William E. Ryan and Stephen G. Wilson, "Two classes of convolutional codes over GF(q) for q-ary orthogonal signaling," *IEEE Trans. Commun.* vol. COM-39, pp. 30-40, Jan. 1991.

[2] Mats Cedervall and Rolf Johannesson, "A fast algorithm for computing distance spectrum of convolutional codes," *IEEE Trans. Inform. Theory.* vol. IT-35, pp. 1146-1159, Nov 1989.

Abelian group codes, duality and MacWilliams identities

Thomas Ericson[1] and Victor Zinoviev[2]

[1]Dept. of Electrical Engineering, Linköping University,
S-581 83 Linköping, Sweden

[2]Institute for Problems of Inf. Transmission
Ermolova Str.19
GSP-4, Moscow, 101447, Russia

Abstract **– The concept of dual codes is formulated in terms of characters and abelian groups. The MacWilliams transform is established under general conditions. It is demonstrated that this transform can naturally be regarded as a partitioning of a fourier transform.**

Let A be a finite abelian group and let $U \stackrel{\Delta}{=} \{ z \in C : |z|=1 \}$ be the set of units in the complex plane C. Any homomorphism $\varphi: A \to U$ is an irreducible character of A. The set $\hat{A}$ of all irreducible characters is an abelian group isomorphic to A (Herstein [1], p 115). Let $\Psi : A \to \hat{A}$ be a fixed isomorphism from A to $\hat{A}$, taking the element a in A to the irreducible character Ψ_a in $\hat{A}$.

By a code we understand a subset C in the group A. The code is a **group code** if it is a subgroup in A. For any group code C in A we define the **dual** $C^{\perp}$ according to

$$C^{\perp} \stackrel{\Delta}{=} \{ a \in A : \Psi_a(C) = \{1\} \}.$$

It is easy to see that $C^{\perp}$ is also a group.

Now suppose there is a weight w(x) associated with each element $x \in A$. More formally, let $w: A \to \Re^+$ be a map from the finite group A into the set $\Re^+$ of non-negative real numbers. Denote by W the range of this map and let C be a code in A. For any $u \in W$ we define

$$A_u = |\{ x \in C : w(x) = u \}|.$$

The **weight distribution** of the code C is $\Delta \stackrel{\Delta}{=} \{ (u, A_u) : u \in W \}$.

A bijection $T: A \to A$ such that $w(Tx) = w(x)$ holds for any element x in A is called a **weight - preserving transformation**. If T and S are two weightpreserving transformations, define the product TS according to $TS(x) = T(S(x))$, $x \in A$. It is clear that the set of all weight preserving transformations forms a group under this product. We denote this group by Ω.

Lemma: Let the characters Ψ_x satisfy $\Psi_{Tx}(Ty) = \Psi_x(y)$; $x, y \in A$; $T \in \Omega$. Under this assumption there is a function $K : A \times A \to \Re^+$ such that

$$J_u(y) = \sum_{x \in A} j_u(x)\, \Psi_x(y) = K(u, w(y)), \qquad x, y \in A,$$

where $j_u(x)$ is the indicator function for the weight w and where $J_u(y)$ is its fourier transform. Morever, under these conditions the weight distribution $A_u^{\perp}$ for the dual code $C^{\perp}$ is given by

$$A_u^{\perp} = \sum_{v \in W} K(u,v)\, A_v .$$

The last relation is the MacWilliams identity. We address the question under what conditions this holds. One general result is as follows.

Denote by $\Re^A$ the set of all functions $f : A \to \Re$ and let L(w) denote the linear subspace in $\Re^A$ spanned by the functions $\{ j_u : u \in A \}$. It is clear that this set forms an orthogonal basis in L(w). The set $\{J_u : u \in A\}$ is of course also an orthogonal basis. We denote by $L^{\perp}(w)$ the linear space spanned by this new basis. In general the spaces L(w) and $L^{\perp}(w)$ are different. Occasionally, however, they might coincide.

Theorem: the MacWilliams identity holds if and only if the weight w is such that $L(w) = L^{\perp}(w)$.

REFERENCES

[1] J.N. Herstein, "Topics in algebra", *Blaisdell Publ. Company*, New York - Toronto - London, 1964.

[2] F.J. MacWilliams, "A theorem on the distribution of weights in a systematic coide", *Bell Syst. Techn, J*, Vol. 42, pp. 79-94, 1963.

[3] T. Ericson and D. Zongduo, "MacWilliam's identity using tensor products", *Proc. of he 13th symposium on information theory and its applications*, SITA '90, Tatesina, Japan, 1991.

[4] B. Sundar Rajan and M.U. Siddiqi, "A generalized DFT for Abelian codes over Z_m", *IEEE Trans. Inform. Theory*, Vol. IT-40, pp. 2082-2090, 1994.

Detection of Spread-Spectrum Signals for Linear Multi-User Receivers

Urbashi Mitra* & H. Vincent Poor**

* Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210

** Department of Electrical Engineering, Princeton University, Princeton, NJ 08544

Abstract– **This paper investigates the performance (differential SNR) of two detectors for spread-spectrum signals modeled as random processes embedded in channel noise. Linear interference suppression is performed on the multiple-access interference prior to detection; thus the noise in the detection problem is comprised of colored noise and residual multiple-access interference. It is observed that a non-linear detector outperforms a purely linear detector.**

1. Introduction

A Code-Division Multiple-Access (CDMA) based digital communication system is considered. Bandwidth efficiency, complexity and security issues motivate the search for schemes to integrate new user information into centralized demodulators. In order to accommodate a new user into a receiver, its presence must be detected.

It is straightforward to show that the locally optimum detector for this detection problem optimizes the differential SNR. However the locally optimal detector is infeasible to implement; thus, simpler, noise-distribution-independent detectors are pursued. A non-linear detector is considered to compensate for the presence of the residual multiple-access interference (RMAI) which is non-Gaussian in nature. Comparison is made with a linear detector which is better suited to Gaussian noise.

2. The Detection Problem

The signal to be used for detection of the spread-spectrum signal is a linear transformation of the received signal. The transformation, V, is chosen to suppress multiple-access interference. This gives rise to a hypothesis testing problem that can be cast as follows:

$$\begin{aligned} H_0 : \underline{x}_i &= V\underline{n}_i + \underline{\delta}_i \\ H_1 : \underline{x}_i &= V\underline{n}_i + \underline{\delta}_i + \theta\underline{s}_i \quad \textit{for } i \in [1, N], \end{aligned}$$

θ is an SNR parameter. $V\underline{n}_i$ is the ambient channel noise which will be modeled as a zero-mean, colored, additive Gaussian process. $\underline{\delta}_i$ is the RMAI. For K existing users, $\underline{\delta}_i$ is drawn from one of 2^K possible random vectors with equal probability. We assume that the stochastic signal $\underline{s}_i$ is zero mean.

We examine detectors based on real-valued detection statistics, $T_N(\underline{x})$, compared to thresholds. The differential SNR for the random signal case is defined as ,

$$\xi(T) \equiv \lim_{N\to\infty} \frac{1}{N} \frac{\left[\lim_{\theta^2\to 0} \frac{\partial^2}{\partial\theta^2} \mathbf{E}_\theta\{T_N(\underline{x})\}\right]^2}{\mathbf{Var_0}\{T_N(\underline{x})\}}.$$

The two detectors under study have the following form:

$$T_N(\underline{x}) = \sum_{i=1}^{\frac{N}{2}} \sum_{k=1}^{L} \Phi(x_{2i-1\,k})\, \Phi(x_{2i\,k}),$$

where $\Phi(x) = x$ for the *simple correlator* (T_{SCO}) and $\Phi(x) = \text{sgn}(x)$ for the non-linear *polarity coincidence correlator* (T_{PCC}). It can be shown that the decision statistics for these two correlators under both hypotheses are asymptotically normal and hence justify the use of the differential SNR as a performance measure.

While one can easily determine the differential SNR for T_{SCO}, calculation of the differential SNR for T_{PCC} involves the evaluation of the following probability: $P[x > 0, y > 0]$ for x and y two jointly Gaussian, correlated random variables with *non-zero means*. No closed form expression exists for this quantity [1]; thus we bound this probability to yield the following,

$$0 \le |\mathbf{E}\{\text{sgn}(x)\text{sgn}(y)\}| \le \frac{2}{\pi} \arcsin\left(\frac{\rho}{\sigma_x \sigma_y}\right),$$

where $\rho = \mathbf{Cov}\{x, y\}$. Use of these bounds yield upper and lower bounds for the differential SNR of T_{PCC} .

3. Performance Example

Performance is studied in the context of a decorrelator [2] based multi-user receiver system. It is assumed that the spreading codes are mismatched between the mobile transmitters and the receiver (e.g. due to multi-path) thus RMAI will be present. We consider an environment where the absolute value of the cross-correlation between the signature sequences is increased by $0 < \zeta < 1$, and the auto-correlation is decreased by ζ; ζ captures the worst-case mismatch due to propagation effects. It is clear from Figure 1 that the non-linear detector maintains a distinct advantage over the linear one. The performance of a multi-user system based on conventional matched filter receivers is also studied, but not presented here.

References

[1] S. S. Gupta, "Probability integrals of multivariate normal and multivariate t," *Annals of Mathematical Statistics*, vol. 34, pp. 792–838, 1963.

[2] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 123–136, January 1989.

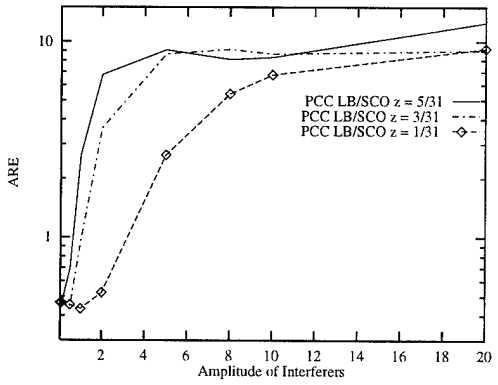


Figure 1: Asymptotic Relative Efficiencies between T_{PCC} (lower bound) and T_{SCO} for 10 users (length 31 codes).

*This research was supported by the U. S. Army Research Office under Grant DAAH04-93-G-0219.

MMSE Interference Suppression for Joint Acquisition and Demodulation in CDMA systems

Upamanyu Madhow[1]

Coordinated Science Laboratory, University of Illinois, 1308 West Main, Urbana, IL 61801, USA

I. Introduction

Minimum Mean Squared Error (MMSE) demodulation for direct-sequence CDMA systems [1] eliminates the near-far problem, and can be implemented adaptively (i.e., without explicit knowledge of the parameters of the multiple-access interference), given a training sequence for the desired transmission. However, prior to *timing acquisition*, the receiver does not know the *phase* of the training sequence, i.e., it does not know, for a given observation interval, which bit of the training sequence contributes the most signal energy. Conceivably, this timing information could be obtained using conventional acquisition techniques by correlating over long enough intervals and applying enough power control to resolve the near-far problem. In this paper, however, we present an adaptive approach to the problem of near-far resistant *joint* acquisition and demodulation.

Our method is to use a training sequence with a short period P, and run P adaptive algorithms either serially or in parallel, one for each assumed phase of the training sequence. The adaptive algorithm that yields the least Mean Squared Error (MSE) corresponds to the correct phase, and yields in addition an MMSE correlator that can be used for continued training or for decision-directed adaptation. Thus, acquisition results in a near-far resistant *demodulator* that implicitly accounts for the timings and amplitudes of *all* the transmissions without explicitly estimating even the timing of the desired signal. (Estimates of the latter can be derived from the resulting MMSE demodulator if required.) We note that a method for joint acquisition and demodulation that does not require a training sequence has also been devised [2].

II. Model and Algorithm

We consider an *equivalent synchronous model* for an asynchronous direct-sequence CDMA system, obtained by chip matched filtering, sampling at (a multiple of) the chip rate, and restricting attention to a finite observation interval for each bit decision. The received vector $\mathbf{r}_n \in \mathcal{R}^d$ used for the nth bit decision is given by

$$\mathbf{r}_n = b_0[n]\mathbf{u}_0 + \sum_{j=1}^{J} b_j[n]\mathbf{u}_j + \mathbf{w}_n \qquad (1)$$

where $\mathbf{u}_0$ is the vector modulating the desired bit $b_0[n]$, and, for $1 \leq j \leq J$, $b_j[n]$ are interfering bits due to intersymbol interference and multiple-access interference, $\mathbf{u}_j$ are interference vectors modulating these bits, and $\mathbf{w}_n$ is additive white Gaussian noise. The received vector for subsequent bits are obtained by sliding the observation interval by T, where T is the bit interval. The vectors $\mathbf{u}_j$ are linear combinations of shifts of the spreading sequences used by the various transmissions; we do *not* assume knowledge of these vectors. Our objective is to arrive at a linear receiver that provides a bit estimate $\hat{b}_0[n] = \mathrm{sgn}(\mathbf{c}^T\mathbf{r}_n)$, where $\mathbf{c}$ is chosen to minimize the MSE $E[(\mathbf{c}^T\mathbf{r}_n - b_0[n])^2]$.

The desired transmission sends a periodic sequence (period P) of training bits $t[n]$. We consider an observation interval of at least $2T$, so that one bit of the desired transmission must fall completely within it. Letting $b_0[n]$ denote this bit, we must have $b_0[n] = t[n+k^*]$ for some unknown integer k^* between 0 and $P-1$. Since the phase k^* of the training sequence is *not* known while in acquisition mode, we run P adaptive MMSE demodulators, each corresponding to one of the following hypotheses about the phase of the training sequence:

$$H_i: \quad b_0[n] = t[n+i], \quad i = 0, 1, ..., P-1 \qquad (2)$$

For example, under a least squares implementation of this algorithm spanning M observation intervals, the correlator for the ith hypothesis is given by

$$\tilde{\mathbf{c}}_i = \hat{\mathbf{R}}^{-1}\tilde{\mathbf{u}}^{(i)} \qquad (3)$$

where $\hat{\mathbf{R}} = (1/M)\sum_{n=1}^{M} \mathbf{r}_n\mathbf{r}_n^T$ is the empirical crosscorrelation matrix for the received vector, and

$$\tilde{\mathbf{u}}^{(i)} = (1/M)\sum_{n=1}^{M} t[n+i]\mathbf{r}_n$$

is the estimate of the desired signal vector $\mathbf{u}_0$ under hypothesis H_i. The estimated MSE under hypothesis H_i is given by $\eta_i = 1 - \tilde{\mathbf{c}}_i^T\tilde{\mathbf{u}}^i$. The best hypothesis is the one with the smallest estimated MSE, and the corresponding correlator $\tilde{\mathbf{c}}_i$ is a near-far resistant demodulator by virtue of the near-far resistance of the MMSE demodulator [1]. Good hypotheses can be combined to further enhance performance. This method relies on the training sequence having good periodic autocorrelation, and on the data bits for the interfering transmissions being uncorrelated with those of the desired transmission. If multiple transmissions are being simultaneously acquired, their training sequences should have good periodic crosscorrelations.

In the conference presentation, we will (a) show via simulation of the least squares algorithm that a near-far resistant demodulator is obtained after a very small number of iterations, (b) provide an approximate analysis of the effect of least squares estimation errors on acquistion performance (i.e., on the probability of choosing the wrong hypothesis), and (c) comment on directions for future research.

[1] This work was supported in part by funds from the University of Illinois Research Board.

References

[1] U. Madhow, M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3178-3188, December 1994.

[2] U. Madhow, "Blind adaptive interference suppression for acquisition and demodulation of direct-sequence CDMA signals," *Proc. Conf. Inf. Sci. Sys. (CISS '95)*, Johns Hopkins University, Baltimore, MD, March 1995.

ORTHOGONALLY ANCHORED BLIND INTERFERENCE SUPPRESSION USING THE SATO COST CRITERION

Michael L. Honig
Department of EECS, Northwestern University
Evanston, IL 60208

We present a blind adaptive interference suppression algorithm for Direct-Sequence Code-Division Multiple-Access, which is based on the Minimum Mean Squared Error (MMSE) criterion. The algorithm is blind in the sense that it does not require a training sequence, although it does require (approximate) knowledge of the user spreading waveform and associated timing. The algorithm is related to the blind interference suppression algorithm presented in [1], and assumes that the MMSE filter is expressed as the sum of two orthogonal components: the matched filter (referred to as the *anchor*) and an adaptive filter. However, instead of using the minimum variance (MV) criterion, as in [1], we consider an alternative cost function which is closer to the actual MSE. This cost criterion was proposed by Sato and Godard [2] for blind equalization of a single-user channel. However, without the orthogonal decomposition presented in [1], this cost function is not suitable for the multi-user application due to the presence of a local minimum associated with each user.

The orthogonally anchored Sato cost function leads to a stochastic gradient (or least squares) algorithm that has the following advantages relative to the MV algorithms in [1]:

- The algorithm is insensitive to mismatch between the anchor and desired signal.
- Multipath components within the window spanned by the filter are coherently combined.
- The stochastic gradient algorithm produces (much) less asymptotic MSE than the MV stochastic gradient algorithm for the same speed of convergence.

A disadvantage associated with this cost function is that there is a local minimum associated with each user. However, if the crosscorrelation between any pair of pulse shapes is small, then the orthogonal anchor ensures that the norm of the coefficient vector that achieves any of these local minima must be very large. These local minima can therefore be excluded by an appropriate norm constraint on the vector of filter coefficients.

Orthogonally Anchored Adaptive Algorithm

Consider a synchronous DS-CDMA system where the vector of received samples corresponding to the ith transmitted bit at the output of the chip matched filter is given by

$$\mathbf{r}[i] = \sum_{k=1}^{K} b_k[i]A_k\mathbf{s}_k + \mathbf{n}[i] \qquad (1)$$

where K is the number of users, $\mathbf{r}$ has N components, N being the processing gain, $\{b_k[i]\}$ is the sequence of binary symbols corresponding to user k, $\mathbf{s}_k$ is the spreading code for user k where $\|\mathbf{s}_k\| = 1$, A_k is the amplitude for user k, and $\mathbf{n}$ is a noise vector.

The linear MMSE detector for user 1 consists of the coefficient vector $\mathbf{c}_1$ that minimizes $E[(b_1[i] - \mathbf{c}_1'\mathbf{r}[i])^2]$. To obtain the blind algorithm $\mathbf{c}_1$ is constrained to be of the form $\mathbf{c}_1' = \hat{\mathbf{s}}_1 + \mathbf{w}_1$ where $\hat{\mathbf{s}}_1$ is an estimate of $\mathbf{s}_1$, and $(\mathbf{w}_1)'\hat{\mathbf{s}}_1 = 0$. $\mathbf{c}_1$ is then chosen to minimize the Sato cost function

$$F(\mathbf{c}_1) = E\left\{\left[\mathbf{c}_1'\mathbf{r}[i] - \text{sgn}\,(\mathbf{c}_1'\mathbf{r}[i])\right]^2\right\} \qquad (2)$$

where $\text{sgn}\,(x) = x/|x|$.

A stochastic gradient algorithm that minimizes (2), subject to the preceding orthogonal decomposition, is given by

$$\mathbf{w}[i] = \mathbf{w}[i-1] - \mu e[i]\left(\mathbf{r}[i] - (\mathbf{r}'[i]\hat{\mathbf{s}}_1)\hat{\mathbf{s}}_1\right) \qquad (3)$$

where $e[i] = \mathbf{c}_1'\mathbf{r}[i] - \text{sgn}\,(\mathbf{c}_1'\mathbf{r}[i])$, and μ is the step-size. (The MV stochastic gradient algorithm presented in [1] simply replaces $e[i]$ by the output sample $\mathbf{c}_1'\mathbf{r}[i]$.) A least squares adaptive algorithm based on the preceding approach is easily derived. Numerical examples comparing the performance of these algorithms with the MV algorithm in [1], and with the conventional LMS algorithm will be presented at the conference.

References

[1] M. L. Honig, U. Madhow, and S. Verdu, "Blind Adaptive Interference Suppression for Near-Far Resistant CDMA", *IEEE Trans. on IT*, July 1995.

[2] D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems", *IEEE Trans. on Comm.*, Vol. COM-28, No. 11, pp. 1867-1875, Nov. 1980.

OPTIMAL SOFT MULTI-USER DECODING FOR VECTOR QUANTIZATION IN A SYNCHRONOUS CDMA SYSTEM

Mikael Skoglund and Tony Ottosson

Department of Information Theory, Chalmers University of Technology, S - 412 96 Göteborg, Sweden

I. INTRODUCTION

During the last few years there has been much work on *multi-user detection* (MUD) for Direct Sequence Code Division Multiple Access (DS/CDMA) systems, and several solutions have been presented [1]. Another active field of research considers methods for *combined source and channel coding* in vector quantization (VQ) [2]. The present paper combines these two areas. We present a method for robust transmission of VQ-data over a CDMA channel. Our approach differs from most prior work in two ways: (1) The decorrelation of the users and the decoding of the VQs are carried out simultaneously; (2) The decoding is based on the unquantized matched filter outputs, and no binary decisions are taken. We use the term *soft decoding* to emphasize this latter fact. Thus, our approach considers estimation based rather than detection based decoding of the channel and the VQs. Similar studies for single-user channels can be found in [3], [4], and [5].

II. SYSTEM MODEL

Consider a symbol synchronous CDMA system with K users. User k produces a sample vector $\mathbf{X}_k$, which is encoded into an index I_k by the VQ encoder of user k. The index is thereafter converted into a block $\mathbf{b}(I_k)$ of L bits in polar format $\{\pm 1\}$. For simplicity we assume that all users have the same block length L. The bits are transmitted one by one on a CDMA channel that is distorted by AWGN. Thus, the matched filter outputs of the received signal at time n can be expressed as, (c.f. [1]), $\mathbf{Y}_n = \mathbf{RW}\cdot\mathbf{b}_n + \mathbf{N}$, where $\mathbf{R}$ is the cross-correlation matrix between the different spreading codes of the users and $\mathbf{W} = \mathrm{diag}(w_1,...,w_K)$ is the amplitude matrix, where w_k denotes the amplitude of user k. All user bits at time n are represented by the vector $\mathbf{b}_n = (b_n(I_1),...,b_n(I_K))^T$ where $b_n(I_k)$ denotes the nth bit of user k. The channel noise vector $\mathbf{N}$ is white and composed of Gaussian zero mean variables with variances σ^2.

III. OPTIMAL SOFT DECODING

For decoder design we adopt the minimum mean square error (MMSE) criterion. That is, the decoder $\hat{\mathbf{X}}_k(\mathbf{Y})$, for user k, is designed to minimize $E\|\mathbf{X}_k - \hat{\mathbf{X}}_k(\mathbf{Y})\|^2$, where $\mathbf{Y} = (\mathbf{Y}_1,...,\mathbf{Y}_L)$ denotes the matrix of matched filter outputs. The main result of this paper is a formulation of the MMSE decoder based on estimates, $\hat{b}_k(\mathbf{y}_n) = \tanh(\sigma^{-2}\{(\mathbf{RW})^T\mathbf{y}_n\}_k)$, of the individual bits, $b_n(I_k)$. Here, $\{\mathbf{a}\}_k$ denotes the kth element of the vector $\mathbf{a}$. The derivation is based on the Hadamard transform description of a VQ [3]. To treat this in some more detail, let $\Im$ be the super-index defined by the binary forms of all the VQ encoder outputs, I_k, such that user 1 defines the L least significant bits and user K the L most significant bits of $\Im$. Also let $\mathbf{c}_\Im = [(\mathbf{c}_{I_1}^{(1)})^T,...,(\mathbf{c}_{I_K}^{(K)})^T]^T$ denote the vector composed by the centroids, $\mathbf{c}_i^{(k)} = E[\mathbf{X}_k | I_k = i]$, of the VQs. Then $\mathbf{c}_\Im$ can be described as $\mathbf{c}_\Im = \mathbf{T}\mathbf{h}_\Im$, where $\mathbf{h}_\Im$ is the $\Im$th column of a size 2^{KL} Hadamard matrix and $\mathbf{T}$ is a sparse transform matrix (c.f. [3]). It is easy to show that the MMSE estimate of the input vectors of all users is $\hat{\mathbf{X}}(\mathbf{y}) = \mathbf{T}\hat{\mathbf{h}}(\mathbf{y})$, where $\hat{\mathbf{h}}(\mathbf{y}) = E[\mathbf{h}_\Im | \mathbf{Y} = \mathbf{y}]$. This leads to the expression

$$\hat{\mathbf{X}}(\mathbf{y}) = [(\hat{\mathbf{X}}_1(\mathbf{y}))^T,...,(\hat{\mathbf{X}}_K(\mathbf{y}))^T]^T = \mathbf{T}\cdot\frac{\mathbf{R}_{\mathbf{hh}}}{\mathbf{m}_\mathbf{h}^T\hat{\mathbf{p}}(\mathbf{y})}\cdot\hat{\mathbf{p}}(\mathbf{y})$$

for the MMSE decoder. Here $\mathbf{R}_{\mathbf{hh}} = E[\mathbf{h}_\Im\mathbf{h}_\Im^T\cdot f(\Im)]$ and $\mathbf{m}_\mathbf{h} = E[\mathbf{h}_\Im\cdot f(\Im)]$, where $f(\Im) = \exp(-(2\sigma^2)^{-1}\sum_n\|\mathbf{RWb}_n\|^2)$. Furthermore, the bit-estimates, $\hat{b}_k$, enter as $\hat{\mathbf{p}}(\mathbf{y}) = \hat{\mathbf{p}}_K\otimes\cdots\otimes\hat{\mathbf{p}}_1$, where $\hat{\mathbf{p}}_k = (1,\hat{b}_k(\mathbf{y}_L))^T\otimes\cdots\otimes(1,\hat{b}_k(\mathbf{y}_1))^T$. Here $\otimes$ denotes the Kronecker matrix product. Thus, the vector $\hat{\mathbf{p}}(\mathbf{y})$ consists of products of bit-estimates, $\hat{b}_k(\mathbf{y}_n)$, for all users at all different times. We name our decoder the Soft Multi-User Decoder (SMUD). The SMUD performs, as noted above, combined MSE-optimal user decorrelation and VQ decoding. Note that, in the decoder expression, *only* the vector $\hat{\mathbf{p}}(\mathbf{y})$ depends on the received signal $\mathbf{y}$. Furthermore, note that the expectations in the expressions for $\mathbf{R}_{\mathbf{hh}}$ and $\mathbf{m}_\mathbf{h}$ are taken over the statistics of the VQ indices. Thus, the a-priori index information is confined to these quantities. Since the SMUD is MSE-optimal it shows how to utilize the a-priori information in an optimal fashion where $\mathbf{R}_{\mathbf{hh}}$ and $\mathbf{m}_\mathbf{h}$ are used to modify the statistic $\hat{\mathbf{p}}(\mathbf{y})$ to account for the source statistics. This is in contrast to systems where VQ decoding is based on an ML-decision, not taking the source statistics into account. Note also that, since the Hadamard transform is a fast transform, the calculation of $\hat{\mathbf{h}}(\mathbf{y})$ from the received signal, $\mathbf{y}$, can be carried out using an order of $KL\cdot 2^{KL}$ operations [6].

IV. NUMERICAL SIMULATIONS

We have compared the SMUD to the Maximum Likelihood Multi-User Decoder (ML-MUD) [1] in combination with table-look-up VQ decoding on a CDMA system with 2 users having the same transmission energy ($w_1^2 = w_2^2$). The cross-correlation between the users is 0.7. A VQ trained for a first order Gauss-Markov source with correlation $\rho = 0.9$ was utilized for both users, and we used the sample vector dimension 6 and the block length L=6 bits. The performance measure is the signal-to-noise ratio (SNR) at the output of the decoder at a given Channel-SNR (CSNR), w_k^2/σ^2. The performance of the decoders is shown in the table below. As seen the SMUD outperforms the ML-MUD with more than 3 dB at low CSNRs. We have also observed that the performance gain increases with increasing cross-correlation between users, lower CSNRs, and lower VQ output entropies. Furthermore near-far resistance for the SMUD has been concluded from simulations.

CSNR (dB)	-1	3	7	11	15
SMUD (dB)	3.10	4.78	6.53	8.34	10.32
ML-MUD (dB)	-0.32	1.32	3.43	6.33	9.94

REFERENCES

[1] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Personal Communications*, vol. 2, no. 2, pp. 46-58, April 1995.

[2] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 799-809, July 1990.

[3] M. Skoglund and P. Hedelin, "A Soft Decoder Vector Quantizer for a Noisy Channel," in *Proc. ISIT '94*, p. 401, Trondheim, Norway, Jun 1994.

[4] V. A. Vaishampayan and N. Farvardin, "Joint Design of Block Source Codes and Modulation Signal Sets," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 1230-1248, July 1992.

[5] F.-H. Liu, P. Ho, and V. Cuperman, "Joint Source and Channel Coding Using a Non-Linear Receiver," in *Proc. ICC '93*, pp. 1502-1507, Geneva, Switzerland, May 1993.

[6] M. Skoglund, "A Soft Decoder Vector Quantizer for a Rayleigh Fading Channel - Application to Image Transmission," in *Proc. ICASSP 95*, pp. 2507-2510, Detroit, May 1995.

Linear Multiuser Detectors for Synchronous Code-Division Multiple-Access Systems with Continuous Phase Modulation

Aris Papasakellariou and Behnaam Aazhang [1]
Elec. & Comp. Eng. Dept., Rice University, Houston, Texas 77251-1892

Abstract — **We consider Code-division multiple-access (CDMA) systems with continuous phase modulation (CPM). In particular, two multiuser detection algorithms with linear computational complexity are proposed for a synchronous system. We demonstrate that the choice of an appropriate set of decision statistics is crucial for detection and we derive an efficient representation. The analysis is performed for two signal formats which exhibit different spectral and error rate characteristics. We determine the code design that maximizes the minimum Euclidean distance and show that the resulting CPM/CDMA signals can achieve significant performance improvements over conventional CDMA signals.**

I. Signal Model

CPM/CDMA signals are an attractive choice for communications over predominantly bandwidth and power limited channels since they combine the merits of both techniques. In particular, CDMA offers a series of desirable properties that include increased capacity, inherent diversity against multipath fading and the ability to coexist with narrowband interference. CPM provides signals with compact spectral characteristics that maintain a constant envelope and hence are immune to nonlinear distortions and easily amplified [1].

We consider a CPM/CDMA system with K active users. The k^{th} transmitted signal is given as

$$s_k(t, \mathbf{b}_k) = \sqrt{2w_k} \cos\left(2\pi f_c t + \theta(t, \mathbf{b}_k, \mathbf{c}_k, h) + \theta_{k,0}\right) \qquad (1)$$

where $\mathbf{b}_k = (\ldots, b_k(-1), b_k(0), b_k(1), \ldots)$ is the transmitted data sequence with $b_k(m) \in \{-1, 1\}$, $\mathbf{c}_k = (c_k(1), \ldots, c_k(N_c))$ is the spreading code of length N_c with $c_k(n) \in \{-1, 1\}$, and h is the modulation index [1]. The signal power is w_k, the carrier frequency is f_c and $\theta_{k,0}$ is an arbitrary constant initial phase. The phase function $\theta(t, \mathbf{b}_k, \mathbf{c}_k, h)$ contains all the information and its construction defines the signal format. The first format examined in this paper is similar to conventional CDMA in the sense that only one code is assigned to each user. Under that scenario however, only modest gains can be achieved in the error rate performance relative to conventional CDMA [2]. Another format introduced in [2] for a memoryless CPM/CDMA system, considers the case where each user has available a distinct pair of codes. The code that is used is determined by the transmitted information bit and the objective is to minimize the error probability. We discuss how to construct such codes, depending on the CPM parameterization, and provide the lower bound for the error rate.

II. Multiuser Detection

We assume that all users employ the same type of CPM [1], and that the signals are transmitted in a synchronous, additive white Gaussian noise channel. Similarly to conventional CDMA, the complexity of the optimum detector increases exponentially with the number of users and the number of trellis states. Clearly, the implementation of the optimum detector is impractical and this motivates the need to develop linear multiuser detectors. To achieve linear complexity and near-optimum performance, a suboptimum detector must decouple the multiuser detection problem and subsequently perform single-user detection by individually recovering the metrics of each user. The optimal single-user detector can then be recursively implemented using the Viterbi algorithm. For single-user detection, each path metric becomes equivalent to the correlation between the received signal and the corresponding estimated transmitted signal. Denoting by $\hat{\theta}_i(m)$ the i^{th} trellis state during the m^{th} bit interval, the branch metric of the k^{th} user that is associated with the transmission of $\hat{b}_k(m) = \pm 1$ from the $\hat{\theta}_i(m)$ state is given as

$$L_k(\hat{b}_k(m), \hat{\theta}_i(m)) = \sum_{j=1}^{K} L_{k,j}(\hat{b}_k(m), \hat{\theta}_i(m)) + n(m) \qquad (2)$$

where $L_{k,k}(\hat{b}_k(m), \hat{\theta}_i(m))$ is the metric of the desired signal, $L_{k,j}(\hat{b}_k(m), \hat{\theta}_i(m))$, $j \neq k$, are the interference metric components and $n(m)$ is zero-mean Gaussian noise. Naturally, the objective of the multiuser detector is to remove the interference component from the metrics of the desired user. However, an attempt to directly evaluate the effects of the interference on the metrics of each user leads to prohibitively complex expressions for the decision statistics and an alternative approach is necessary. We prove that the decision statistics can be considerably simplified if they are expressed in terms of the difference and the sum of the two branch metrics that emanate from a common trellis state. That linear transformation reduces the complexity of the multiuser detector while preserving the metric information required by the Viterbi algorithm.

We propose two linear multiuser detection algorithms that are based on properties of the decision statistics and utilize concepts applied in multiuser detection of conventional CDMA signals. Both algorithms achieve near-optimum performance and can be employed for either signal format. We derive the conditions that maximize the minimum Euclidean distance and evaluate the optimum performance which can exhibit a gain that approaches 3 dB over binary antipodal signaling. The strict dependence between spectral and error probability performance that exists in typical CPM signals is largely decoupled and CPM/CDMA signals allow considerable flexibility in selecting a parameterization that satisfies certain spectral, error rate and complexity constraints.

[1] This work is supported by the Advanced Technology Program of the Texas Higher Education Coordinating Board under Grant 003604-018

References

[1] J. B. Anderson, T. Aulin and C.-E. Sundberg, "Digital Phase Modulation", Plenum, New York, N.Y., 1986.

[2] A. Svensson, C.-E. Sundberg and G. Lindell, "On Direct Sequence Spread Spectrum Systems with Continuous Phase Modulation," *in Proc. 1985 CISS*, pp. 526-531.

A Near Ideal Whitening Filter for M-algorithm Detection in an Asynchronous Time-Varying CDMA System[1]

Lei Wei
Telecommunications Engineering Group, RSISE, The Australian National University, Canberra, ACT 0200, Australia
Lars K. Rasmussen
Mobile Communications Research Centre, ITR, University of South Australia, The Levels, SA 5095, Australia

Abstract — **In this paper a near ideal noise whitening filter for a time-varying CDMA system is considered. The structure of the ideal noise whitening filter is studied and the metric function for tree search detection is derived. The ideal noise whitening filter for a time-varying CDMA system depends on unknown, future system parameters and is therefore difficult to realize. A near ideal, realizable noise whitening filter is proposed as a solution.**

I. Introduction

Recently joint multiuser detection, in which the multiuser interference is treated as a part of the information rather than noise, has attracted much attention. The work of Verdú [1] has shown that optimum near-far resistance and a significant performance improvement over the conventional detector is achieved by an optimum maximum likelihood multiuser detector. The substantial improvements, however, are obtained at the expense of a dramatic increase in computational complexity. The complexity grows exponentially with the number of users. Thus, when the number of users is large, the optimum detector becomes infeasible. It is therefore desirable to use a near optimum, low complexity detector for CDMA systems with a large number of users. Many low complexity multiuser detectors have been proposed (see references in [2]). Sub-optimal tree search algorithms such as sequential detection and the M-algorithm are especially promising. The IDDFD detector suggested by Wei and Schlegel [3] is essentially the M-algorithm applied over all users in a given time slot.

II. M-Algorithm Detection

Wei et al. [2] have shown that, in contrast to the case of the optimum multiuser detector, the choice of the receiver filter severely influences the performance of sub-optimum multiuser detectors. Detectors based on the M- or the T-algorithms and a noise whitening receiver filter generally perform better than similar detectors using only the matched filter. The M- and T-algorithm detectors based on noise whitening filter outputs can achieve near optimum performance at a very low complexity compared to the optimum detector. The M-algorithm can easily be applied to a time-invariant, asynchronous CDMA system, assuming that the noise whitening filter exists. In a practical system the noise whitening filter is related to time-varying system parameters. Time variations such as arrival and departure of users, random signature waveforms, and multipath effects make it necessary to derive the noise whitening filter following each system change. Wijayasuriya et al. [4] have suggested the sliding window decorrelating receiver in. However, the derivation of adaptive filters is not easily accommodated using this technique. In the control theory area, a factorization method has been suggested by Youla and Kazanjian [5]. An alternative method has been suggested by Alexander and Rasmussen to factorize the CDMA multiuser channel [6].

III. Near ideal filter

In this paper, we show that the method of Youla and Kazanjian can be generalized to derive a near ideal noise whitening filter for a time-varying asynchronous CDMA system. The structure of the ideal noise whitening filter is studied and the metric function for the M-algorithm based on the ideal noise whitening filter is derived. A near ideal, realizable noise whitening filter is then introduced. The convergence of the factorization method for a time-varying CDMA system is considered. The truncation of the number of taps of the ideal noise whitening filter is studied and the metric function for the M-algorithm based on the near ideal noise whitening filter is formulated. Simulation results are obtained for 5, 7 and 10-user time-varying CDMA systems with binary random signature sequences of length 10 and a rectangular chip waveform. The results show that the near ideal noise whitening filter can accurately approximate the ideal noise whitening filter at a low complexity level. The performance degradation of a time-varying, asynchronous CDMA system using a typical near ideal noise whitening filter is minimal compared to a system using the ideal noise whitening filter.

[1] This work was supported in parts by Telecom Australia under Contract No. 7368 and by the Commonwealth of Australia under International S & T Grant No. 56. The results of this work form parts of Australian Provisional Patent Application No. PM9548/94.

References

[1] S. Verdú, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels," *IEEE Trans. Inform. Theory*, vol. 32, pp.85-96, Jan. 1986.

[2] L. Wei and C. Schlegel, "Synchronous DS-SSMA System with Improved Decorrelating Decision-Feedback Multiuser Detection," *IEEE Trans. on Veh. Technol.*, vol. 43, pp. 767-772, Aug. 1994.

[3] L. Wei, L. K. Rasmussen and R. Wyrwas, "Bit-Synchronous CDMA Systems With Tree-Search Detection Over a Flat Fading Rayleigh Channel," in the *Proc. 1994 Int. Symp. Inform. Theory Appl.*, pp 91-95, Sydney, Australia, Nov. 1994.

[4] S. Wijayasuriya, G. Norton and J. McGeehan, "Sliding Window Decorrelating Receiver for DS-CDMA Receivers, *IEE Elec. Letters*, vol.-28, pp.1596-1597, Aug. 1992.

[5] D. Youla and N. N. Kazanjian, "Bauer-Type Factorization of Positive Matrices and the Theory of Matrix Polynomials Orthogonal on the Unit Circle," *IEEE Trans. on Circuits and Systems*, vol. 25, pp.57-69, Feb. 1978.

[6] P. D. Alexander and L. K. Rasmussen, "An Efficient Technique for Deriving Receiver Filters in Multiuser Asynchronous DS/SSMA," in the *Proc. PIMRC 1994*, pp 519-523, The Hague, The Netherlands, Sep. 1994.

A New Projection Receiver for Coded Synchronous Multi-User CDMA Systems

Christian Schlegel
Division of Engineering
The University of Texas at San Antonio
San Antonio, Texas 78249

Tel: (210) 691 5939, Fax: (210) 691 5589
Email: chris@beast.eng.utsa.edu

Zengjun Xiang
INRS Telecommunications
University of Quebec
Verdun, H3E 1H6, Canada

Tel: (08) 302 3881, Fax: (08) 302 3873
Email: xiang@inrs-telecom.uquebec.ca

SUMMARY: Over the past few years research into multi-user receivers for code-division multiple-access (CDMA) networks has become increasingly more popular. Multi-user detectors treat the interference from other users in the same frequency band as an information bearing part of the signal, rather than as noise.

It is known that optimal near-far resistance and significant performance improvements can be achieved by an optimal multi-user detector [1]. These improvements, however, are achieved at the expense of a dramatic increase in computational complexity, which grows *exponentially* with the number of users, making the optimum detector an unachievable theoretical concept. It becomes desirable to use near-optimum, low complexity detectors instead, and a number of sub-optimal approaches to the detection problem have been studied in detail. Surprisingly, near-optimal performance for uncoded CDMA can be achieved with a non-linear tree-search detector, whose complexity increases only *linearly* in the number of users [2].

While all these studies were undertaken for uncoded systems, the application of forward error control coding (FEC) to improve performance and system capacity remains a largely open area of future research. In this paper we study a very promising detector structure for coded CDMA, termed projection receiver (PR), whose structure is built on the decorrelating detector [3]. In the PR the effect of the interfering users is accounted for by metric adjustments, upon which the error control decoder operates. The actual interference resolution has a complexity which is proportional to the number of users, and all that is required are single user error control decoders.

The simple addition of an FEC system to an uncoded multi-user receiver may not lead to the best performance. This is evidenced in the performance plots shown in Figure 1, where three systems are compared for synchronous CDMA using length $N = 31$ random signature sequences and 64-state convolutional error control codes. Application of FEC to the conventional detector (correlation detector) leads to the poorest performance. FEC and decorrelation works better, but the PR performs best. It virtually achieves the single user bound, which is the theoretical performance limit for multi-user CMDA, for an arbitrary number of users up to a fully loaded system. That is, the PR effectively eliminates multi-user interference.

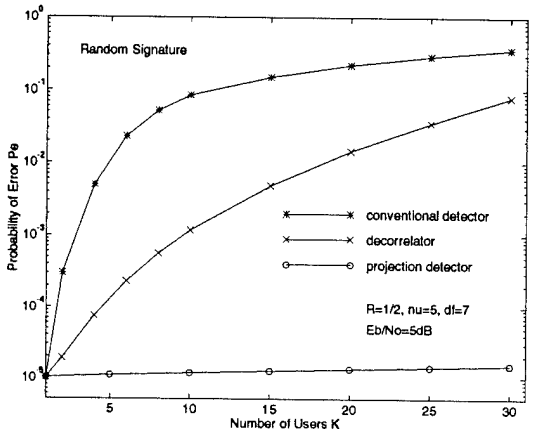


Figure 1: Performance of coded CDMA multi-user systems as a function of system load.

The projection receiver linearly projects the effects of the interfering users onto the complement of the subspace spanned by those users. In effect the PR decorrelates the unwanted users. From this an adjusted metric results, which has the form of a diversity metric, and is used in the FEC decoder of the desired user. As evidenced by Figure 1, this approach achieves single-user performance on an additive white Gaussian noise (AWGN) channel. In this presentation we will present the theory of the PR, performance results and discuss adaptive implementations of the detector which are suitable for VLSI implementations.

ACKNOWLEDGMENT: The continued interest and support of our colleagues Paul Alexander, Lars Rasumssen and Sumit Roy is gratefully acknowledged.

REFERENCES

[1] S. Verdu, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels", *IEEE Trans. Inform. Theory*, Vol. 32, No. 1, January 1986.

[2] W. Lei and C. Schlegel, "Synchronous DS-SSMA With Improved Decorrelating Decision-Feedback Multiuser Detection", *IEEE Trans. Veh. Tech.*, Vol. 43, No. 3, August 1994,

[3] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channel", *IEEE Trans. Inform. Theory*, Vol. IT-35, pp.123–136, Jan., 1989

Adaptive multilevel coding associated with CCI cancellation for CDMA

Ahmed Saifuddin †, Ryuji Kohno †

† Div. of Elect. & Comp. Eng., Yokohama National University,
156 Tokiwadai, Hodogaya-ku, Yokohama-240, Japan

Abstract — **For transmission of speech or multimedia information in a time varying mobile channel fixed rate codes are normally used designed for average or worst channel conditions. However, fixed rate codes fail to explore the time varying nature of the mobile channel. In this report we propose an adaptive multilevel coding scheme for Code Division Multiple Access (CDMA) which is associated with co-channel interference (CCI) cancellation to explore the time varying nature of the radio link.**

I. Introduction

For efficient usage of available spectrum and to explore the time varying nature of the mobile radio link, adaptive coding/modulation (codulation) scheme may be employed [1] [2]. In this paper, which is essentially the extension of our previous work [3], we took into account CSI and propose an adaptive multilevel coding scheme associated with multi user interference cancellation for CDMA, which yields significant performance improvement.

II. System Model

The information stream of each user is stored in a buffer prior to transmission from where informations are sent adaptively according to the channel condition. Adaptation can be done symbol by symbol or block by block according to CSI. We assume both transmitter and receiver can sense any change in channel condition at discrete instants of symbol transmission. For adaptation we changed the number of encoded levels according to CSI with modulation format held fixed. The transmitter decides what overall rate should be transmitted according to a set of thresholds chosen to keep the BER below a certain level. All the transmitters first look at the immediate values of fading multiplicative distortions. For a three level 8PSK if $0 \leq Max(\alpha_k^2) \leq \mu_1$ all the three level coding is done. The overall transmission rate is low in this case for worst channel conditions. If $\mu_1 \leq Max(\alpha_k^2) \leq \mu_2$, first two rows are encoded. Otherwise only one level is encoded with much higher rate. The receiver of any arbitrary kth user correlates the complex signal with each of the possible signal points of the partitioned signal constellation after despreading. Using the channel history, corresponding decoding scheme is chosen and the first component code is decoded. From all such precise decoded information of all users, CCI is estimated and is subtracted subsequently from the delayed version of the received signal to have more accurate estimate of the received signal. This process is carried on till all the component codes are decoded.

III. Result and Conclusion

For good channel condition we used one level coding using rate 1/2, M=4 convolutional code with $d_{free} = 7$. As the channel condition deteriorates we use 2nd and 3rd level codes which are rate 2/3 convolutional code with M = 4 and $d_{free} = 4$. The BER and throughput graphs are shown in Fig. 1 for a total 10 users. Fig. 2 shows the performance of in terms of throughput. We found that about 10dB gain can be achieved by using adaptive scheme compared to fixed rate scheme at a BER of 10^{-3}.

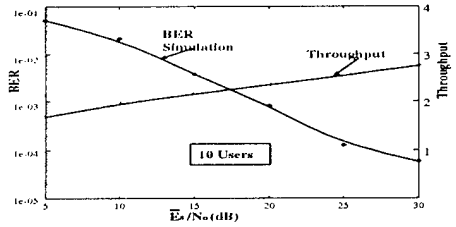


Fig. 1 BER and throughput curve of the proposed scheme in Rayleigh fading

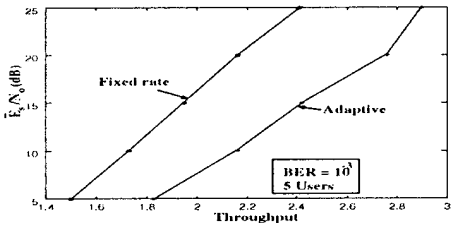


Fig. 2 Throughput comparison with fixed rate scheme

Acknowledgements

The authors would like to express their appreciation to Prof. Hideki Imai of University of Tokyo for valuable discussion.

References

[1] S. M. Alamouti, S. Kallel, " Adaptive Trellis-Coded Multiple-Phase-Shift Keying for Rayleigh Fading Channels, " *IEEE Trans. on Comm.*, Vol. 42, No. 6, pp. 2305-2314, June 1994.

[2] A. Goldsmith, " Variable-Rate Coded MQAM for Fading Channels, " *Proceedings of GLOBECOM*, pp. 186-190, December 1994.

[3] A. Saifuddin, R. Kohno, H. Imai, " Integrated Receiver Structure of Staged Decoder and CCI Canceller for CDMA with Multilevel Coded Modulation, " *Europ. Trans. on Telecomm. and Related Technol.*, Vol.6, No. 1, pp. 9-19, Jan-Feb, 1995

Performance Bounds for Decorrelator Detectors in a QS-CDMA System

Ronald A. Iltis[1]
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract — **A linear decorrelator detector is considered for use in a quasi-synchronous code-division multiple access (QS-CDMA) system. For long code lengths, the evaluation of bit-error rate can be computationally expensive, due to the need for exhaustive search to determine the worst case relative delays. An upper bound on the error rate based on eigenvalue bounds is presented for the linear decorrelator detector, and which can be computed solely in terms of the maximum cross-correlation between codes and the number of users.**

I. Introduction

A QS-CDMA communication system is considered in which decorrelators are employed for multiuser detection at the base station. In contrast to the decorrelator-based receiver in [1], the delays are assumed unknown a-priori, although confined to a subinterval of the bit duration due to the quasi-synchronous assumption. The worst-case bit-error rate (BER) of the decorrelator detector can be evaluated [2] by exhaustive search over the relative times-of-arrival (code delays) of the users. However, for long code lengths, such an approach may be extremely time-consuming, due to the computational burden of evaluating repeated correlations. Thus, we seek an upper bound on BER that can be used for long PN codes, and that does not require determination of the worst-case code delays.

The signal model for the QS-CDMA system is first described. Let $s_n(t)$ represent the direct-sequence signal transmitted by the n-th user. The received Nyquist samples, where the sampling interval is T_c sec., are given by

$$r(kT_c) = \sum_{n=1}^{N} a_n s_n(kT_c - T_n) + n(kT_c), \tag{1}$$

where $a_n \in \mathcal{C}$ is the complex amplitude associated with the n-th user, and T_n is the n-th delay. The additive noise sequence $n(kT_c) \in \mathcal{C}$ is discrete-time white Gaussian. Due to the quasi-synchronous assumption, $T_n \in [-MT_c, MT_c]$, where $MT_c << T$, with T the bit duration. It will be convenient to work with the following vector model of the received signal during the k-th bit duration.

$$\mathbf{r}(k) = a_1 d_1(k)\mathbf{s}_1(T_1) + \sum_{n=2}^{N} a_n \mathbf{s}_n(T_n) + \mathbf{n}(k), \tag{2}$$

where the elements of $\mathbf{s}_n(T_n) \in \mathcal{C}^L$ are the Nyquist samples $s_n(kT_c - T_n)$.

[1]This work was sponsored in part by Rockwell International Co. and the UC MICRO program.

II. Description of the Decorrelator Detector and Upper Bound

An approximate maximum-likelihood receiver for the signal model (2) has been previously derived [2], and is described by the following decision variable, where it is assumed that $\mathbf{s}_1(T_1)$ is the desired signal.

$$U = Re\{\mathbf{r}(k)^H[\mathbf{I} - \mathbf{P}_{S_1'}]\mathbf{s}_1(T_1)e^{iarg a_1}\}, \tag{3}$$

where $[\mathbf{I} - \mathbf{P}_{S_1'}]$ is an orthogonal projection matrix which rejects the undesired users. The projection matrix $\mathbf{P}_{S_1'}$ corresponds to the subspace spanned by the signals $\mathbf{s}_n(mT_c)$. As shown in [2], undesired vectors with delays T_n falling between the discretized values mT_c are nearly rejected, since they fall approximately in the subspace spanned by the columns of $\mathbf{S}_1'$.

A bound is obtained for an SNR loss factor, defined in terms of $SNR = E\{U\}/\sqrt{2Var\{U\}}$. Then the loss factor is given by $LF = SNR/\sqrt{E_b/N_0}$. Hence, if no loss in SNR when compared with ideal BPSK occurs, $LF = 1$. A lower bound on the SNR is found in terms of the following quantities γ_1 and γ_2 derived in [3], where t_{max} denotes the maximum normalized cross-correlation between codes.

$$\gamma_1 = 1 - \frac{t_{max}^2(2M+1)(N-1)}{1-(2M+1)(N-1)t_{max}}, \tag{4}$$

$$\gamma_2 = t_{max} + \Sigma_{max}\frac{t_{max}^2(2M+1)(N-1)}{1-((2M+1)(N-1)-1)t_{max}}. \tag{5}$$

The term Σ_{max} is given by

$$\Sigma_{max} = \sup_{\{m:-M\le m\le M\},\epsilon} \sum_{k=M-m+1}^{L-M-m-1} \sum_{n=-\infty}^{\infty} sinc(k + nL - \epsilon/T_c). \tag{6}$$

The final expression for the loss factor is then

$$LF = \sqrt{\gamma_1} - \sum_{n=2}^{N} \sqrt{\frac{J}{S}}\frac{\gamma_2}{\sqrt{\gamma_1}}. \tag{7}$$

Note that when the actual delays T_n equal the dicretized values mT_c, the term $\gamma_2 = 0$. Specfic results for the loss factor are evaluated for varying Gold code lengths and SNRs in [3]. In general, the bound is useful for long PN codes, where exhaustive search to find the worst-case relative delays is computationally prohibitive.

References

[1] R. Lupus and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Transactions on Communications*, vol. 38, pp. 496–508, April 1990.

[2] R. Iltis and L. Mailaender, "Multiuser detection for quasi-synchronous signals." Submitted to the IEEE Transactions on Communications.

[3] R. Iltis, "An upper bound for the error rate of linear decorrelator detectors." Submitted to the IEEE Transactions on Information Theory.

A Non-Orthogonal Synchronous DS-CDMA Case, Where Successive Cancellation and Maximum-Likelihood Multiuser Detectors are Equivalent

Peter Kempf
Institut fuer Netzwerk- und Signaltheorie, Technische Hochschule Darmstadt, Germany
E-Mail: pkempf@nesi.e-technik.th-darmstadt.de, Phone: (+49)-06151/16-4796

Abstract — **A non-orthogonal synchronous direct sequence code division multiple access (DS-CDMA) system with additive white Gaussian noise channel (AWGN) is presented where the suboptimal successive cancellation detector performs optimal.**

I. Introduction

Due to recent advances in cellular technology [1], DS-CDMA has been considered as multiple access method. It is well-known that joint detection of the users improves system capacity considerably [2]. The maximum likelihood (ML) decision rule over all active users is optimal in the sense of the estimation error rate, but in general too complex due to the exponential dependence on the number of users. The compexity (in terms of operations per bit decision) of successive cancellation is linear in the number of users.

II. System Description

Fig. 1 shows the considered synchronous DS-CDMA system modell with AWGN-channel. On account of the synchronism and the memoryless channel, each bit period can be considered independently. All vectors denote column vectors. The users $1 \dots K$ transmit bits $b_1, b_2, \dots, b_K \in \{-1, 1\}$ by modulating them onto user-specific spreading-vectors $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K$ of length N. The components of the spreading vectors are considered to be real numbers and the Euclidean norm of the vectors is equal for all vectors

$$|\mathbf{c}_k|^2 = \mathbf{c}_k^T \mathbf{c}_k = 1. \tag{1}$$

The superscript $()^T$ denotes the transpose. The noise vector $\mathbf{n}$ is assumed to be Gaussian with covariance matrix $\sigma_n^2 \mathbf{I}$ ($\mathbf{I}$ is the identity matrix). With $\mathbf{b}$ the transmitted bit vector and $\mathbf{C}$ the spreading code matrix, the received vector $\mathbf{r}$ can be written as

$$\mathbf{r} = \mathbf{C}\mathbf{b} + \mathbf{n}, \quad \mathbf{b} \triangleq (b_1\ b_2 \cdots b_K)^T, \quad \mathbf{C} \triangleq (\mathbf{c}_1 | \mathbf{c}_2 | \cdots | \mathbf{c}_K). \tag{2}$$

The receiver's task is to estimate the bits $b_1, \dots, b_K$ from the observation of $\mathbf{r}$. The optimal decision rule for equiprobable input bits is the ML rule, which minimizes the Euclidean distance between $\mathbf{r}$ and $\mathbf{C}\hat{\mathbf{b}}$ where $\hat{\mathbf{b}}$ is the estimate of $\mathbf{b}$.

Definition 1 *ML rule: Choose the estimated bit vector* $\hat{\mathbf{b}}^{ML}$ *such that the Euclidean distance e ist minimized with*

$$e^2 \triangleq \left| \mathbf{C}\hat{\mathbf{b}}^{ML} - \mathbf{r} \right|^2 = (\mathbf{C}\hat{\mathbf{b}}^{ML} - \mathbf{r})^T (\mathbf{C}\hat{\mathbf{b}}^{ML} - \mathbf{r}). \tag{3}$$

A suboptimal decision rule called "successive interference cancellation" (SC rule) uses first a bank of matched filters MF1...MFK to produce decision variable $d_1, d_2, \dots, d_K$ (see Fig. 1). The decision variable vector $\mathbf{d}$ can be written as

$$\mathbf{d} \triangleq (d_1\ d_2 \cdots d_K)^T = \mathbf{C}^T\mathbf{r} = \mathbf{C}^T\mathbf{C}\mathbf{b} + \mathbf{C}^T\mathbf{n} = \mathbf{R}\mathbf{b} + \mathbf{C}^T\mathbf{n}. \tag{4}$$

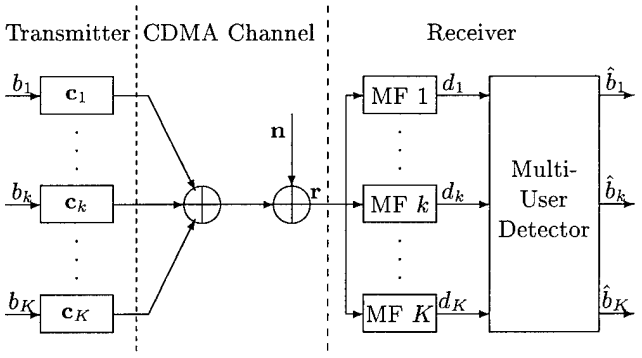


Figure 1: System modell (DS-CDMA, AWGN-channel)

Definition 2 *SC rule: Let the reliability of a decision variable* $\mathrm{rel}(d_k)$ *be defined by the absolute value:* $\mathrm{rel}(d_k) = |d_k|$. *Let* $\mathbf{r}^{(1)} = \mathbf{r}$ *and* $\mathcal{S}_s = \{k_1, k_2, \dots, k_{s-1}\}$, *where* $\mathcal{S}_s$ *is the set of indices for which a decision has been taken in steps* 1 *up to step* $s-1$. *Initially* $\mathcal{S}_1 = \{\}$. *Then successive cancellation chooses the estimates* $\hat{b}_k^{SC}$ *in* K *steps as follows:*
At step s *compute* $\mathbf{d}$ *from* $\mathbf{r}^{(s)}$ *and choose the decision variable* d_k *with highest reliability taking into consideration only indices* $k \notin \mathcal{S}_s$. *Decide on bit* $\hat{b}_k^{SC}$ *using the sign function*

$$\hat{b}_k^{SC} = \mathrm{sgn}(d_k), \tag{5}$$

form the set $\mathcal{S}_{s+1} = \mathcal{S}_s \cup \{k\}$ *and compute* $\mathbf{r}^{(s+1)}$ *as*

$$\mathbf{r}^{(s+1)} = \mathbf{r}^{(s)} - \hat{b}_k^{SC}\mathbf{c}_k. \tag{6}$$

III. Result

Theorem: *SC and ML rules are equivalent if the cross-correlation* $\mathbf{R} \triangleq \mathbf{C}^T\mathbf{C}$ *satisfies for a constant* q *with* $|q| < 1$

$$R_{i,j} = \begin{cases} q & i \neq j \\ 1 & else \end{cases}. \tag{7}$$

Proof: The proof has two steps, of which the first shows that for the first decision in successive cancellation the resulting bit estimate is equal to the maximum likelihood estimate. The second step shows that after subtracting the influence of the estimated bit the problem is principally the same, only the dimension has decreased by 1.

References

[1] Qualcomm Inc., "Proposed EIA/TIA interim standard," tech. rep., Qualcomm Inc., Apr. 1992. Document Number 80-7814 Rev - DCR 03567.

[2] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

An Inequality on Guessing and Its Application to Sequential Decoding

ERDAL ARIKAN
Electrical Engineering Department, Bilkent University, 06533 Ankara, Turkey

Abstract — **Let (X,Y) be a pair of discrete random variables with X taking values from a finite set. Suppose the value of X is to be determined, given the value of Y, by asking questions of the form 'Is X equal to x?' until the answer is 'Yes.' Let $G(x|y)$ denote the number of guesses in any such guessing scheme when $X=x$, $Y=y$. The main result is a tight lower bound on nonnegative moments of $G(X|Y)$. As an application, lower bounds are given on the moments of computation in sequential decoding. In particular, a simple derivation of the cutoff rate bound for single-user channels is obtained, and the previously unknown cutoff rate region of multi-access channels is determined.**

I. THE INEQUALITY

Theorem 1 *For arbitrary guessing functions $G(X)$ and $G(X|Y)$, and any $\rho \geq 0$,*

$$E[G(X)^\rho] \geq (1+\ln M)^{-\rho}[\sum_{x\in\mathcal{X}} P_X(x)^{\frac{1}{1+\rho}}]^{1+\rho} \qquad (1)$$

and

$$E[G(X|Y)^\rho] \geq (1+\ln M)^{-\rho}\sum_{y\in\mathcal{Y}}[\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^{\frac{1}{1+\rho}}]^{1+\rho} \qquad (2)$$

where $P_{X,Y}$, P_X are the probability distributions of (X,Y) and X, respectively, the summations are over all possible values of X, Y, and M is the number of possible values of X.

This result is a simple consequence of the following variant of Hölder's inequality.

Lemma 1 *Let a_i, p_i be nonnegative numbers indexed over a finite set $1 \leq i \leq M$. For any $0<\lambda<1$,*

$$\sum_{i=1}^{M} a_i p_i \geq \left[\sum_{i=1}^{M} a_i^{\frac{-\lambda}{1-\lambda}}\right]^{\frac{1-\lambda}{-\lambda}} \left[\sum_{i=1}^{M} p_i^{\lambda}\right]^{\frac{1}{\lambda}}$$

Proof. Put $A_i = a_i^{-\lambda}$, $B_i = a_i^{\lambda}p_i^{\lambda}$, in Hölder's inequality $\sum_i A_i B_i \leq \left(\sum_i A_i^{\frac{1}{1-\lambda}}\right)^{1-\lambda}\left(\sum_i B_i^{\frac{1}{\lambda}}\right)^{\lambda}$.

Proof of Theorem. Inequality (1) is obtained by taking $a_i = i^\rho$, $p_i = \Pr(G(X)=i)$, $\lambda = 1/(1+\rho)$ in the lemma, and noting that $\sum_{i=1}^{M} 1/i \leq (1+\ln M)$. Inequality (2) follows readily:

$$\begin{aligned} E[G(X|Y)^\rho] &= \sum_y P_Y(y) E[G(X|Y=y)^\rho] \\ &\geq \sum_y P_Y(y)(1+\ln M)^{-\rho}[\sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}}]^{1+\rho} \\ &= (1+\ln M)^{-\rho}\sum_y[\sum_x P_{X,Y}(x,y)^{\frac{1}{1+\rho}}]^{1+\rho} \end{aligned}$$

II. APPLICATION TO SEQUENTIAL DECODING

To relate sequential decoding to guessing, let $\mathcal{X}$ denote the set of nodes in a tree code at some level N channel symbols into the tree from the tree origin. Let X be a random variable uniformly distributed on $\mathcal{X}$, indicating the node in $\mathcal{X}$ which lies on the transmitted path. Let Y denote the received channel output sequence when X is transmitted. Let $G(x|y)$ denote the rank order in which node $x \in \mathcal{X}$ is hypothesized (for the first time) by a sequential decoder when $X=x$ and $Y=y$. Moments of $G(X|Y)$ serve as measures of complexity for sequential decoding.

Let M be the size of $\mathcal{X}$, and $R = (1/N)\ln M$ denote the code rate. By Theorem 1 and the fact that $P_X(x) = 1/M$ for $x \in \mathcal{X}$, for $\rho > 0$,

$$E[G(X|Y)^\rho] \geq (1+NR)^{-\rho}\exp[\rho NR - E_0(\rho, P_X)]$$

where

$$E_0(\rho, P_X) = -\ln \sum_y [\sum_x P_X(x) P_{Y|X}(y|x)^{\frac{1}{1+\rho}}]^{1+\rho}.$$

Gallager [1, p. 149] shows that for discrete memoryless channels

$$E_0(\rho, P_X) \leq N E_0(\rho)$$

where $E_0(\rho)$ equals the maximum of $E_0(\rho, Q)$ over all single-letter distributions Q on the channel input alphabet. Thus, at rates $R > E_0(\rho)/\rho$, the ρth moment of computation performed at level N of the tree code must go to infinity exponentially as N is increased. The infimum of all real numbers R' such that, at rates $R > R'$, $E[G(X|Y)^\rho]$ must go to infinity as N is increased is called the cutoff rate (for the ρth moment) and denoted by $R_{cutoff}(\rho)$. We have thus obtained the following bound.

Theorem 2 *For any discrete memoryless channel with a finite input alphabet,*

$$R_{cutoff}(\rho) \leq E_0(\rho)/\rho, \quad \rho > 0. \qquad (3)$$

This result was proved earlier (for $\rho = 1$ only) in [2]; the present proof is much simpler. Moreover, the above method extends to the case of multiaccess channels, yielding their previously unknown cutoff rate region [3].

ACKNOWLEDGEMENTS

I am indebted to J.L. Massey and M. Burnashev for discussions on this problem.

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[2] E. Arikan, 'An upper bound on the cutoff rate of sequential decoding,' *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 55-63, Jan. 1988.

[3] E. Arikan, 'An inequality on guessing and its application to sequential decoding,' submitted to *IEEE Trans. Inform. Theory*, Nov. 1994.

Weighted Coding Methods for Binary Piecewise Memoryless Sources

Frans Willems and Franco Casadei[1]

Electrical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract — **Coding methods for piecewise memoryless sources have recently been studied by Merhav [2]. While Merhav mainly concentrated on the single-transition case, we describe and analyse here coding techniques that allow multiple transitions.**

I. Introduction

A binary memoryless source generates the sequence $x_1 \cdots x_T$ with probability

$$P_a(x_1 \cdots x_T) = \Pi_{t=1,T} P_a(x_t).$$

The source parameter is piecewise constant. Suppose that the instants **before** which transitions appear are $t_1, t_2, \cdots, t_C$, i.e.

$$P_a(X_t = 1) = \theta_c, \text{ if } t_c \leq t < t_{c+1},$$

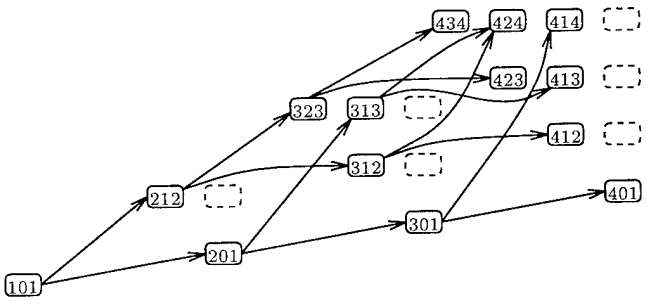
with $c = 0, 1, \cdots, C$, and $t_0 = 1$ and $t_{C+1} = T + 1$. We describe a method for compressing the sequences generated by this source based on arithmetic coding techniques (see [1]). It was our objective to construct a coding distribution $P_c(\cdot)$ such that the maximal individual redundancy

$$\max_{x_1 \cdots x_T} \log \frac{P_a(x_1 \cdots x_T)}{P_c(x_1 \cdots x_T)}.$$

is as small as possible. The base of the log is 2. Information quantities are expressed in bits.

II. Coding method

We assume that the source moves in a graph (see below) from state to state. It starts in state $(1, 0, 1)$ and generates the



symbol x_1 according to parameter θ_0. When the source is in state (t, c, t_c) it first generates symbol x_t according to parameter θ_c. After this, its parameter may change to θ_{c+1}, in that case the source moves to state $(t+1, c+1, t_{c+1} = t+1)$, or it may not change, then the source moves to state $(t+1, c, t_c)$. When the source is in state (t, t_c) we assume that

$$P_c(X_t = 1 | (t, c, t_c)) = \frac{b(x_{t_c} \cdots x_{t-1}) + 1/2}{t - t_c + 1},$$

where $b(x_{t_c} \cdots x_{t-1})$ is the number of ones in $x_{t_c} \cdots x_{t-1}$, and

$$P_{tr}((t+1, c+1, t+1) | (t, c, t_c)) = \frac{c + 1/2}{t}.$$

The source output x_t and next state $(t+1, p, q)$, where $(p, q) = (c, t_c)$ or $(c+1, t+1)$ are assumed to be independent of each other given the current state (t, c, t_c).

Under these assumptions the coding probability for sequence $x_1 \cdots x_t$ is the probability that the source moves along a certain path to a state at depth t and generates $x_1 \cdots x_t$, summed over all paths and states.

Two remarks can be made about this coding distribution. The first is that the coding distribution can easily be updated sequentially using the graph structure. The second remark is that the coding distribution can be regarded as a weighting (see [3]) over all coding distributions corresponding to fixed transition patterns. The weighting coefficient of a pattern is determined by the Krichevsky-Trofimov estimator. This makes it easy to study the redundancy behavior of our method.

III. Performance analysis

If the actual source made C transitions (pattern $\mathcal{T}$), our method yields

$$\begin{aligned} \log \frac{P_a(x_1 \cdots x_T)}{P_c(x_1 \cdots x_T)} &\leq \log \frac{P_a(x_1 \cdots x_T)}{P_c(x_1 \cdots x_T | \mathcal{T})} + \log \frac{1}{P_{tr}(\mathcal{T})} \\ &\leq \frac{C+1}{2} \log \frac{T}{C+1} + C + 1 \\ &\quad + C \log \frac{(T-1)e}{C} + \frac{1}{2} \log(T-1) + 1. \end{aligned}$$

Parameter redundancy is as usual (i.e. roughly $\frac{1}{2} \log T$ per parameter), the *transition* redundancy is roughly $\log T$ per transition, plus a bias term of $\frac{1}{2} \log T$. Apart from this bias term our method achieves the Merhav [2] lower bound. The bias term however is a consequence of the fact that the number of transitions is assumed to be unknown.

IV. Complexity

The storage complexity of this method grows quadratically in the sequence length T. The computational complexity is cubic in T.

There exists a simplified version of the method described here which has storage behavior linear in T and for which the number of computations is quadratical in T. For this method the redundancy is roughly $\frac{3}{2} \log T$ per transition however.

References

[1] J. Rissanen and G.G. Langdon, Jr. "Universal Modeling and Coding," *IEEE Trans. on Inform. Theory*, vol. IT-27, pp. 12-23, January 1981.

[2] N. Merhav, "On the Minimum Description Length Principle for Sources with Piecewise Constant Parameters," *IEEE Trans. on Inform. Theory*, vol. IT-39, pp. 1962-1967, November 1993.

[3] F.M.J. Willems, Y.M. Shtarkov and Tj.J. Tjalkens, "Context Tree Weighting : A Sequential Universal Source Coding Procedure for FSMX Sources," *IEEE Int. Symp. Inform. Theory*, San Antonio, Texas, January 17-22, 1993, p. 59.

[1] On leave from the University of Bologna, Italy.

A D-ary Huffman code for a class of sources with coutably infinite alphabets

Akiko KATO, Te Sun HAN, and Hiroshi NAGAOKA
Graduate School of Inform. Systems, Univ. of Electro-Communications,
Tokyo, Japan

Abstract — **We briefly describe the results in [4].**

I. Introduction and the main result

Let us consider an information source S with probability distribution $P = \{p(i)\}_{i=0}^{\infty}$ on the *infinite* source alphabet $\mathcal{X} = \{0, 1, 2, \ldots\}$, where we assume throughout the positivity: $p(i) > 0$ for all $i \in \mathcal{X}$ as well as the monotonicity:

$$\min_{0 \le i \le k_0 - 1} p(i) \ge p(k_0) \ge p(k_0 + 1) \ge p(k_0 + 2) \ge \cdots \quad (1)$$

for some $k_0 \in \mathcal{X}$.

We shall consider prefix codings for the distribution P using the D-ary code alphabet $\mathcal{D} = \{0, 1, \ldots, D-1\}$ where D is an integer such that $D \ge 2$. Our concern is how to construct an *optimal* D-ary prefix code given a probability distribution P on $\mathcal{X}$, where "optimal" means a code with the *minimum* expected codeword length over all the possible prefix codes for the P. Although the Huffman coding algorithm with *finite* source alphabet is known to achieve the optimal code, it is not applicable in general to the case with *infinite* source alphabet. However, some specific properties of the given distribution P are enough to ensure the applicability of Huffman-type coding algorithms as well also to the infinite alphabet case: for instance, see Gallager and Van Voorhis[2], Humblet[3], and Abrahams[1]. The sufficient conditions shown in these papers are all written in terms of inequalities that must hold for *all* $m \in \mathcal{X}$ larger than some integer.

In [4] we provide a new type of sufficient condition that merely includes inequalities for *infinitely many* m's in $\mathcal{X}$, which is stated as

Condition 1. *There exist* infinitely many m*'s in* $\mathcal{X}$ $(m \ge k_0)$ *such that*

$$m \equiv -1 \pmod{D-1} \quad (2)$$

and

$$p(m) \ge \sum_{i=m+D}^{\infty} p(i). \quad (3)$$

Remark 1. *It is evident that, in the binary case* $(D = 2)$*, condition* (2) *holds for all* $m \in \mathcal{X}$.

Our main result is

Theorem 1. *If the probability distribution* P *on* $\mathcal{X}$ *satisfies* Condition 1*, then there exists an algorithm to recursively construct an optimal* D*-ary prefix code* C^{H}.

II. Coding algorithm under Condition 1

To describe how to construct the code C^{H}, let us first introduce some notations. Let $m_1 < m_2 < \cdots$ be those integers m that satisfy conditions (2) and (3) where $m_1 \ge k_0$. For each $j = 1, 2, \cdots$, let us define a partition of $\mathcal{X}$ by $\mathcal{A}_j^0 = \{A_j^{(k)}\}_{k=0}^{m_j+1}$ where

$$A_j^{(k)} = \begin{cases} \{k\} & \text{for} \quad 0 \le k \le m_j, \\ \{i \mid i \ge k\} & \text{for} \quad k = m_j + 1. \end{cases}$$

For each $j = 1, 2, \cdots$ define the information source $\tilde{S}_j$ with the finite alphabet

$$\tilde{\mathcal{A}}_j \stackrel{\text{def}}{=} \mathcal{A}_j^0 \setminus \mathcal{A}_{j-1}^0 \{A^{(k)}\}_{k=m_{j-1}+1}^{m_j+1}$$
$$= \{\{m_{j-1}+1\}, \{m_{j-1}+2\}, \cdots, \{m_j\}, \{i \mid i \ge m_j + 1\}\}$$

and the probability distribution $\tilde{P}_j = \{\tilde{p}_j(A_j^{(k)})\}_{k=m_{j-1}+1}^{m_j+1}$ such that

$$\tilde{p}_j(A_j^{(k)}) = \frac{p(A_j^{(k)})}{\tilde{a}_j} \quad (m_{j-1} + 1 \le k \le m_j + 1),$$

where $p(B) = \sum_{i \in B} p(i)$ for a subset $B \subset \mathcal{X}$ and we have set $m_0 = -1$ $(\mathcal{A}_0^0 = \{\mathcal{X}\})$ and $\tilde{a}_j \stackrel{\text{def}}{=} \sum_{k=m_{j-1}+1}^{m_j+1} p(A_j^{(k)})$.

Now we are in the place to describe the coding algorithm to construct the code C^{H}.

Coding algorithm

Step 0 Set $j := 1$ and $C(*) := \lambda$ (null string).

Step 1 Construct a D-ary Huffman code $\tilde{C}_j^{\mathrm{H}}$ for the information source $\tilde{S}_j$.

Step 2 If $\{i\} \in \tilde{\mathcal{A}}_j$ then define the codeword for $i \in \mathcal{X}$ by

$$C^{\mathrm{H}}(i) := C(*) \cdot \tilde{C}_j^{\mathrm{H}}(\{i\}),$$

where "$\cdot$" denotes the concatenation of strings.

Step 3 Set $C(*) := C(*) \cdot \tilde{C}_j^{\mathrm{H}}(A_j^{(m_j+1)})$.

Step 4 Set $j := j + 1$ and go to **Step 1**.

Remark 2. *For each* $j \ge 1$*, just after* **Step 3** *the resulting code (the codewords are* $C^{\mathrm{H}}(i)$ *for* i $(0 \le i \le m_j)$*, and* $C(*)$ *for* $i = m_j + 1$ *) is a Huffman code for the source* $S_{\mathcal{A}_j^0}$ *(with finite alphabet* $\mathcal{A}_0^j$ *and probability distribution* $P_{\mathcal{A}_j^0} = \{p(A_j^{(k)})\}$*). Generalizing this property, we get a new definition of an optimal* D*-ary prefix code which is meaningful even for the case where* $H(P) = \infty$*. See [4] for details.*

References

[1] J. Abrahams: Huffman-type codes for infinite source distributions, *Journal of the Flanklin Institute,* to appear.

[2] R. A. Gallager and D. C. Van Voorhis: Optimal source coding for geometrically distributed integer alphabets, *IEEE Trans. Inform. Theory,* vol. 21, no. 2, pp. 228–230 (1975).

[3] P. A. Humblet: Optimal source coding for a class of integer alphabets, *IEEE Trans. Inform. Theory,* vol. 24, no. 1, pp. 110–112 (1978).

[4] A. Kato, T. S. Han, and H. Nagaoka: Huffman coding with infinite alphabet, *submitted to IEEE Trans. Inform. Theory.*

Data Expansion with Huffman Codes

Jung–Fu Cheng, Sam Dolinar[1], MichelleEffros[2], Robert McEliece
Electrical Engineering Department and Jet Propulsion Laboratory
California Institute of Technology

Abstract — **"In-place" Huffman coding of a file can cause the file to temporarily expand. In this paper we investigate this phenomenon,**

I. Introduction

Huffman codes are widely used for data compression. In a typical application, a file consisting of N m-bit symbols is compressed by an adaptive version of Huffman's algorithm, in which the required symbol probabilities are determined by the relative frequencies of the symbols in the file. Each m-bit symbol in the file is then replaced by the corresponding Huffman codeword. It is clear that if such a strategy is used, the file cannot expand, since the "worst case" is when all the Huffman codewords have length m bits, and in all other cases the file will indeed be strictly compressed.

However, if the compression is done sequentially and "in place," that is, if the first symbol in the file is encoded, then the second, etc., the file may temporarily expand if many low-probability symbols occur at or near the beginning of the file. In space-critical implementations of Huffman's algorithm, it will then be important to know the amount of extra storage space that must be allocated to allow for this temporary expansion.

The general problem we address in this paper, then, is this. For a file consisting of N letters from a source alphabet of 2^m symbols, what is the maximum possible "temporary expansion" possible for a Huffman code, in units of *bits per file symbol*? We denote this quantity by $\delta(m)$.

II. Example

Consider an 8-letter symbol alphabet $\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F},\mathtt{G},\mathtt{H}\}$, and a file consisting of the following 16 symbols from the alphabet.

HGFEDCBBBBAAAAAA.

If each symbol in the alphabet is given a 3-bit representation, the file length is 48 bits. The relative frequencies, and a set of appropriate Huffman codewords, for this file is given in the following table.

symbol	rel. freq.	codeword	length
A	3/8	0	1
B	1/4	10	2
C	1/16	1100	4
D	1/16	1101	4
E	1/16	11100	5
F	1/16	11101	5
G	1/16	11110	5
H	1/16	11111	5

After each of the 16 symbols in the file has been replaced by its Huffman codeword, a simple calculation shows that the fully compressed file is only 42 bits long. However, since the low-frequency symbols C, D, E, F, G, H all occur at the beginning of the file, after these six symbols have been Huffman encoded, the file's length will be 58 bits, which represents an expansion of 5/8 bits per source symbol. In fact, it follows from our results that for an eight-letter source alphabet this is the worst case, so that $\delta(3) = 5/8$. (It is easy to see that $\delta(1) = 0$ and $\delta(2) = 2/5$.)

III. Alternative definition of $\delta(m)$

There is a alternative definition of $\delta(m)$ that makes the problem easier to deal with.

Definition. *For $m = 1, 2, 3, \ldots$, define*

$$\delta(m) = \max\left\{\sum_{j=1}^{2^m} p_j(n_j - m)_+\right\},$$

where the maximum is taken over all pairs of lists $(p_1 \geq p_2 \geq \cdots \geq p_{2^m})$, $(n_1 \leq n_2 \leq \cdots \leq n_{2^m})$, where p_j's are an ordered list of probabilities, and the n_j's are the lengths of a corresponding Huffman code for the p_j's. (The symbol "$(x)_+$" is shorthand for $\max(x, 0)$.)

IV. Statement of Results

Theorem 1. *There is a universal constant Δ such that $\delta(m) \leq \Delta$ for all m.*

The proof of Theorem 1 relies on a recent result of Schack [3]. We conjecture that $\Delta = 4/5$ (incidentally, for "Shannon" codes it is quite easy to show that the corresponding quantity is $\Delta_{\text{Shannon}} = 1$), but so far we have only been able to prove that $4/5 \leq \Delta \leq 4$. The upper bound comes from a careful examination of the proof of Theorem 1. The lower bound comes from an explicit construction of a family of Huffman trees, using techniques similar to those developed in [1] and [2], for which the quantity δ equals $(4 \cdot 2^m - 12)/(5 \cdot 2^m - 8)$, for $m \geq 3$. The probabilities in the tree are proportional to $(2^m - 2, 2^{m-1}, 1, \ldots, 1)$, and the corresponding Huffman lengths are $(1, 2, m+1, m+1, m+2, \ldots, m+2)$. Furthermore, we believe that this construction gives the largest possible value of δ, so we have the following conjecture.

Conjecture 1. *$\delta(m) \uparrow 4/5$, as $m \to \infty$.*

Acknowledgement

We thank Dr. Douglas Whiting of STAC, Inc., for posing this interesting problem.

References

[1] Y. S. Abu-Mostafa and R. J. McEliece, "Maximal codeword lengths in Huffman codes," *JPL TDA Progress Report*, vol. 42-110 (August 1992), pp. 188–192.

[2] G. O. H. Katona and T. O. H. Nemetz, "Huffman codes and self-information," *IEEE Trans. Inform. Theory*, vol. IT-22 (May 1976), pp. 337–340.

[3] R, Schack, "The length of a typical Huffman codeword," *IEEE Trans. Inform. Theory*, vol. IT-40 (July 1994), pp. 1246–1247.

[1]Work done at JPL under contract with the National Aeronautics and Space Administration.

[2]Work done with partial support from the National Science Foundation

Minimizing the Maximum Codeword Cost

Julia Abrahams
julia@hornet.onr.navy.mil
Mathematical, Computer, and Information Sciences Division
Office of Naval Research, Arlington, VA 22217-5660

I. INTRODUCTION

Varn [8] introduced the problem of finding the prefix condition variable length source code which minimizes average cost when the code symbols are of unequal cost and the source symbols are equiprobable. Other authors have also addressed this problem from the algorithmic point of view [3,4,7]. There are two versions of this problem, exhaustive in which the r-ary code tree is constrained to be a full tree and nonexhaustive in which that constraint is not imposed. Recently the author [1] was able to show, based on previous work by Horibe [5] and Chang [2], for the exhaustive case that for integer code symbol costs there exists a very close relationship between a subsequence of the sequence of Varn code trees indexed by the number of leaves in the tree and a recursively generated sequence of trees called generalized Fibonacci trees. In particular the kth tree in the recursively generated sequence of trees has as its ith leftmost subtree, i=1,...,r, the tree previously generated in the sequence with index k-c(i) where the code symbol costs are c(i) ordered monotonically nondecreasing in i and the associated code symbols are associated with the code tree branches from left to right. The initialization is that the first c(r) trees are all single root nodes. Then, when the number of leaves in the exhaustive Varn code tree is the same as the number of leaves in the generalized Fibonacci tree for the same code symbol costs, they are the same tree. The recursive construction is nice in that it reveals an elegant structure underlying the sequence of Varn code trees and also because recurrence relations derived from the recursive construction permit the evaluation of the resulting minimum average cost codes without actually constructing the tree.

The problem addressed in this abstract is to identify a similar recursive construction for the nonexhaustive case. It turns out that it is possible to do this not for Varn's original problem, but for a close variant of it. While Varn looks for optimum codes in the minimum average codeword cost sense, the problem of interest here will be to look for optimum codes in the sense of minimizing the maximum codeword cost. It is not hard to see that in the exhaustive case, Varn's algorithm finds optimum code trees in both senses, that is, the minimum average codeword cost tree is also the minimax tree. But this is not the case for nonexhaustive codes. Perl et al. [7] give a simple algorithm for the minimax problem as a "remark" in their paper otherwise concerned with the minimum average codeword cost case. So, as we'll see, it is the minimax version of Varn's problem which has the Fibonacci-like structure. It will also turn out that under certain conditions on the code symbol costs, minimax and minimum average codeword cost trees are identical. One nonexhaustive special case for which this is true, c(i)=i, i=1,2,..., was treated previously in the literature by Patt [6] motivated by a computer file search problem.

II. CONSTRUCTING NONEXHAUSTIVE TREES RECURSIVELY

As in the exhaustive case, the kth tree in our sequence, T(k), will have T(k-c(i)) as its ith leftmost subtree. However now the initialization will be T(1)=...=T(c(2)) each consisting of a single root node. One example is given in this abstract. The costs are c(1)=2, c(2)=c(3)=3, c(4)=5. The trees are described by labeling leaf nodes with their costs, listing them in left to right order with sibling nodes separated by + signs, and using parentheses to indicate depth in the tree from the root T(1)=T(2)=T(3)=0,T(4)=(2+3+3), T(5)=(2+3+3),T(6)=((4+5+5)+3+3+5),T(7)=((4+5+5)+(5+6+6)+(5+6+6)+ 5),.... It is not hard to give recurrence expressions for the number of leaves in the kth tree and for its unnormalized cost. These can be solved by the method of generating functions.

III. PROOF OF MINIMAX OPTIMALITY

The idea of the proof that the trees constructed in the previous section are Varn minimax trees is outlined here. First it is shown that the nonexhaustive generalized Fibonacci trees are the same as the exhaustive generalized Fibonacci trees with a certain number of highest cost leaves removed. This can be done by induction. Then we make use of an argument, like Varn's for the minimum average cost case, that optimal minimax nonexhaustive code trees are obtained by deleting highest cost leaves from a particular "correct" optimal exhaustive tree while maintaining the same number of interior nodes. The hard part is to show that the exhaustive generalized Fibonacci tree in its sequence beginning with T(c(r)+1) is the "correct" tree for the corresponding nonexhaustive generalized Fibonacci tree in its sequence beginning with T(c(2)+1). To do this we need to show that if we started with any other optimal exhaustive tree and deleted the appropriate number of highest cost leaves, we would either get a more costly tree in the minimax sense or would have to remove leaves in such a way as to leave what was an interior node childless. The details of this demonstration are omitted in this abstract for conciseness.

IV. WHEN ARE MINIMAX CODE TREES MINIMUM AVERAGE COST?

The algorithm of Perl et al. [7] for minimum average cost trees involves two stages, extension and mending, and their algorithm for minimax trees is a variant of the extension stage. They also give sufficient conditions on the code symbol costs for the mending stage to be unnecessary in the minimum average cost problem. Thus, whenever any of these sufficient conditions is satisfied by the costs, and the costs are such that the variant of the extension algorithm for minimax codes and the original extension algorithm for minimum average cost codes both yield the same tree, minimax and minimum average cost code trees are the same and the minimum average cost tree sequence shares the nice recursive structure of the minimax tree sequence. Patt's [6] costs are one such example, and his paper includes the recursive tree sequence structure.

V. CONCLUSION

The highly structured recursively constructed subsequence of optimal exhaustive code trees has been extended to the nonexhausive case by focusing on the minimax optimality criterion instead of Varn's original minimum average codeword cost criterion. When these two criteria give the same sequence of code trees, as they do under certain conditions on the costs, omitted here for conciseness, the recursive structure applies to Varn's original problem as well.

REFERENCES

1. J. Abrahams, "Varn codes and generalized Fibonacci trees," Fibonacci Quarterly, Feb. 1995.
2. D.K. Chang, "On Fibonacci k-ary trees," Fibonacci Quarterly, 24, 258-262, 1986.
3. N. Cot, Characterization and Design of Optimal Prefix Codes, Stanford Electrical Eng. Dept. Thesis, 1977.
4. M.J. Golin and N. Young, "Prefix codes: equiprobable words, unequal letter costs," Proc. 21st Intl. Colloq. Automata, Languages, and Programming, Lecture Notes in Computer Science 820, Springer, 605-617, 1994.
5. Y. Horibe, "Note on Fibonacci trees and their optimality," Fibonacci Quarterly, 21, 118-128, 1983.
6. Y.N. Patt, "Variable length tree structures having minimum average search time," Comm. ACM, 12, 72-76, 1969.
7. Y. Perl, M.R. Garey, and S. Even, "Efficient generation of optimal prefix code: equiprobable words using unequal cost letters," JACM, 22, 202-214, 1975.
8. B.F. Varn, "Optimal variable length codes (arbitrary symbol cost and equal code word probabilities)," Information and Control, 19, 289-301, 1971.

Entropy reduction, ordering in sequence spaces, and semigroups of non-negative matrices

Henk D.L. Hollmann and Peter Vanroose

Philips Research Laboratories, Eindhoven, The Netherlands; Katholieke Universiteit Leuven, Heverlee, Belgium

Abstract — **Given a programmable finite-state input/output device, what program(s) maximally reduces the "diversity" of the possible output sequences of the device? This question is made precise, and a method is developed to determine this minimum achievable diversity.**

I. Introduction

In this paper, a *(time-invariant) finite-state entropy-reduction algorithm*, or briefly, an *algorithm*, is a synchronous finite-state input/output device (see e.g. [1]); such a device takes inputs from a given source alphabet $\mathbf{B}$, and, depending on the input symbol and its internal state, produces an output symbol in an output alphabet $\mathbf{F}$ and moves to a new internal state.

In the context of channel codes (modulation codes), such a device is usually referred to as a (synchronous) finite-state encoder [1] and is used to translate or encode arbitrary sequences of source symbols into sequences that have certain desirable properties. In that context, it is of course required that decoding is possible.

Here, we think of such devices as performing some sort of data compression on sequences, and we are interested in algorithms that have a "small" output space. A natural measure of the efficiency of an algorithm is thus the *topological entropy* of the output space, which measures the growth rate of the number of output sequences of a given length. (In the context of channel codes, the entropy is usually called the *(Shannon) capacity* [1].) We allow the case where the input sequences are restricted to a given constrained system over $\mathbf{B}$.

The use of the term "data compression" may cause confusion, since we do not consider the reconstruction problem. Instead, it might be better to speak of *entropy-reduction*: the algorithm transforms data sequences and the efficiency of the entropy-reduction is measured by the number of distinct output sequences that can be produced by the algorithm.

Now let $\mathcal{F}$ be a finite collection of algorithms, all having the same source alphabet and sharing a common set of internal states. A *time-varying (entropy-reduction) algorithm over* $\mathcal{F}$ is a sequence

$$\mathbf{f} = f_1, f_2, f_3, \cdots$$

of algorithms $f_t \in \mathcal{F}$. We think of such a sequence $\mathbf{f}$ as an algorithm whose action at time t is directed by algorithm f_t. So at time t, $t = 1, 2, \ldots$, the algorithm $\mathbf{f}$ takes an input from the source alphabet and, depending on this input and its internal state, produces an output and moves to another internal state according to algorithm f_t. The collection of all entropy-reduction algorithms over $\mathcal{F}$ (all sequences over $\mathcal{F}$) will be denoted by $\mathcal{F}^\infty$.

Interestingly, it may happen that some time-varying algorithm in $\mathcal{F}^\infty$ performs better than the best algorithm in $\mathcal{F}$. (This will be shown by some examples.) So the question now arises how to produce *lower bounds* for the efficiency of algorithms in $\mathcal{F}^\infty$ and how to find the best time-varying algorithm in $\mathcal{F}^\infty$. We will refer to this problem as the *optimal entropy-reduction problem* for $\mathcal{F}$.

The motivation to investigate time-varying entropy-reduction stems from a problem in [2] and [3] on ordering in sequence spaces, a subject introduced in [4] to study certain types of organization processes. We will outline these ordering problems, and we will show that they may be considered as special instances of the optimal entropy-reduction problem considered here.

We show how the optimal entropy-reduction problem can be transformed into a problem for a related finitely-generated semigroup of non-negative matrices. Briefly stated, we will show that with each algorithm f in $\mathcal{F}$ we may associate a non-negative matrix D_f such that the efficiency of an algorithm $\mathbf{f} = f_1, f_2, \ldots$ in $\mathcal{F}^\infty$ is measured by the number

$$\mu(\mathbf{f}) = \limsup_{t\to\infty} \lambda(D_{f_1} D_{f_2} \cdots D_{f_t})^{1/t},$$

where $\lambda(D)$ denotes the largest real (Perron-Frobenius) eigenvalue of a non-negative matrix D. The number

$$\mu(\mathcal{F}) = \liminf_{\mathbf{f}\in\mathcal{F}^\infty} \mu(\mathbf{f}),$$

which can be thought of as the minimum growth rate of matrix products in the semigroup generated by the matrices D_f, $f \in \mathcal{F}$, then provides the solution to the optimal entropy-reduction problem.

We then investigate this semigroup problem. We present a method to obtain lower bounds for the optimal efficiency $\mu(\mathcal{F})$. In fact, we conjecture that in many cases our method will be able to determine the *exact* value of $\mu(\mathcal{F})$. Our results generalise (part of) Perron-Frobenius theory for non-negative matrices to *semigroups* of such matrices.

Later, we return to the ordering problem. We will use our method to determine $\tau_2(0,2,1)$, the optimal efficiency of a time-varying binary ordering algorithm in the class $(0,2,1,T^+,O^-)$ [4]. We will show that $\tau_2(0,2,1) = \frac{1}{3}^2\log(2+\sqrt{3})$, as conjectured in [2]. Our approach suggests that (at least in principle) other values of $\tau_q(\pi,\beta,\phi)$ may be computed similarly.

Finally, we discuss our results.

References

[1] B.H Marcus, P.H. Siegel, and J.K. Wolf, "Finite-state modulation codes for data storage", IEEE J-SAC, vol. 10, no. 1, pp. 5–37, Jan. 1992.

[2] P. Vanroose, "Een ordeningsresultaat voor de situatie $(0,2,1,T^+)$" (in Dutch), report, Katholieke Universiteit Leuven, September 1989.

[3] J-P. Ye, "Towards a theory of ordening in sequence spaces", thesis, Universität Bielefeld, July 1988.

[4] R. Ahlswede, J-P. Ye, and Z. Zhang, "Creating order in sequence spaces with simple machines", Inform. and Comp. 89, pp. 47-94, 1990.

Variable-to-Fixed Length Codes and the Conservation of Entropy

Serap A. Savari
Laboratory for Information and Decision Systems, MIT, Cambridge, MA 02139 USA

Abstract — **For Markov sources, we consider a generalization of variable-to-fixed length codes and find the optimal code and its performance as the dictionary size approaches infinity.**

I. Introduction

A variable-to-fixed length coder can be decomposed into a parser and a string encoder. The parser segments the source output into a concatenation of variable-length strings, each of which belongs to a dictionary with M entries. The string encoder maps each dictionary entry into a fixed-length codeword. Variable-to-fixed length codes can take advantage of the memory of the source when the dictionary entries are roughly equiprobable. Furthermore, the Lempel-Ziv codes can be viewed as universal variable-to-fixed length codes.

II. Problem Formulation

A Markov source with finite alphabet $\{0,\ldots,K-1\}$ and set of states $\{0,\ldots,R-1\}$ is defined by specifying, for each state s and letter j,

1. the probability $p_{s,j}$ that the source emits j from state s
2. the next state $S[s,j]$ after j is issued from state s.

Given any initial state s_0, these rules inductively specify both the probability $P(\sigma|s_0)$ that any given source string σ is emitted and the resulting state $S[s_0,\sigma]$ after σ is output; they also determine $\mathcal{H}$, the entropy of the source in natural units, $\mathcal{H}(s)$, the entropy in natural units of the next source symbol emitted from state s, and π_s, the long-run proportion of time that the source is in state s.

The dictionaries that we consider are *uniquely parsable*; i.e., every source sequence has exactly one prefix in the dictionary. This condition implies that $M=\alpha(K-1)+1$ for some integer α; here, α is the number of intermediate nodes in the dictionary tree, including the root.

The best variable-to-fixed length code has a dictionary that maximizes the steady-state expected number, $E[L]$, of source letters per dictionary entry. For the special case of a discrete, memoryless source, $E[L]$ is the sum of the probabilities associated with each intermediate node in the tree, including the root. For more general Markov sources, it is considerably harder to evaluate $E[L]$ since the probability of a dictionary entry, starting at a parsing point, depends on the state probabilities at parsing points, which in turn depend on the dictionary itself.

To gain insights into codes for Markov sources, we will consider a broader class of codes in which there is a uniquely parsable dictionary $\mathcal{D}_s$ of size M associated with *each* state s. For these codes, the parser determines the source state s after each parsing point and then uses $\mathcal{D}_s$ to find the next parsed string. We would like to find a good way to design the dictionaries $\mathcal{D}_s$. Let $\mathcal{L}_s$ represent the expected length of a dictionary entry for $\mathcal{D}_s$; then, for all s,

$$\mathcal{L}_s = \sum_{\text{intermediate nodes } \sigma \text{ for } \mathcal{D}_s} P(\sigma|s).$$

In [1], $\mathcal{D}_s$ was chosen to maximize $\mathcal{L}_s$ for each state s. This code, called the *generalized Tunstall code*, maximizes the expected number of source symbols per parse for each state, but does not necessarily lead to good parsing probabilities.

III. New Contributions

There is another way to address the problem of selecting the dictionaries $\{\mathcal{D}_s\}$. Let H_s denote the entropy of the entries in $\mathcal{D}_s$ and let H represent the steady-state average self-information between successive parsing points. We have

Theorem 1 $H = \mathcal{H}\cdot E[L]$.

For memoryless sources, this "conservation of entropy" theorem was established for codes with one uniquely parsable dictionary in [2]; our proof indicates that it applies for a much larger set of codes than the ones we discuss here. Theorem 1 suggests that we may get good dictionaries by maximizing H_s for each s. The "leaf entropy" theorem of [3] implies that

$$H_s = \sum_{\text{intermediate nodes } \sigma \text{ for } \mathcal{D}_s} P(\sigma|s)\mathcal{H}(S[s,\sigma]),\ 0\le s\le R-1;$$

the expressions for $\mathcal{L}_s$ and H_s suggest that we should consider dictionaries that maximize

$$X(s) = \sum_{\text{intermediate nodes } \sigma \text{ for } \mathcal{D}_s} P(\sigma|s)x_{S[s,\sigma]}$$

for some choice of $\mathbf{x}=\{x_0,\ldots,x_{R-1}\}$. x_i is called the *weight* of state i and $P(\sigma|s)x_{S[s,\sigma]}$ is the *state s return* of string σ. A desirable feature of the generalized Tunstall code is that it can be constructed in a *greedy* manner; i.e., for each state s and string σ, the state s return of σ is at most the state s return of any proper prefix of σ, so the nodes with the largest state s return can be selected one by one starting with the null string. A necessary and sufficient condition for a greedy construction is that the weight vector $\mathbf{x}$ is in the set

$$\mathcal{G} = \{\mathbf{x}=(x_0,\ldots,x_{R-1}):\ \mathbf{x}>\mathbf{0},\ p_{i,j}x_{S[i,j]}\le x_i,\ \forall\, i,\ j\}.$$

We have the following results.

Theorem 2 *Let* $f(\mathbf{x}) = \sum_{i=0}^{R-1}\pi_i\mathcal{H}(i)\ln(\sum_r \pi_r x_r/x_i)$ *and* $C=\mathcal{H}\ln((K-1)/\mathcal{H})-\sum_{i=0}^{R-1}\sum_{j=0}^{K-1}\pi_i p_{i,j}(-\ln p_{i,j})^2/2$. *The weight vector* $\mathbf{x}^*=(x_0^*,\ldots,x_{R-1}^*)$ *of the asymptotically best greedy code is given by* $\mathbf{x}^*=\arg\min_{\mathbf{x}\in\mathcal{G}} f(\mathbf{x}^*)$; *the asymptotic compression achieved by this code satisfies*

$$(\ln M)\cdot\left(\frac{\ln M}{E[L]}-\mathcal{H}\right)\longrightarrow C+f(\mathbf{x}^*).$$

Corollary 1 *If* $p_{i,j}\mathcal{H}(S[i,j])\le\mathcal{H}(i)$ *for all* i *and* j, *then the greedy code with weight vector* $(\mathcal{H}(0),\ldots,\mathcal{H}(R-1))$ *is asymptotically the best generalized variable-to-fixed length code.*

References

[1] S. A. Savari and R. G. Gallager, "Variable-to-fixed length codes for sources with memory," I.S.I.T., 1994.

[2] F. Jelinek and K. S. Schneider, "On variable-length-to-block coding," *I.E.E.E. Trans. Inform. Theory* IT-18, 765-774, 1972.

[3] J. L. Massey, "The entropy of a rooted tree with probabilities," I.S.I.T., 1983.

An Inequality On Entropy

Robert J. McEliece Zhong Yu[1]

Dept. of Electrical Engineering, Mail Code 116-81, California Institute of Technology, Pasadena, CA 91125

Abstract — **The entropy $H(X)$ of a discrete random variable X of alphabet size m is always non-negative and upper-bounded by $\log m$. In this paper, we present a theorem which gives a non-trivial lower bound for $H(X)$. We will show that for any discrete random variable X with range $R = \{x_0, \ldots, x_{m-1}\}$, if $p_i = Pr\{X = x_i\}$ and $p_0 \geq p_1 \geq \ldots p_{m-1}$, then**

$$H(X) \geq \frac{2\log m}{m-1} \sum_{i=0}^{m-1} i p_i, \tag{1}$$

with equality iff
(i) X is uniformly distributed, i.e., $p_i = \frac{1}{m}$ for all i, or trivially
(ii) $p_0 = 1$, and $p_i = 0$ for $1 \leq i \leq m-1$.

I. Introduction

For a discrete random variable X with range $R = \{x_0, \ldots, x_{m-1}\}$ with $p_i = Pr\{X = x_i\}$ for $0 \leq i \leq m-1$, the entropy of the random variable is defined as [1]

$$H(X) = -\sum_{i=0}^{m-1} p_i \log p_i. \tag{2}$$

The upper bound, $H(X) \leq \log m$, with equality iff the random variable X is uniformly distributed, is well known. In this paper, we will prove a theorem that gives a useful lower bound on entropy for finitely valued discrete random variables. In [2], an upper bound for a constrained entropy of infinitely valued discrete random varibles is shown under certain condition. Our theorem provides a lower bound for the constrained entropy.

II. The Theorem

Let us define a convex region K_m and a set A:

$$K_m = \{(x_1, \ldots, x_m) : 1 \geq x_i \geq x_j \geq 0, i \leq j, \textstyle\sum_{i=1}^m x_i = 1\},$$
$$A = \{\bar{a}_1, \ldots, \bar{a}_m\},\ \bar{a}_i = (\underbrace{\tfrac{1}{i}, \ldots, \tfrac{1}{i}}_{i}, 0, \ldots, 0),\ 1 \leq i \leq m.$$

Clearly, $A \subset K_m$. The following lemmas can be shown:
Lemma 1. K_m is the convex hull of A.
Lemma 2. If $f(\bar{x})$ is convex $\cup$ on K_m, then

$$f(\bar{x}) \leq \max\{f(\bar{a}_1), \ldots, f(\bar{a}_m)\}. \tag{3}$$

If $f(\bar{x})$ is strictly convex, then equality holds iff $\bar{x} = \bar{a}_i$ for some i such that $f(\bar{a}_i)$ is the maximum among all $f(\bar{a}_1), \ldots, f(\bar{a}_m)$. Consider the function

$$f(\bar{x}) = f(x_1, \ldots, x_m) = \sum_{i=1}^{m} (\log x_i + \frac{2(i-1)}{m-1} \log m) x_i, \tag{4}$$

defined on the convex region K_m. Since $\frac{\partial^2 f(\bar{x})}{\partial^2 x_i} = \frac{1}{x_i} > 0$ for $1 \leq i \leq m$, $f(\bar{x})$ is a strictly convex $\cup$ function on K_m. We can show $f(\bar{x}) \leq 0$, which implies our theorem:
Theorem. For any discrete random variable X with range $R = \{x_0, \ldots, x_{m-1}\}$, if $p_i = Pr\{X = x_i\}$ and $p_0 \geq p_1 \geq \ldots p_{m-1}$, i.e., p_i are in a non-increasing order, then

$$H(X) \geq \frac{2\log m}{m-1} \sum_{i=0}^{m-1} i p_i, \tag{5}$$

with equality iff
(i) X is uniformly distributed, i.e., $p_i = \frac{1}{m}$ for all i, or
(ii) $p_0 = 1$, and $p_i = 0$ for $1 \leq i \leq m-1$.

III. Examples

We can compute lower bounds for two specific examples using the above theorem.
(1) Geometrical Distribution:

$$H(X) \geq \frac{2\log m}{m-1} \cdot (1 - \frac{1}{2^{m-1}}) = \frac{2\log m}{m-1} + o(1). \tag{6}$$

(2) Binomial Distribution:

$$H(X) \geq \log m \cdot (\frac{\binom{m-2}{\lfloor \frac{m-1}{2} \rfloor}}{2^{m-3}} - \frac{1}{m-1}) = \sqrt{\frac{8}{\pi}} \frac{\log m}{\sqrt{m}} + o(1). \tag{7}$$

IV. Remark

Let us define

$$H(m, \alpha) = \{-\sum_{i=1}^{m} p_i \log p_i : \sum_{i=1}^{m} i \cdot p_i = \alpha, p_j \geq p_k, j \leq k\}. \tag{8}$$

$\alpha = \sum_{i=1}^m i \cdot p_i$ can be viewed as the average number of guesses with an optimum strategy needed to guess the value of a random variable X [2]. Let $H_L(m, \alpha) = \min H(m, \alpha)$ and $H_U(m, \alpha) = \max H(m, \alpha)$. Clearly, $H_U(m, \alpha)$ is a monotonically increasing function of m. An upper bound on $H(m, \alpha)$ is [2]

$$H_U(m, \alpha) \leq \lim_{m \to \infty} H_U(m, \alpha) \leq \log(\alpha - 1) + 1, \tag{9}$$

when $\alpha \geq 2$. From our theorem, we can provide a lower bound on $H(m, \alpha)$:

$$H_L(m, \alpha) \geq \frac{2\log m}{m-1}(\alpha - 1). \tag{10}$$

References

[1] R. J. McEliece, The Theory of Information and Coding, Encyclopedia of Mathematics and Its Applications, Volume 3, Cambridge University Press, 1985.

[2] J. L. Massey, "Guessing and Entropy", Proceedings ISIT94, p. 204.

[1]This work was support by a Grant from Pacific Bell

On entropies, divergences, and mean values

Michèle Basseville and Jean-François Cardoso[1]

IRISA/CNRS, Campus de Beaulieu, 35042 Rennes Cedex, France - E-mail: basseville@irisa.fr,
and Télécom Paris/CNRS, 46 rue Barrault, 75634 Paris Cedex 13, France - E-mail: cardoso@sig.enst.fr.

Abstract — **Two entropy-based divergence classes are compared using the associated quadratic differential metrics, mean values and projections.**

I. Two classes of divergences

The design concepts of divergences are of interest because of the key role they play in statistical inference and signal processing. Most of the existing divergences $\mathbf{D}$ between two probability distributions may be associated with an integral or non integral entropy functional $\mathbf{H}_\nu(\mu)$ with respect to some reference measure ν. We distinguish two different classes of divergences built on entropies. The first one is the well known class of f-divergences $\mathbf{I}_f$ [4] which are based on the likelihood ratio and are formally identical to the above entropies $\mathbf{I}(\mu,\nu) \triangleq -\mathbf{H}_\nu(\mu)$. In the integral case, this yields the relative entropy class, which includes Kullback information as its most prominent member [4]. The most important instance of non integral f-divergences is Rényi information [13].

The second class of divergences builds upon the concavity of an entropy functional, which entails that, for $0 < \alpha < 1$, $\mathbf{J}_H^{(\alpha)}(\mu,\nu) \triangleq \mathbf{H}((1-\alpha)\mu+\alpha\nu) - (1-\alpha)\mathbf{H}(\mu) - \alpha\mathbf{H}(\nu)$ is positive. One can then construct $\mathbf{C}_H(\mu,\nu) = \max_\alpha \mathbf{J}_H^{(\alpha)}(\mu,\nu)$, a Jensen difference $\mathbf{J}_H(\mu,\nu) = \mathbf{J}_H^{(1/2)}(\mu,\nu)$, and a Bregman distance $\mathbf{D}_H(\mu,\nu) = \lim_{\alpha\to 0} \alpha^{-1}\mathbf{J}_H^{(\alpha)}(\mu,\nu)$. Bregman distances enjoy an Euclidian-like property, similar to the Pythagorean theorem [9, 7], when involved in projections onto exponential or mixture families. This may be further generalized to projections onto 'α-families' as shown in [2] where families of distributions are dealt with as differential manifolds. Still in this geometrical vein, the interplay between $\mathbf{C}_H, \mathbf{J}_H$ and $\mathbf{D}_H$ can be understood via Thales theorem.

A local quadratic differential metric is associated with any divergence measure [12, 2]. Based on the fact that f-divergences are locally equivalent to the Riemannian metric defined by the Fisher information matrix, we characterize the intersection of the two above divergence classes. In particular, it is easily found that the only Bregman distance $\mathbf{D}_H$ which is a f-divergence is Kullback information [9, 7]. Similarly, it is found that the only f-divergences which can be written as a Jensen difference $\mathbf{J}_H^{(\alpha)}$ are those introduced in [11, 10].

II. Associated mean values and projections

Mean values can be associated with entropy-based divergences in two different ways. The first way [13, 1] consists in writing explicitly the *generalized mean values* $\phi^{-1}\left(\sum_{i=1}^n \beta_i\, \phi(p_i)\right)$ underlying integral and non integral f-divergences. Here the β's are normalized positive weights. For Rényi information, $\phi(u) = u^\alpha$, and this results in *α-mean values* $\left(\sum_{i=1}^n \beta_i\, p_i^\alpha\right)^{1/\alpha}$.

The second way [3] consists in defining mean values by $\arg\min_v \sum_{i=1}^n \beta_i\, d(v,u_i)$, namely as *projections*, in the sense of distance d, onto the half-line $u_1 = \ldots = u_n > 0$ [7]. When d is an integral f-divergence $d(v,u_i) = u_i f\left(\frac{v}{u_i}\right)$, this gives the *entropic means* [3], which are characterized implicitly by $\sum_{i=1}^n \beta_i\, f'\left(\frac{v}{u_i}\right) = 0$, and necessarily homogeneous (scale invariant). The class of entropic means includes all available integral means and, when applied to a random variable, contains most of centrality measures (moments, quantiles). When d is a Bregman distance $d_h(u,v) = h(u) - h(v) - (u-v)h'(v)$, the corresponding mean values are exactly the above generalized mean values (for $\phi = h'$), which are generally not homogeneous.

The only generalized mean value which is also an entropic mean, and thus both an f-divergence-projection and a Bregman-projection, is the above α-mean value, corresponding to Rényi information [3]. This agrees with invariant properties of means [8] and the axiomatic of inference in [5].

Finally, we mention that mutual information (viewed both as relative entropy and Jensen difference) and the related concepts of channel capacity [6] and information radius [14], can be seen as another manner of investigating the intersection of the above two divergence classes.

[1] The authors are also with GDR CNRS no 134 'Traitement du Signal et Images'.

References

[1] J. Aczél and Z. Daròczy, "*On Measures of Information and Their Characterizations*," Academic Press, 1975.

[2] S-I. Amari, "*Differential-Geometrical Methods in Statistics*," Lecture Notes in Statistics, vol.28, Springer-Verlag, 1985.

[3] A. Ben-Tal, A. Charnes and M. Teboulle, "Entropic means," *Jal Math. Anal. Appl.*, vol.139, pp.537-551, 1989.

[4] I. Csiszár, "Information measures: a critical survey," *7th Prague Conf. Inf.Th., Stat.Dec.Funct. and Rand.Proc.*, pp.73-86, 1974.

[5] I. Csiszár, "Why least-squares and maximum entropy ? An axiomatic approach to inference for linear inverse problems," *Annals Statistics*, vol.19, pp.2032-2066, Jul. 1991.

[6] I. Csiszár, "Generalized cutoff rates and Rinyi's information measures," *IEEE Trans. Information Theory*, vol.IT-40, pp.26-34, Jan. 1995.

[7] I. Csiszár, "Generalized projections for non-negative functions," *Acta Mathematica Hungarica*, vol.68, pp.161-185, Jan. 1995.

[8] G.H. Hardy, J.E. Littlewood and G. Polya, "*Inequalities*," Cambridge Univ. Press, 1952.

[9] L.K. Jones and C.L. Byrne, "General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis," *IEEE Trans. Inform. Theory*, vol.IT-36, pp.23-30, Jan. 1990.

[10] L. Knockaert, "A class of statistical and spectral distance measures based on Bose-Einstein statistics," *IEEE Trans. Signal Proc.*, vol.SP-41, pp.3171-3174, Nov. 1993.

[11] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inform. Theory*, vol.IT-37, pp.145-151, Jan. 1991.

[12] C.R. Rao, "Differential metrics in probability spaces," in *IMS Lecture Notes*, vol.10, S. Gupta (ed.), pp.217-240, 1987.

[13] A. Rényi, "On measures of entropy and information," *4th Berkeley Symp. Math. Stat. Proba.*, vol.I, pp.547-561, 1961.

[14] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitsth. Werw. Gebiete*, vol.14, pp.149-160, 1969.

When Are the MLSD Respectively the Matched Filter Receiver *Optimal* with Respect to the BER ?

Per Ödling[1] Håkan B. Eriksson[1] Timo Koski[2] Per Ola Börjesson[1]

[1] Division of Signal Processing, Luleå University of Technology, S-971 87 Luleå, Sweden

[2] Department of Mathematics, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Abstract—**We reconsider the minimum/optimal bit-error probability receiver (OBER) for intersymbol interference channels with Gaussian noise and the reception of finite blocks of bits. We view the OBER as a function with two inputs: the received sequence and an expected signal-to-noise ratio; and one output: the estimated block of bits. Assuming that all sequences are equally probable to be transmitted we prove two results about the behaviour of the OBER. We show that the OBER coincides with the maximum likelihood sequence detector when designed for high signal-to-noise ratios and that it collapses to a matched filter followed by a hard-limiting device for low expected signal-to-noise ratios.**

I. A BLOCK TRANSMISSION SYSTEM MODEL

After the introduction of the Viterbi detector as a *Maximum Likelihood Sequence Detector (MLSD)* [3], the optimal, or minimum, bit-error probability receiver *(OBER)* [1], [2], [4] for intersymbol interference (ISI) channels has not been given much attention as a practical receiver. We reconsider the OBER for *block transmission systems*, to gain insight to its properties and its relation to the MLSD.

Consider the transmission of blocks of binary data through a channel with known ISI and additive Gaussian noise at the receiver. Let the vector $\mathbf{b} \in \{-1,+1\}^N$ denote the block of independent bits to be transmitted. We represent the transmission system in matrix notation as:

$$\mathbf{y} = \mathbf{H}\mathbf{b} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ is a deterministic and known matrix representing the ISI, the noise $\mathbf{n} \in N(\mathbf{0}, \sigma_{\mathbf{n}}^2\mathbf{I})$ and $\mathbf{y}$ is the $(N+L)\times 1$ stochastic vector observed by the receiver. Further, let $\boldsymbol{\eta}$ denote the outcome of $\mathbf{y}$.

II. THE OPTIMAL BIT-ERROR PROBABILITY RECEIVER

Let us consider the detection of $(\mathbf{b})_{\text{bit } k}$. A geometric interpretation of this binary hypothesis testing is that of choosing the correct halfcube:

$$H_0: \ \mathbf{b} \in \mathcal{C}_k^+, \qquad H_1: \ \mathbf{b} \in \mathcal{C}_k^-, \tag{2}$$

where $\mathcal{C}_k^+$ and $\mathcal{C}_k^-$ are the halfcubes with $(\mathbf{b})_{\text{bit } k} = +1$ and $(\mathbf{b})_{\text{bit } k} = -1$, respectively. The Bayes decision rule minimizing the probability of detection error is given by [1], [2], [4]

$$\Lambda_k(\mathbf{y}) \triangleq \frac{f_{\mathbf{y}|H_1}(\boldsymbol{\eta}|H_1)}{f_{\mathbf{y}|H_0}(\boldsymbol{\eta}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\Pr\{H_0\}}{\Pr\{H_1\}}, \tag{3}$$

where $\Pr\{H_1\} = 1 - \Pr\{H_0\}$ is the *a priori* probability that H_1 is true, and $f_{\mathbf{y}|H_0}(\cdot)$ and $f_{\mathbf{y}|H_1}(\cdot)$ designate the probability density functions for $\mathbf{y}$ given H_0 and H_1, respectively.

Proposition 1: Let $\psi(\mathbf{y}|\boldsymbol{\beta})$ denote the conditional density of $\mathbf{y}$ given that the sequence $\boldsymbol{\beta}$ was transmitted (here multi-dimensional Gaussian). Furthermore, let

$$\widehat{\mathbf{b}}_{\text{OBER}}(\mathbf{y}) \triangleq \Gamma(\mathbf{y}, \sigma_{\mathbf{n}}) \triangleq \operatorname{sign}\left(\sum_{\boldsymbol{\beta}\in\mathcal{C}_1^+} w(\mathbf{y},\boldsymbol{\beta})\,\boldsymbol{\beta}\right) \tag{4}$$

and $w(\mathbf{y},\boldsymbol{\beta}) = \psi(\mathbf{y}|\boldsymbol{\beta})\Pr\{\mathbf{b}=\boldsymbol{\beta}\} - \psi(\mathbf{y}|-\boldsymbol{\beta})\Pr\{\mathbf{b}=-\boldsymbol{\beta}\}$, where $\Pr\{\mathbf{b}=\boldsymbol{\beta}\}$ is the probability for the sequence $\boldsymbol{\beta}\in\mathcal{C}$ being transmitted. Then $\widehat{\mathbf{b}}_{\text{OBER}}(\mathbf{y})$ is the detector of the transmitted bits that minimizes the bit-error probability. ∎

Note that (4) represents a parallel block processor structure, *simultaneously* detecting all the individual bits.

As indicated by (4), we find it instructive to view the OBER as a function $\Gamma(\mathbf{y},\alpha)$ with two inputs, $\mathbf{y}$ and α. The parameter α^2 is the variance the OBER is designed for, and controls the decision regions in $\mathbb{R}^{N+L}$ where $\mathbf{y}$ takes its values. Thus, the OBER depends on the expected SNR and is only optimal when the expected variance α^2 and the true variance $\sigma_{\mathbf{n}}^2$ agree.

III. THE BEHAVIOUR OF THE OBER

We will discuss asymptotical properties of the OBER by studying the function $\Gamma(\cdot,\cdot)$ as defined in (4). Assuming that all sequences are equally probable to be transmitted we show that

$$\lim_{\alpha\to 0}\Gamma(\mathbf{x},\alpha) = \widehat{\mathbf{b}}_{\text{MLSD}}(\mathbf{x}), \quad \text{for all } \mathbf{x}\in\mathbb{R}^{N+L}, \tag{5}$$

and that

$$\lim_{\alpha\to\infty}\Gamma(\mathbf{x},\alpha) = \operatorname{sign}(\mathbf{H}^T\mathbf{x}), \quad \text{for all } \mathbf{x}\in\mathbb{R}^{N+L}. \tag{6}$$

Equation (5) means that the OBER designed for a high SNR *becomes* the MLSD. It is because of this is that the MLSD will achieve the minimum attainable bit-error probability when used in systems with a high SNR, *cf.* [3]. In equation (6) we find a similar comparison for low SNR between the OBER and the matched filter with hard decisions. If the true SNR is low, the best possible receiver is actually the matched filter receiver as comes to the BER.

REFERENCES

[1] K. Abend, T.J. Harley and B.D. Fritchman, "On optimum receivers for channels having memory," *IEEE Trans. Inf. Th.*, vol. 14, no. 6, pp. 818–820, November 1968.

[2] R.R. Bowen, "Bayesian decision procedure for interfering digital signals," *IEEE Trans. Inf. Th.*, vol. 15, no. 4, pp. 506–507, July 1969.

[3] D.G. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Th.*, vol. 18, no. 3, pp. 363–378, May 1972.

[4] P. Ödling, H.B. Eriksson, T. Koski and P.O. Börjesson, "Representations for the minimum bit-error probability receiver for block transmission systems with intersymbol interference channels," Research Report 1995:11, Luleå University of Technology, Luleå, Sweden, April 1995.

A Genie-Aided Detector with a Probabilistic Description of the Side Information

Håkan B. Eriksson[1] Per Ödling[1] Timo Koski[2] Per Ola Börjesson[1]

[1] Division of Signal Processing, Luleå University of Technology, S-971 87 Luleå, Sweden

[2] Department of Mathematics, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Abstract— **Building on Forney's concept of the genie [4], [5], and introducing the idea of an *explicit statistical description* of the side information provided to the genie-aided detector, we develop a generic tool for derivation of lower bounds on the bit-error rate of any actual receiver [3]. With this approach, the side information statistics become *design parameters*, which may be chosen to give the resulting bound a desired structure. To illustrate this, we choose statistics in order to obtain a special case: the lower bound derived by Mazo [6]. The statistical description of the side information makes the lower bounding a transparent application of Bayesian theory.**

I. Introduction

The idea of a *good genie* with a corresponding *genie-aided detector (GAD)* has, in particular, often been used to determine a lower performance bound for the probability of bit-error [1], [2], [3], [4], [5]. The GAD has access to more information than any actual detector: it has access to the side information supplied by the genie and is expected to handle all information *optimally*. Because of this, it is argued, it cannot have a worse performance than any detector working without the side information. However, in order that optimal processing of the side information be well-defined in the sense of Bayesian detection theory, an explicit (statistical) description of the side information is required. This paper introduces such a representation of the genie, augmenting the foundational ideas of Forney's work in [4], [5]. Our aim is to introduce the side information supplied by the genie as the output of a "side information channel" parallel to the original channel and governed by a *probabilistic rule* with free parameters.

II. The Side Information Channel

Consider a transmission system where binary data is sent through a discrete-time, additive Gaussian channel with intersymbol interference, and where additional side information is carried to the detector through a parallel channel (representing the genie).

We discuss the detection of bit number k in the important special case when the side information consists of a pair of sequences and one of the sequences is equal to the transmitted sequence, *cf.* [4], [5]. Define $\mathcal{C}_k^+$ and $\mathcal{C}_k^-$ as the sets of sequences with the bit in position k as $+1$ and -1, respectively, and denote the side information with z and its outcome with ζ. Let ζ consist of pairs in $\mathcal{C}_k^+ \times \mathcal{C}_k^-$ of the form $\zeta_{i,j} \triangleq (\boldsymbol{\beta}_i^+, \boldsymbol{\beta}_j^-)$, for $1 \le i,j \le 2^{N-1}$. With the transmitted sequence being $\boldsymbol{\beta}$, let the additional sequence be chosen at random among the sequences differing from $\boldsymbol{\beta}$ in bit k, according to the *known, probabilistic* transition rule:

$$\Pr\{z = \zeta_{i,j} | \mathbf{b} = \boldsymbol{\beta}\} = \begin{cases} p(j|i) & \text{if } \boldsymbol{\beta} = \boldsymbol{\beta}_i^+ \in \mathcal{C}_k^+ \\ q(i|j) & \text{if } \boldsymbol{\beta} = \boldsymbol{\beta}_j^- \in \mathcal{C}_k^- \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the properties of the genie, or equivalently, the properties of the output of the side information channel, are defined by the statistics (or transition probabilities) $p(j|i)$ and $q(i|j)$.

III. The Genie-Aided Detector

With the complete statistical description of the transmission system, including a set of transition probabilities, the GAD with minimum bit-error probability is derived in terms of a binary Bayesian hypothesis test. By evaluating the performance of this GAD, a lower bound on the probability of bit-error of any detector, with or without access to the side information, is obtained as

$$\begin{aligned} P_{\text{BER},k} \;\ge\; & \textstyle\sum_{i,j} Q\left(\frac{d_{i,j}}{2} + \frac{\ln \gamma_{i,j}}{d_{i,j}}\right) q(i|j) \Pr\{\mathbf{b} = \boldsymbol{\beta}_j^-\} + \\ & \textstyle\sum_{i,j} Q\left(\frac{d_{i,j}}{2} - \frac{\ln \gamma_{i,j}}{d_{i,j}}\right) p(j|i) \Pr\{\mathbf{b} = \boldsymbol{\beta}_i^+\}, \end{aligned}$$

where $d_{i,j}$ is the Euclidian distance between $\boldsymbol{\beta}_i^+$ and $\boldsymbol{\beta}_j^-$,

$$\gamma_{i,j} = \frac{q(i|j) \Pr\{\mathbf{b} = \boldsymbol{\beta}_j^-\}}{p(j|i) \Pr\{\mathbf{b} = \boldsymbol{\beta}_i^+\}}$$

and $Q(x) \triangleq (1/\sqrt{2\pi}) \int_x^{+\infty} e^{-t^2/2} dt$.

The transition probabilities $\{p(j|i), q(i,j)\}$ are free parameters which can be chosen to optimize the performance of the GAD. They might for example be chosen to make the corresponding bound tight, or to give the bound a simple structure. We choose several sets of transition probabilities as examples in order to discuss the properties of their respective performance bounds. In this, we also discuss the relation of the attainable performance bounds to the works by Forney [4], [5] and Mazo [6].

References

[1] T. Aulin, "Genie-aided sequence detection under mismatch," In *Proc. ISITA '94*, pp. 869–873, Sydney, Australia, Nov. 1994.

[2] R.E. Blahut, *Digital Transmission of Information.* New York: Addison-Wesley Publ. Comp., 1990.

[3] H.B. Eriksson, P. Ödling, T. Koski and P.O. Börjesson "A genie-aided detector based on a probabilistic description of the side information," Subm. to *IEEE Trans. Inf. Th.*, May 1995.

[4] G.D. Forney, "Lower bounds on error probability in the presence of large intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 1, pp. 76–77, Feb. 1972.

[5] G.D. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Th.*, vol. 18, no. 3, pp. 363–378, May 1972.

[6] J.E. Mazo, "Faster-than-Nyquist-signaling," *The Bell System Technical Journal*, vol. 54, Oct. 1975, pp. 1451–1462.

A New Family of Decision Delay-Constrained MAP Decoders for Data Transmission over Noisy Channels with ISI and Soft-Decision Demodulation

E. Baccarelli, R. Cusani, G. Di Blasio

INFOCOM Dpt., University of Rome 'La Sapienza', Via Eudossiana 18, 00184 Roma, Italy

Abstract **- A new family of nonlinear decision delay-constrained receivers minimizing the symbol-decoding-error probability of QAM- or PSK-modulated digital information sequences transmitted over time-dispersive time-varying noisy waveform channels is presented. New (generally) time-varying Bhattacharyya-type upper bounds for the performance evaluation of the proposed receivers are also presented.**

SUMMARY

In this work a novel solution for the optimal synthesis of nonlinear receivers for the detection of digitally modulated (QAM or PSK) information sequences transmitted over generally time-varying channels impaired by known ISI and additive noise is presented for the case when the decoding-decision-delay Δ is limited and finite. Receivers which minimize the symbol error probability (i.e., symbol-by-symbol MAP decoders) are considered.

The known solution presented in [1] for a similar problem holds for the case of data transmission systems with time-invarying waveform channels and unquantized soft-decision demodulation. Moreover, the algorithm in [1] has been obtained by means of a direct application of Bayes' rule so that the resulting receiver complexity grows exponentially with the decision-delay Δ; as a consequence, the implementation of such receivers for multilevel digital signalling seems to be unattractive even for small values of Δ [4, Sect.6.6].

In this work a M-level quantizer is assumed present at the output of the noisy waveform channel so that the finite word-length effects of digital receivers can be suitably taken into account. Moreover, the approach followed to synthesize the MAP decoder is completely different from that in [1] and is based on the modeling of the ISI channel as a sequential Moore-type finite-state-machine. This allows to adopt the recursive Kalman-like algorithms of [2] for the computation of the sequence $\{\pi(k \mid k+\Delta),\ k \geq 1\}$ of the A Posteriori Probabilities (APPs) of the so-called "channel state" Markov chain $\{x(k) \in \mathcal{U} = \{u_1, u_2, \dots, u_N\},\ k \geq 0\}$ (defined as in [3,Sect.II]) for every assigned decision-delay Δ. The main advantage of this approach is that the implementation of the resulting MAP decoders exhibits a complexity which grows only linearly with the value assumed by the decision-delay Δ. In fact, the following expression for the computation of the APP sequence (recursive with respect to the k-index) holds:

$$\pi(k \mid k+\Delta) = A\,\pi(k-1 \mid k+\Delta-1) + \sum_{m=k}^{k+\Delta} \beta(k;m)\Theta(m). \qquad (1)$$

In (1) the matrix A is the probability transition matrix of the Markov chain $\{x(k)\}$ and the sequences $\{\beta(k;m)\}$ and $\{\Theta(m)\}$ can be recursively calculated as in [2]. Starting from the above APPs sequence, the corresponding MAP estimate sequence $\{\hat{a}_{MAP}(k|k+\Delta) \in \mathcal{A}\}$ of the transmitted S-ary information sequence $\{a(k) \in \mathcal{A} = \{a_1, a_2, \dots, a_S\},\ k \geq 0\}$ can be easily computed following quite standard procedures (see, for example, [3,Sect.IV]).

As far as the performance evaluation of the mentioned nonlinear MAP decoders is concerned we observe that, from the authors' knowledge, *no explicit* analytical expressions are known in literature (see [4, Sect.6.6]). Starting from Eq.(1), new (generally) time-varying Bhattacharyya-type upper bounds for the performance evaluation of the proposed MAP decoders have been derived as follows:

$$P(\hat{a}_{MAP}(k|k+\Delta) \neq a(k)) \leq$$

$$\frac{1}{S}\sum_{j=1}^{S}\sum_{\substack{r=1 \\ r\neq j}}^{S}\left\{\sum_{m=1}^{M^{k+\Delta+1}}\sqrt{P\left(Y_0^{k+\Delta}=y_0^{k+\Delta}(m)\middle|a(k)=a_r\right)P\left(Y_0^{k+\Delta}=y_0^{k+\Delta}(m)\middle|a(k)=a_j\right)}\right\}, \qquad (2)$$

where $y_0^{k+\Delta}(m)$ is the m-th determination assumed by the ordered random sequence $Y_0^{k+\Delta}$, constituted by the quantized noisy data received at the channel output from step 0 to step $k+\Delta$. Simulation results proved that the upper bounds of Eq.(2) are quite tight for error probabilities below 10^{-2}.

The performance of the proposed symbol-by-symbol MAP receivers have been compared to that pertaining to the conventional sequence Maximum Likelihood (ML) receivers (based on the classic Viterbi Algorithm with optimized branch metric). Computer simulations showed that the performance of the presented receivers overcomes that of the ML sequence receivers when the transmission channel is largely time-dispersive and the signal-to-noise ratio (SNR) at the receiver site is quite low, so that the proposed decoders could be attractive, in particular, for HF channel equalization. Moreover, for the MAP decoders at hand a decision-delay Δ of the order of the length $\mathcal{L}$ of the channel impulse response (measured as multiples of the signalling period T) results in a negligible performance loss with respect to the ideal case $\Delta = \infty$, while a delay Δ of 5-6 times the length $\mathcal{L}$ is in general required for the corresponding ML decoders.

As illustrative example, in Table I the bit-error-rate (BER) for the case of a BPSK-modulated binary message sequence crossing the discrete-time baseband ISI channel of [4], Tab.6.7.1, of length $\mathcal{L}$=6 are reported. Hard-decision demodulation and AWGN are assumed; the signal-to-noise ratio is evaluated at the input of the receiver' quantizer. In Tab.II the corresponding steady-state values of the Bhattacharyya-like bound (2) are reported. In [5] the symbol-by-symbol MAP decoders described in this work are employed for decoding Trellis-encoded data sequences. It is finally observed that if the transmitted sequences are equiprobable, the proposed MAP receivers coincide with the corresponding symbol-by-symbol ML receivers.

REFERENCES

[1] K. Abend, B.D. Fritchman, "Statistical detection for communication channels with intersymbol interference", *Proc. of the IEEE*, vol.58, pp.779-785, May 1970.

[2] E. Baccarelli, R. Cusani, "Universal Optimal Estimation and Detection of Markov Chains over Noisy Discrete Channels", INFOCOM Dpt Tech. Rep. n.005-02-94, Univ. Rome La Sapienza, Italy, Feb. 1994.

[3] G.K. Kaleh, R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels", *IEEE Trans. on Comm.*, vol.42, no.7, pp.2406-2413, July 1994.

[4] J.G. Proakis, *Digital Communication*, 2nd ed., McGraw-Hill, 1989.

[5] E. Baccarelli, R. Cusani, L. Piazzo, "A Novel General Approach to the Optimal Synthesis of Trellis-Codes for Arbitrary Noisy Discrete Memoryless Channels", *Proc. Int. Symp. Inf. Theory*, ISIT95, Sept. 1995.

[6] E.Baccarelli,R.Cusani,S.Galli, "Nonlinear Decision-Delay Constrained MAP Decoders with Soft-Decision Demodulation for Noisy Time-Dispersive Time-Variant Channels", submitted to *IEEE Trans. on Comm.*

Channel of length $\mathcal{L}$=6	*Sequence detectors(VA)*			*Proposed detectors*		
	$\Delta = 0$	$\Delta = \mathcal{L}$-1	$\Delta = \infty$	$\Delta = 0$	$\Delta = \mathcal{L}$-1	$\Delta = \mathcal{L}$+1
SNR = 7	0.2822	0.2587	0.249	0.2488	0.2239	0.223
SNR = 15	0.0459	0.0255	0.0122	0.0428	0.0229	0.0153

Tab.I

Upper bounds	$\Delta = 0$	$\Delta = \mathcal{L}$-1	$\Delta = \mathcal{L}$+1
SNR = 7	0.4935	0.4718	0.3910
SNR = 15	0.0995	0.0502	0.0314

Tab.II

MMSE-Optimal Feedback and its Applications

Felix Tarköy
Institute for Signal and Information Processing
ETH Zürich, Switzerland
e-mail: tarkoey@isi.ee.ethz.ch

Abstract — **Decision-feedback suffers from the problem that wrong decisions deteriorate further decisions by increasing the interference in the observation. MMSE-optimal feedback minimizes this residual interference power. Applications include decision-feedback equalization and delay estimation in code-division multiple-access (CDMA) systems.**

I. Introduction

In many applications, one observation (e.g., a sequence) can give rise to decisions on many random variables. For optimal results in the maximum-likelihood (ML) sense, all random variables have to be estimated jointly. *Decision-feedback* can be used as a less complex but suboptimal method. The estimate of each random variable in turn is fed back to the observation with the aim of reducing the influence of this random variable on further decisions. One application is decision-feedback equalization: data estimates are appropriately filtered and fed back to cancel out the interference from the corresponding data symbol on future decisions. However, wrong decisions can *increase* the influence of a previously decided symbol instead of diminishing it. E.g., a wrong decision on a binary antipodal symbol increases the interference power of that symbol in the observation by a factor of four. The problem arises from the implicit assumption of decision-feedback equalization that all decisions are correct.

II. MMSE-Optimal Feedback Strategy

Our purpose is to mitigate the detrimental effects of decision-feedback by an improved feedback scheme. We will treat the case where an observation Y (e.g., an infinite-length sequence $Y[\cdot]$) can be expressed as a sum of two real-valued terms, one of which is independent of the random variable X (e.g., a data symbol $X[n]$) to detect. The observation can be written as

$$Y = Y_0 + f(X), \tag{1}$$

where $f(\cdot)$ denotes an arbitrary function. Every feedback scheme subtracts some function $r(Y)$ from the observation Y, and hence the latter becomes

$$Y' = Y_0 + f(X) - r(Y). \tag{2}$$

Being interested in minimizing the impact of X on the observation Y, a reasonable criterion of goodness is the average residual power due to X after cancellation. Therefore, one is interested in finding

$$r_0(Y) = \arg\{\min_{r(\cdot)} E[\|f(X) - r(Y)\|^2 \mid Y = y]\}, \tag{3}$$

where $\|\cdot\|^2$ denotes the squared Euclidean norm.

The problem raised by (3) is an instance of the well-understood Bayesian (nonlinear) minimum mean-squared error (MMSE) estimation problem (see, e.g., [1, Section 7-5]). It follows that the MMSE-optimal feedback function is

$$r_0(Y) = E[f(X) \mid Y = y], \text{ all } y. \tag{4}$$

III. MMSE-Optimal Feedback Equalization

In a decision-feedback scheme, the observation Y corresponds to the received sequence $Y[\cdot] = \sum_{m=0}^{M} g[m]X[\cdot - m] + Z[\cdot]$. After deciding on a transmitted symbol $\hat{X}[k]$, a decision-feedback scheme subtracts the sequence

$$r^{(k)}[\cdot] = g[\cdot - k]\hat{X}[k]. \tag{5}$$

On the other hand, the MMSE-optimal scheme subtracts [3]

$$r_0^{(k)}[\cdot] = g[\cdot - k] \cdot E[X[k] \mid Y[\cdot] = y[\cdot]]. \tag{6}$$

IV. Delay Estimation in CDMA

MMSE-optimal feedback can also be applied to estimate the relative transmission delays of the users of an asynchronous code-division multiple-access (A-CDMA) system. A possible scheme can be derived from *successive cancellation* [2]: the users' delays are estimated in turn and appropriately subtracted from the observation to improve subsequent estimates. Again, better performance for this general feedback scheme is achieved by using MMSE-optimal feedback. Figure 1 illustrates the gain for a specific A-CDMA system with randomly chosen, repeatedly emitted synchronization sequences.

Other promising multi-user applications include MMSE-optimal multi-user decision-feedback and interference cancellation in CDMA data detection (cf. [4] for a related approach).

References

[1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. Singapore: MacGraw-Hill, 1984.

[2] A.J. Viterbi, "Very Low Rate Convolutional Codes for Maximum Theoretical Performance of Spread-Spectrum Multiple-Access Channels," *IEEE J. Sel. Areas in Comm.*, vol. SAC-8, pp. 641–649, May 1990.

[3] D.P. Taylor, "The Estimate Feedback Equalizer: A Suboptimum Nonlinear Receiver," *IEEE Trans. Comm.*, vol. COM-21, pp. 979–990, Sept. 1973.

[4] M. Bossert, Th. Frey, "Interference Cancellation in the Synchronous Downlink of CDMA Systems," presented at COST 231, TD(94)70, Prag, 1994.

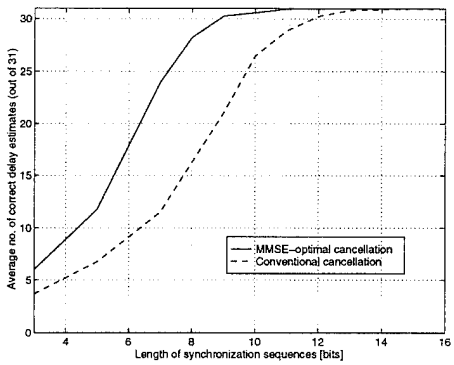


Fig. 1: Successive cancellation for A-CDMA; 31 equal-energy users with length-31 spreading sequences.

A Comparison of a Single-Carrier System using a DFE and a Coded OFDM System in a Broadcast Rayleigh-Fading Channel

Sarah Kate Wilson [1] and John M. Cioffi [2]

School of Electrical and Computer Engineering, Purdue University,
West Lafayette, Indiana 47907-1285

Abstract — **This paper compares the Coded Orthogonal Frequency Division Multiplexing (COFDM) system and a single-carrier system using decision feedback equalization (DFE) in a Rayleigh-fading environment assuming perfect knowledge of the channel and ignoring error propagation in the DFE. Analytic techniques are introduced to bound the average probability of error of a single-carrier system using decision-feedback equalization and the average probability of error of a COFDM system in a two-path fading channel.**

I. Introduction

We compare a single-carrier broadcast system using decision feedback equalization (DFE) to COFDM in a slowly Rayleigh-fading environment. Analytic techniques are introduced for bounding probability of error for systems using DFE and COFDM. Diversity is a well-known technique to reduce the average probability of error in a fading channel [1], [2]. This paper shows how the inherent diversity of a single-carrier system using a DFE is equivalent to the inherent diversity of a COFDM system in a two-path fading channel with proper coding and interleaving.

II. Diversity Calculations for a DFE in a Two-Tap Fading Channel

We consider the performance of a DFE when the received channel pulse, after any receiver filtering and symbol-spaced sampling, is a two-tap channel. To upper-bound the probability of error of the DFE, we consider the zero-forcing DFE, because a zero-forcing DFE will have a higher probability of error than a DFE [2]. The zero-forcing DFE will convert the channel pulse response to one that is causal, monic and minimum-phase. It will then subtract the precursor ISI. Consider a two-tap channel pulse response, $h(D) = h_0 + h_1 D$. If $|h_0| > |h_1|$, the feedforward section of the equalizer (including matched filtering) will be simply $\frac{1}{h_0}$ and the feedback section will be $\frac{h_1}{h_0} D$. Without loss of generality, assume $E[\mathrm{x}^2]/\sigma^2 = 1$. The resulting instantaneous SNR will be $|h_0|^2$. Now suppose that $|h_1| > |h_0|$. In this case, the D-transform of the feedforward section of the equalizer will be $W_{ZF-DFE}(D) = \frac{h_0^* + h_1^* D^{-1}}{h_1^*(h_1 + h_0 D^{-1})}$ and the feedback section will be $\frac{h_0^*}{h_1^*} D$. In this case, the resulting SNR will be $|h_1|^2$. Therefore, the zero-forcing DFE in the two-tap channel case selects the larger of the two paths. This is equivalent to selection diversity for a two-antenna channel. Therefore, selection diversity provides an upper-bound to the probability of error for a DFE in a Rayleigh-fading channel.

On the other hand, the matched-filter-bound for a two-tap fading channel can be used to lower-bound the probability of error for a DFE [3]. Both the upper and lower bounds for a DFE in a fading two-path fading channel exhibit two-path diversity. Therefore a DFE exhibits two-path diversity in a fading channel.

III. Diversity Calculations for a COFDM System

Given a COFDM system with convolutional coding and interleaving across frequency tones, we can find the probability that a codeword is mistaken for its nearest neighbor by recognizing that the coded SNR is a quadratic sum of complex Gaussian random variables. This will give a conservative approximation to the amount of diversity inherent in a COFDM system. Given two paths in the channel separated by time τ, we can write the SNR at tone i by

$$w_i = |h_0|^2 + |h_1|^2 + h_0 h_1^* e^{j\omega_i \tau} + h_0^* h_1 e^{-j\omega_i \tau}, \quad (1)$$

where $\omega_i = \frac{2\pi i}{T}$ and $\frac{1}{T}$ is the width of each tone in the OFDM symbol.

Now the coded SNR can be written as:

$$\sum_{i \in I} \alpha_i w_i = \sum_{i \in I} \alpha_i \left(|h_0|^2 + |h_1|^2\right) + 2\mathrm{Real}\left(\sum_{i \in I} \alpha_i h_0 h_1^* e^{j\omega_i \tau}\right), \quad (2)$$

where I is a set that indexes the differing tones between nearest-neighbor codewords and α_i adjusts the SNR to reflect the distance between coded symbols on a given branch of the trellis. Equation (2) has the same form as the instantaneous SNR of the matched filter bound for a two-tap fading channel found in [3]. We can use this to show that the diversity of the system is at most 2 for a two-path channel, regardless of the number of diversity branches of the code.

IV. Conclusions

This paper introduces analytic techniques to bound the probability of error for both a single-carrier system using a DFE and a COFDM system. It shows analytically that in a two-path channel, both a single-carrier system using a DFE and a COFDM system with interleaving across the tones exhibit two-path diversity in the average probability of error.

References

[1] William C. Jakes, *Microwave Mobile Communications*, John Wiley and Sons, 1974.

[2] John Proakis, *Digital Communications*, McGraw-Hill, 1989.

[3] J.E. Mazo, "Exact matched filter bound for two-beam Rayleigh fading", *IEEE Transactions on Communications*, vol. 39, pp. 1027–1030, July 1991.

[1] This work was supported by NSF grant 2DPL133

[2] of Information Systems Laboratory
Stanford University
Stanford, California 94305-4055

A Decision Feedback Filter

Saul B. Gelfand[1]
and Michael P. Fitz[1]
School of Electrical & Computer Engineering, Purdue University,
West Lafayette, IN, USA 47907-1285

Due to its low complexity and robust performance, the decision feedback equalizer (DFE) [1] continues to play an important role in high data rate and/or low cost systems, e.g., digital subscriber lines, magnetic recording and (possibly) mobile radio. Here we examine the possibility of recursive equalizers which perform soft-decisioning with complexity comparable to the DFE. The approach taken here involves initially making decisions like a DFE, followed by post filtering of these decisions using a recursive (conditionally) linear filter structure; we call this a decision feedback filter (DFF). Significantly, the DFF can be set-up to provably retain the performance capability of the DFE at high SNR, and empirically has improved performance over a wide range of SNR.

The DFF is derived assuming the usual AWGN FIR equivalent baseband model (it is possible to modify the DFF to take into account correlated noise as would arise from using a mean-square whitened matched filter in the front end). Starting with the fixed-lag Kalman filter (KF) [2], the current symbol estimate and error variance are isolated from the previous symbol estimates and error covariance. Then the current symbol (linear) estimate is replaced by a (nonlinear) MAP estimate (based on the approximation that the current observation is conditionally Gaussian, conditioned on the current symbol and past data), and the error variance is adjusted accordingly. The current symbol estimate is thus filtered and fed back, and eventually (after a fixed number of additional observations) thresholded to obtain the final estimate. Some simulation results demonstrate the improved BER of the DFF compared with the DFE.

A rigorous analysis of the DFF is performed. It turns out that the stability and performance of the DFF is related to the magnitude of the (computed) conditional error variance p_k of the current symbol estimate. We identify two critical constants $\alpha, \beta (\beta < \alpha)$ with the following properties:

(i) If $\sup_k p_k < \alpha$ then the DFF state is mean square bounded, uniformly as $SNR \rightarrow \infty$ (this is true even for nonminimum phase channels, in contrast to the KF which tends toward instability as $SNR \rightarrow \infty$ [3]).

(ii) If $\sup_k p_k < \beta$ then the DFF BER is asymptotically upper bounded by the DFE BER as $SNR \rightarrow \infty$.

Since p_k is random in the DFF (since the current symbol estimate is nonlinear) these conditions would generally have to be imposed in order to guarantee one or both of the above properties, i.e.,p_k would be replaced by $\max(p_k, \gamma)$ for some $\gamma < \alpha$. These results are proved using variations on comparison techniques familiar in the analysis of recursive stochastic algorithms, and some basic results on DFEs [4]. The novely in the analysis lies in the fact that the continuous-state DFF can be effectively compared to the discrete-state DFE.

[1]This work was supported by NSF under Grant NCR-9406073

References

[1] C.A. Belfiore and J.H. Park, Jr., "Decision Feedback Equalization," *Proc. IEEE*, vol., 67, August 1979, pp. 1143-1156

[2] R.E. Lawrence and H. Kaufman, "The Kalman Filter for the Equalization of a Digital Communications Channel," *IEEE Trans. Commun.*, vol. COM-19, December 1971, pp. 1137-1141.

[3] S. Benedetto and E. Biglieri, "On Linear Receivers for Digital Transmission Systems," *IEEE Trans. Commun.*, vol. COM-22, September 1974, pp. 1205-1215.

[4] D.L. Duttweiler, J.E. Mazo and D.G. Messerschmitt, "An Upper Bound on the Error Probability in Decision-Feedback Equalization," *IEEE Trans. Info. Theory*, vol. IT-20, July 1974, pp. 490-497.

Combined Decision Feedback Equalization and Coding for High SNR Channels

Daniel Yellin
Electrical Engineering – Systems
Tel-Aviv University, Israel 69978
danny@eng.tau.ac.il

Alexander Vardy
Coordinated Science Laboratory
University of Illinois, IL 61801
vardy@golay.csl.uiuc.edu

Ofer Amrani
Electrical Engineering – Systems
Tel-Aviv University, Israel 69978
ofera@eng.tau.ac.il

Abstract — **We present a novel scheme that combines decision feedback equalization (DFE) with high-rate error-detection coding in an efficient manner. The proposed scheme is evaluated both analytically and by means of comprehensive computer simulations. In our analysis, we introduce an approximate mathematical model taking into account the error propagation phenomenon. Both evaluation methods show that power savings of 2.5 dB to 3 dB over the conventional DFE can be achieved at the expense of a moderate complexity increase.**

I. Introduction

Motivated by the desire to transmit the maximum possible data rate through a band-limited additive-noise channel with intersymbol interference (ISI), a considerable research effort has been devoted to equalization techniques for such channels. Various approaches to the equalization problem can be roughly divided into three classes: linear equalization, decision-feedback equalization (DFE), and maximum-likelihood sequence estimation (MLSE). DFE can significantly outperform the linear equalizer on channels with severe frequency attenuation. A major problem with the DFE, however, is the error-propagation. On the other hand, while MLSE is the most powerful technique, it is also the most complex to implement. Recently, a number of schemes — generally known as reduced-state sequence estimation (RSSE) — were proposed [1, 2, 3] in an attempt to approach the performance of the MLSE at reduced complexity. Both [1] and [2, 3] are based on the idea of pruning the MLSE trellis, namely constructing only a small subset of all the paths in the trellis, and then selecting the most likely of these paths as the estimated sequence. The proposed scheme would be a further step in this direction, with the following two major differences. First, the path generation mechanism of RSSE schemes is controlled by some *a priori* determined rule. In contrast, we propose to generate the subset of paths in the trellis in accordance with the actual noise samples in the channel. Since with this approach, additional complexity is introduced only where it is needed, one would have to consider, on the average, very few paths. Second, we propose to significantly improve upon the performance of both RSSE and MLSE by integrating a simple high-rate error-detection code into the receiver structure.

II. The Proposed Scheme

The following is a simplified overview of the general ideas underlying the proposed scheme. The source data stream is partitioned into blocks of k symbols, which are subsequently encoded into the codewords of a cyclic code of length n. Let a_t denote the transmitted symbols, ν_t the noise samples, and $y_t = \sum_{i=0}^{M} a_{t-i}h_i + \nu_t$ the output sequence of an ideal zero-forcing feed-forward equalizer (FFE), where $\{h_t\}_{t=0}^{M}$ stands for the (postcursor) channel impulse response. Then the conventional DFE operates as follows:

$$z_t = a_t + \sum_{i=1}^{M}(a_{t-i} - \widehat{a}_{t-i})h_i + \nu_t, \qquad (1)$$

where z_t is the signal at the slicer input, and $\widehat{a}_t$ denotes the estimated symbol. We shall refer to the sequence $\{\widehat{a}_t\}$ as the *standard path*. Note that at each time instance the channel noise may be estimated as $\widehat{\nu}_t = z_t - \widehat{a}_t$. The basic idea is to diverge from the standard path, i.e. open a new path in the trellis, only when the estimated noise value $\widehat{\nu}_t$ is large. The same principle may then be employed for branching from each of the paths that are already followed.

Once all the paths have been generated as described above, they are processed in some fixed order and the first one that happens to be in the code is selected as the estimated sequence. Note that the total number of paths to be considered could still be quite large. However, we employ the structure of cyclic codes to implement the selection process with very low computational effort.

III. Performance Analysis

In order to analyze the performance, we introduce an approximate mathematical model, which takes into account the error propagation using a Gilbert-Elliot channel model. That is, we assume that the signal z_t in (1) can be described by a two-state Markov process, where one of the states is error-free while the other is the error-propagation state. Based on this model upper and lower bounds on the overall probability of error are derived. These show that with the proposed method, the probability of error can be made several orders of magnitude lower than that obtained with the conventional DFE. In addition, comprehensive computer simulations have been performed. The simulations concur with the theoretical analysis, indicating a significant improvement over the conventional DFE. More specifically, simulation results for the HDSL channel test-loop #4, which is considered to be a difficult test channel with a considerable amount of ISI, show a reduction of three to four orders of magnitude in the overall system BER, which converts into savings of some 2.5 dB. For other HDSL channels, power savings of up to 3 dB were achieved.

References

[1] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, pp. 428–436, 1989.

[2] M.V. Eyuboğlu and S.U.H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, pp. 13–20, 1988.

[3] M.V. Eyuboğlu and S.U.H. Qureshi, "Reduced-state sequence estimation for coded modulation on intersymbol interference channels," *IEEE J. Sel. Areas Commun.*, vol. SAC-7, pp. 989–995, 1989.

Maximum Likelihood Sequence Estimation for Non-Gaussian Band-Limited Channels

M. Cordier and E. Geraniotis[1]
Dept. of Electrical Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742

Abstract — **This work aims at providing near-optimal and sub-optimal receiver designs for digital communications in the presence of Non-Gaussian noise and intersymbol interference (ISI). Potential applications include wireless indoor (office or factory floor) communications (e.g., [1]) which are characterized by ISI due to multipath fading and limited channel bandwidth and by non-Gaussian background noise.**

In our problem the received signal, corrupted by ISI and additive non-Gaussian noise is given as $r(t) = \sum_{m=-\infty}^{+\infty} a_m h(t - mT) + n(t)$ where $a_m = \pm 1$ (binary signaling) and $h(t)$ represents *the impulse response of the channel.* We have to determine a 2J bit sequence of transmitted bits $a_{-J} \dots a_{J-1}$ from the received wave-form over the observation interval. The problem of sequence estimation is posed as a *M-ary hypothesis testing problem*: $\forall i, 1 \leq i \leq M = 2^{2J}, H_i$ corresponds to the fact that the sequence A_i was sent, i.e.:

$$H_i : r(t) = \sum_{m=-J}^{J-1} a_m^{(i)} h(t - mT) + n(t) = x(A_i, t) + n(t)$$

$r(t)$ is sampled with the sampling rate $\frac{1}{T'} = \frac{L}{T}$ where L denotes the number of samples over a single bit interval. Thus, we have: $t_k = kT'; r(t_k) = x(t_k) + n(t_k) = x_k + n_k$ with $x_k = \sum_m a_m h(t_k - mT)$. Then, we can form a discrete representation of the form: $R_k = X_k + N_k$, where $R_k = [r_1, r_2, \dots, r_P]^T$, $X_k = [x_1, \dots, x_P]^T$, $N_k = [n_1, \dots n_P]^T$. P is chosen such that all the bits creating ISI for the sequence considered are observed. If we assume that the impulse response $h(t)$ becomes zero for $t > NT$ and $t < -NT$, that is the ISI is assumed only over $K = 2N + 1$ adjacent bit sequences, then we have: $P = 2(J + N)L$ We consider sufficiently long sequences that $P >> N$. The decision rule which minimizes the probability of error is: Choose the sequence $A_i = \{a_k^{(i)}\}$ if $P_{r/Hi}(R/Hi) > P_{r/Hj}(R/Hj) \ \ \forall i \neq j$.

Under very low SNR and i.i.d conditions, it can be shown that

$$P_{r/Hi}(R/Hi) = [\sum_{k=1}^{P} g(r_k) x_k(A_i)] P_{r/Ho}(R/Ho)$$

where $Ho : r(t) = n(t)$, $g(rk) = \frac{d}{dr_k} \ln P(r_k/Ho) = \frac{d}{dr_k} \ln P_n(r_k)$, and $x_k(A_i) = \sum_{m=-J}^{J-1} a_m^{(i)} h(t_k - mT)$. Thus, a sufficient statistic for detection is $\lambda_i = \sum_{k=1}^{P} g(r_k) x_k(Ai)$. It should be emphasized that the problem treated here is that of coherent reception and that the knowledge of the impulse response of the channel is required to compute λ_i.

In case of correlated noise samples, the maximization of $P_{r/Hi}(R/Hi)$ can be replaced by a M-ary classification problem involving binary hypothesis testing and pairwise likelihood ratios. For $L_{i,j} = \frac{P(R/H_i)}{P(R/H_j)}$, the decision process is

$$\begin{cases} \text{- compute } L_{i,j} & i \neq j \quad i,j = 1,2,\dots M \\ \text{- decide } H_i \text{ if } & \forall j \neq i \quad L_{i,j} > \eta_{i,j}. \end{cases}$$

Since the computation of $L_{i,j}$ is intractable in a non-Gaussian environment, suitable approximations have to be employed. Indeed, two approaches are followed: (i) The *Generalized Correlator* (GC), as in the iid case. Here we extend the work of [2]; low SNR conditions and large sample sizes are assumed. And (ii) the *Linear Quadratic Detector* (LQD) of [3], which can be designed to match $L_{i,j}$ under any SNR conditions and without having to resort to large noise samples. The generalized likelihood ratio is used here rather than the deflection. Only the knowledge of 1^{st} to 4^{th} order statistics of the observations, under both hypotheses, is required.

We derive the appropriate GC to fit $L_{i,j}$. Each likelihood ratio is then approximated by a statistic of the form:

$$T_{i/j} = \sum_{k=1}^{P} (x_k(A_i) - x_k(A_j)) g_{i/j}(r_k) = \sum_{k=1}^{P} s_k^{(i/j)} g_{i/j}(r_k)$$

and the discrimination test between Hi and Hj becomes: $T_{i/j} >_{<H_j}^{H_i} \eta_{i/j}$. This memoryless discriminator is characterized by the non-linearity $g_{i/j}$ and the threshold $\eta_{i/j}$. When all the non-linearities $g_{i/j}$ are given, the corresponding thresholds can be determined so that the different tests form an appropriate partition of the observation space. Each nonlinearity is selected by maximizing the appropraite efficacy functional and solving the resulting integral equation numerically. However, we also need here an estimate of the impulse response of the multipath channel. Actually for $T_{i/j}$ we need the sample power and sample autocorrelation functions of the signal components under the sequences i and j and the marginal and bivariate pdfs of the background noise. The latter noise distributions can be obtained via histograms or Kernel estimation; noise estimation can be done on-line as long as the signal level remains of sufficiently low SNR. On the other hand, the channel impulse response can be estimated by filtering out the background noise during the training stage.

[1] C. Cordier is with the Ecole Nationale Superiere des Telecommunications, France; he held an internship with the ISR at the University of Maryland during the summer and Fall of 1994

References

[1] H. Hashemi, "The indoor Radio Propagation Channel," *IEEE Proceedings*, vol. 81, Jul. 93.

[2] S. Prasad and S.S. Pathak "Optimum data receivers for low SNR data signals in non-Gaussian noise and intersymbol interference," *IEE Proc. Pt F*, no 5, Oct.88.

[3] B. Picinbono and P. Duvaut, "Optimal linear-quadratic systems for detection and estimation," *IEEE Trans. Inform. Theory*, vol. IT-34, no. 2, p 304-311, Mar. 1988.

Incoherent diversity detection of fading signals in correlated non-Gaussian noise

Luciano Izzo and Mario Tanda

Università di Napoli Federico II, Dipartimento di Ingegneria Elettronica, Via Claudio 21 I-80125 Napoli, Italy

Abstract — **The paper deals with the synthesis of an asymptotically optimum diversity detector for the incoherent detection of a bandpass signal subject to slow and nonselective fading and embedded in correlated spherically invariant noise.**

Summary

The detection problem under consideration can be represented by the hypothesis test

$$\begin{aligned} H_0: \quad \tilde{\boldsymbol{r}}_p &= \tilde{\boldsymbol{n}}_p, \\ H_1: \quad \tilde{\boldsymbol{r}}_p &= \frac{\gamma}{\sqrt{N}} A_p e^{j\theta_p} \tilde{\boldsymbol{v}} + \tilde{\boldsymbol{n}}_p, \end{aligned} \qquad p = 1, 2, ..., L, \quad (1)$$

where $\tilde{\boldsymbol{r}}_p$ and $\tilde{\boldsymbol{n}}_p$ are N-dimensional row vectors whose components are samples drawn from the complex envelopes of the received signal and the noise (respectively) on the *pth* diversity branch. The vector $\tilde{\boldsymbol{v}}$ represents the vector of the samples drawn from the complex envelope of the bandpass signal to be detected. The random variable (RV) A_p, which assumes nonnegative real values, accounts for the presence of a slow amplitude fading on the *pth* channel. The RV θ_p is assumed to be uniformly distributed over a 2π interval (incoherent detection). The RV's A_p and θ_p, and the noise vector $\tilde{\boldsymbol{n}}_p$ are mutually independent on each channel. Furthermore, amplitude fadings, phases, and noises on the different diversity channels are mutually independent. Finally, the signal amplitude is assumed to decrease as $\gamma/\sqrt{N}$ (with γ a positive constant) so that the signal-to-noise ratio (SNR) is finite and not zero for any value of N.

The assumed spherically invariant (SI) noise model allows one [1,2] to write $\tilde{\boldsymbol{n}}_p = a_p \tilde{\boldsymbol{g}}_p$, where a_p is a nonnegative RV independent of $\tilde{\boldsymbol{g}}_p$, which is a zero-mean complex Gaussian vector characterized by a 2N × 2N correlation matrix $\sigma_{gp}^2 \boldsymbol{K}_p$ with σ_{gp}^2 the common variance of the inphase and quadrature components.

The matrices $\boldsymbol{K}_p$ $(p = 1, 2, ..., L)$ admit the Cholesky decomposition $\boldsymbol{K}_p = \boldsymbol{C}_p \boldsymbol{C}_p^T$, where T denotes transpose operation and $\boldsymbol{C}_p$ are 2N × 2N invertible lower triangular matrices. Therefore, the theorem of reversibility and the closure property of the SI vectors under deterministic linear transformations [1] assure that the detector synthesized on the basis of the hypothesis test

$$\begin{aligned} H_0: \quad \tilde{\boldsymbol{x}}_p &= \tilde{\boldsymbol{w}}_p, \\ H_1: \quad \tilde{\boldsymbol{x}}_p &= \frac{\gamma}{\sqrt{N}} A_p e^{j\theta_p} \tilde{\boldsymbol{s}}_p + \tilde{\boldsymbol{w}}_p \end{aligned} \qquad p = 1, 2, ..., L, \quad (2)$$

retains the optimality properties of the detector synthesized starting from (1). In (2), $\tilde{\boldsymbol{w}}_p = \boldsymbol{w}_{pc} + j\boldsymbol{w}_{ps}$ is a white SI vector with modulating RV a_p, which is obtained by the transformation $(\boldsymbol{w}_{pc}, \boldsymbol{w}_{ps}) = (\boldsymbol{n}_{pc}, \boldsymbol{n}_{ps})(\boldsymbol{C}_p^{-1})^T$. Moreover, $(\boldsymbol{x}_{pc}, \boldsymbol{x}_{ps}) = (\boldsymbol{r}_{pc}, \boldsymbol{r}_{ps})(\boldsymbol{C}_p^{-1})^T$ and $(\boldsymbol{s}_{pc}, \boldsymbol{s}_{ps}) = (\boldsymbol{v}_c, \boldsymbol{v}_s)(\boldsymbol{C}_p^{-1})^T$.

The asymptotically optimum (AO) detector can be synthesized starting from an asymptotic expression of the likelihood ratio on the *pth* channel conditioned to A_p and θ_p, which can be derived following an approach similar to that considered in [3]. The resulting decision statistic for the AO detector is

$$T^{AO}(\tilde{\boldsymbol{x}}) = \sum_{p=1}^{L} \ln \left\{ E_{A_p} \left[\exp\left(-\frac{A_p^2 \gamma^2 N P_p}{\| \tilde{\boldsymbol{x}}_p \|^2} \right) \cdot I_0 \left(\frac{2 A_p \gamma \sqrt{N}}{\| \tilde{\boldsymbol{x}}_p \|^2} \mid \tilde{\boldsymbol{s}}_p^* \tilde{\boldsymbol{x}}_p^T \mid \right) \right] \right\}, \quad (3)$$

where $E_{A_p}[\cdot]$ denotes the statistical expectation with respect to A_p, $P_p \triangleq \| \tilde{\boldsymbol{s}}_p \|^2 / N$ provides a measure of the signal power on the *pth* channel, $\| \cdot \|$ denotes Euclidean norm, $I_0(\cdot)$ is the modified Bessel function of the first kind and zero order, and $*$ denotes complex conjugation.

If one assumes that the fading RV's A_p are Rayleigh distributed, it results that

$$T_{Ray}^{AO}(\tilde{\boldsymbol{x}}) = \sum_{p=1}^{L} \frac{\gamma^2 E_{A_p}(A_p^2) N \mid \tilde{\boldsymbol{s}}_p^* \tilde{\boldsymbol{x}}_p^T \mid^2}{\| \tilde{\boldsymbol{x}}_p \|^2 \left[\| \tilde{\boldsymbol{x}}_p \|^2 + \gamma^2 E_{A_p}(A_p^2) N P_p \right]} - \sum_{p=1}^{L} \ln \left[1 + \frac{\gamma^2 E_{A_p}(A_p^2) N P_p}{\| \tilde{\boldsymbol{x}}_p \|^2} \right]. \quad (4)$$

The main advantage of the proposed AO detector is that its structure does not depend on the univariate probability density functions (PDF's) of the noises on the diversity channels. The synthesis of the detection structure, however, requires a priori knowledge of the noise correlation matrices and the fading PDF's. Then, its complexity is just that of the fully optimum detector for a correlated Gaussian noise environment.

The detection probability and the false-alarm rate of the proposed AO detector in white SI noise depend on the signal to be detected only through the mean (over fading) SNR's on the diversity branches, resulting so unaffected by the signal shape. Consequently, the closure property of the SI vectors under deterministic linear transformations assures that the performance in correlated noise can be easily assessed by exploiting the relationship between the mean SNR's at the input and the output of the whitening filters.

References

[1] K. Yao, " A representation theorem and its applications to spherically invariant random processes," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 600–608, 1973.

[2] L. Izzo and M. Tanda, " Array detection of random signals in spherically invariant noise," *J. Acoust. Soc. Am.*, vol. 94, pp. 2682–2690, 1993.

[3] E. Conte, M. Lops, G. Ricci, "Distribution-free radar detection in compound-Gaussian clutter," *Proc. Int. Conf. "RADAR '92", Brighton, GB,* , pp. 98–101, Oct. 1992.

Statistics of Error Recovery Times of Decision Feedback Equalizers

Wendy W. Choy and Norman C. Beaulieu[1]
Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada

Abstract — **New upper and lower bounds to the mean recovery time of decision feedback equalization (DFE) are derived. The recovery time is defined as the time it takes the decision feedback equalizer (DFEQ) to reach the error free state after an error has corrupted an error free DFEQ. The derivations of the bounds assume a causal channel response, independent data symbols, and independent noise samples. The bounds are found to be tighter, especially at large SNR, than previous bounds in a numerical example.**

Intersymbol interference (ISI) in a communication system has a deleterious effect on system performance. The ISI arises because insufficient channel bandwidth causes the pulses to spread into adjacent pulse intervals at the receiver end. This spreading may increase or decrease the noise margin of the received signal depending on the relative polarities of the pulses. On the average, however, ISI increases the bit error probability.

One of the methods often used to combat the effects of ISI is to use DFE. A DFEQ operates by reconstructing the portion of the ISI due to previously transmitted symbols and then subtracting out this portion from the received signal. The reconstruction is based on estimating the previously transmitted symbols and the channel characteristics.

Assuming that the past decisions are correct, a DFEQ (with perfect channel identification) can eliminate ISI due to previously transmitted symbols in the span of the feedback filter completely. However, decision errors will result in residual ISI which may increase the probability of decision error in the future detected symbols. This leads to error propagation in the DFEQ. Analysis of a DFEQ is difficult because little is known about the distribution of the past decision errors.

It is important to know how fast a DFEQ can recover from an error; that is, how many symbol intervals it takes to clear up an initial error introduced into the feedback filter. Then one knows how many future decisions will be affected by the error. When the DFEQ has a finite number of taps in the feedback filter and the system response has a finite time duration, the communications system can be modelled as a finite state Markov chain as shown by Monsen [1] and Austin [2]. Austin, in [2], showed how to obtain the mean recovery time exactly through quasi-simulations and discussed bounding the mean recovery time. However, both of Austin's approaches require computational efforts that grow exponentially with the length of the DFEQ. The mean recovery time of a DFEQ with error state transition probabilities of 1/2 was also computed in [2]. Cantoni and Butler [3] derived an upper bound for the mean number of symbols required to reach the zero error state, starting from an arbitrary initial state and subject to noise. The bound depends only on the number of taps in the DFE feedback filter and the number of signal levels. Kennedy and Anderson [4] extended, generalized, and clarified the contributions in [3], and gave a class of channels for which the upper bound in [3] is exactly the mean recovery time.

Duttweiler, Mazo and Messerschmitt in [5] developed an aggregated states model of a DFEQ which was used to upper bound the average error probability. Beaulieu [6] modified the model in [5] to compute upper and lower bounds for the mean recovery time by writing difference equations for conditional, state dependent, mean recovery times. He also provided analytical proofs of some known results that previously were justified with intuitive arguments. Altekar and Beaulieu developed models in [7] that lead to tighter upper bounds on the average probability of error of a DFEQ than those of [5].

In this paper, new, tighter bounds on the recovery times of DFE are derived by modifying the models of [7] used for error probability upper bounds. The channel is modelled as a linear, shift-invariant, discrete-time filter. Using appropriate choices for defining states, a number of aggregated states models of the DFEQ can be constructed. Good choices for state models lead to improved bounds on recovery time statistics. Three models are constructed here, each of which leads to new and tighter bounds. A single errors model, a double consecutive errors model and an arbitrary double errors model are defined and used to derive bounds on recovery time statistics.

For the numerical example considered, the arbitrary double errors model gives the tightest bounds for the mean recovery time. At small SNR values, the new bounds from the three models and previous bounds coincide. At large SNR values, the new bounds are much tighter than previous ones. In particular, the new bounds from the three models all approach N, the length of the DFEQ, at large SNR.

References

[1] P. Monsen, "Adaptive equalization of the slow fading channel," *IEEE Trans. Commun.*, vol. COM-22, No. 8, pp. 1064-1075, Aug. 1974.

[2] M. E. Austin, "Decision feedback equalization for digital communication over dispersive channels," *Res. Lab. of Electronics. Mass. Inst. Technol.*, Cambridge, Tech. Rep. 461, Aug. 11, 1967.

[3] A. Cantoni and P. Butler, "Stability of decision feedback inverses," *IEEE Trans. Commun.*, vol. COM-24, No. 9, pp. 970-977, Sept. 1976.

[4] R. A. Kennedy and B. D. O. Anderson, "Recovery times of decision feedback equalizers on noiseless channels," *IEEE Trans. Commun.*, vol. COM-35, No. 10, pp. 1012-1021, Oct. 1987.

[5] D. L. Duttweiler, J. E. Mazo and D. G. Messerschmitt, "An upper bound on the error probability in decision feedback equalization" *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 490-497, July 1974.

[6] N. C. Beaulieu, "Bounds of recovery times of decision feedback equalizers," to appear in *IEEE Trans. Commun.*.

[7] S. A. Alketar and N. C. Beaulieu, "Upper bounds to the error probability of decision feedback equalization," *IEEE Trans. Inform. Theory*, vol. 39, No. 1, pp. 145-156, Jan. 1993.

[1]This work was supported by NSERC Grant OGP0003986

A Lower Bound on the Undetected Error Probability of Block Codes [1]

Khaled A. S. Abdel-Ghaffar
University of California, Dept. of Elec. & Comp. Eng., Davis, CA 95616, USA

Abstract — **A new lower bound on the undetected error probability of block codes is presented and codes that meet this lower bound are characterized.**

I. Introduction

Let C be an $(n, M)_q$ code, i.e., C is a q-ary code of length n and size M. We assume that each codeword is transmitted with probability $1/M$ and that each letter is equally likely to suffer from an error that changes it into any one of the other $q-1$ letters with probability $\epsilon/(q-1)$ independently of other letters. We assume in the following that $\epsilon \leq (q-1)/q$, i.e., the probability that a letter is received correctly is at least equal to the probability that it is received as any given erroneous letter. For $w = 0, 1, \ldots, n$, let $A_w(\mathbf{c})$ be the number of codewords at distance w from the codeword $\mathbf{c}$ and $A_w(C) = \sum_{\mathbf{c}\in C} A_w(\mathbf{c})/M$. The undetected error probability of the code C is given by

$$P_{\rm ud}(C,\epsilon) = \sum_{w=1}^{n} A_w(C) \left(\frac{\epsilon}{q-1}\right)^w (1-\epsilon)^{n-w}.$$

In this paper, we derive a lower bound on the undetected error probability for linear and nonlinear block codes and present codes that meet this lower bound. In particular, these codes are optimal for error detection.

II. Lower Bound

The following theorem gives a lower bound on $P_{\rm ud}(C,\epsilon)$ for any $(n, M)_q$ code C.

Theorem 1 *Let C be an $(n, M)_q$ code and $0 \leq \epsilon \leq (q-1)/q$. Then, the undetected error probability of the code C satisfies the bound*

$$\begin{aligned} P_{\rm ud}(C,\epsilon) \geq \frac{1}{M}\sum_{w=1}^{n}\binom{n}{w}\left(\left\lceil\frac{M}{q^{n-w}}\right\rceil - 1\right) \\ \times \left(2M - q^{n-w}\left\lceil\frac{M}{q^{n-w}}\right\rceil\right) \\ \times \left(\frac{\epsilon}{q-1}\right)^w \left(1-\frac{q\epsilon}{q-1}\right)^{n-w}. \end{aligned}$$

Wolf, Michelson, and Levesque derived a lower bound on $P_{\rm ud}(C,\epsilon)$ for any binary linear code C of length n and dimension k [2]. This bound has been recently generalized by Klove [1] to linear codes over arbitrary finite fields of size q. The new bound described in this paper is not only more general than the Klove-Wolf-Michelson-Levesque (KWML) bound, in the sense that it holds for linear and nonlinear codes while the KWML bound holds only for linear codes, but it is also tighter. In fact, the new lower bound equals the KWML bound only if $k = n-1$, $k = n$, $\epsilon = 0$, or $\epsilon = (q-1)/q$. In all other cases, the new lower bound is larger than the KWML bound.

[1] This work was supported in part by NSF under grant NCR 91-15423.

III. Strictly Optimal Codes

We say that a code C is strictly optimal if its undetected error probability equals the lower bound stated in Theorem 1 for all $0 \leq \epsilon \leq (q-1)/q$. The following result gives a combinatorial characterization of strictly optimal codes.

Theorem 2 *An $(n, M)_q$ code C is strictly optimal if and only if C contains at least $\lfloor M/q^s \rfloor$ and at most $\lceil M/q^s \rceil$ codewords that agree on any given s indices, where $s = 1, \ldots, n$.*

Hence, a necessary condition for a code to be strictly optimal is that its Hamming distance is $n - \lceil \log_q M \rceil + 1$. The following result shows that this condition is also sufficient if M is an integer power of q. In this case, an $(n, M)_q$ code C is called maximum distance separable (MDS) if its Hamming distance equals $n - \log_q M + 1$.

Theorem 3 *If M is an integer power of q, then an $(n, M)_q$ code is strictly optimal if and only if it is MDS. In particular, an $(n, M)_q$ linear code is strictly optimal if and only if it is MDS.*

If M is not an integer power of q, then an $(n, M)_q$ code of Hamming distance $n - \lceil \log_q M \rceil + 1$ may not be strictly optimal. The following result determines necessary and sufficient conditions for the existence of strictly optimal binary codes.

Theorem 4 *A strictly optimal $(n, M)_2$ code, where n and M are positive integers and $M \leq 2^n$, exists if and only if one of the following conditions holds:*

- $n \in \{1, 2, 3\}$.
- $n = 4$ *and* $M \notin \{3, 4, 12, 13\}$.
- $n \geq 5$ *is odd and* $M \in \{1, 2, (2^n-2)/3, (2^n+1)/3, 2^{n-1}-1, 2^{n-1}, 2^{n-1}+1, (2^{n+1}-1)/3, (2^{n+1}+2)/3, 2^n-2, 2^n-1, 2^n\}$.
- $n \geq 6$ *is even and* $M \in \{1, 2, (2^n-1)/3, (2^n+2)/3, 2^{n-1}-2, 2^{n-1}-1, 2^{n-1}, 2^{n-1}+1, 2^{n-1}+2, (2^{n+1}-2)/3, (2^{n+1}+1)/3, 2^n-2, 2^n-1, 2^n\}$.

As an application in which M is not an integer power of q, we consider binary-coded-decimal codes where $q = 2$ and $M = 10$. It is interesting to note that the widely known 2-out-of-5 code, consisting of the ten binary sequences of length $n = 5$ with exactly two ones is not strictly optimal. On the other hand, Theorem 4 indicates the existence of a $(5, 10)_2$ strictly optimal code. Indeed, the code consisting of all binary sequences with exactly one or four ones is strictly optimal. This 1-or-4-out-of-5 code has undetected error probability of $4\epsilon^2 - 8\epsilon^3 + 4\epsilon^4 + \epsilon^5$, while the undetected error probability of the 2-out-of-5 code is $6\epsilon^2 - 18\epsilon^3 + 21\epsilon^4 - 9\epsilon^5$.

References

[1] T. Klove, "The weight distribution of cosets", *IEEE Trans. Inform. Theory*, vol. IT-40, pp. 911–913, May 1994.

[2] J. K. Wolf, A. M. Michelson, and A. H. Levesque, "On the probability of undetected error for linear block codes," *IEEE Trans. Commun.*, vol. COM-30, pp. 317–324, February 1982.

The worst-case probability of undetected error for linear codes on the local binomial channel

Torleiv Kløve[1]
Department of Informatics, University of Bergen, HIB, N-5020 Bergen, Norway

Abstract — **The worst-case probability of undetected error for a linear $[n,k;q]$ code used on a local binomial channel is studied. For the two most important cases it is determined in terms of the weight hierarchy of the code. The worst-case probability of undetected error for simplex codes is determined explicitly. A conjecture about Hamming codes is given.**

I. Background

The *local binomial channel* was defined implicitly by Korzhik and Fink [2, page 193] and explicitly by Korzhik and Dzubanov [1]. It is a channel which is a q-ary symmetric channel for each transmitted symbol, but the symbol error probability may vary from one transmitted symbol to the next.

Let $P_{\rm ue}(C,\bar{p}) = P_{\rm ue}(C,p_1,p_2,\ldots,p_n)$ denote the probability of undetected error when a codeword from a linear $[n,k;q]$ code C is transmitted over a local binomial channel with symbol error probability p_i for i'th transmitted symbol. Let the *worst-case error probability* be defined by

$$P_{\rm wc}(C,v) = \max\Big\{P_{\rm ue}(C,\bar{p}) \mid 0 \le p_i \le v \text{ for } 1 \le i \le n\Big\}.$$

The *support* of a vector $\bar{c}$ is given by

$$\chi(\bar{c}) = \{i \mid c_i \neq 0\}.$$

For a vector $\bar{c} = (c_1,c_2,\ldots,c_n)$ and a set $X = \{i_1,i_2,\ldots,i_r\}$, where $1 \le i_1 < i_2 < \cdots < i_r \le n$, we let

$$\bar{c}_X = (c_{i_1},c_{i_2},\ldots,c_{i_r}).$$

For an $[n,k;q]$ code C and a set X as above, we define

$$C_X = \{\bar{c}_X \mid \bar{c} \in C \text{ and } \chi(\bar{c}) \subseteq X\}.$$

We use the notation $P_{\rm ue}^{\rm S}(C,p) = P_{\rm ue}(C,p,p,\ldots,p)$ for the probability of undetected error when C is used on a q-ary symmetric channel with error probability p.

II. New results

Theorem 1 *Let C be an $[n,k;q]$ code. Then*

$$P_{\rm wc}(C,v) = \max\Big\{P_{\rm ue}^{\rm S}(C_X,v) \;\Big|\; X \subseteq \{1,2,\ldots,n\}\Big\}.$$

Theorem 2 *Let C be an $[n,k,d;q]$ code. Then*

$$P_{\rm wc}(C,1) = \frac{1}{(q-1)^{d-1}}.$$

Theorem 3 *Let C be an $[n,k,d;q]$ code. Let*

$$s = \max\Big\{r \;\Big|\; 1 \le r \le k \text{ and } d_r = d_1 + (r-1)\Big\},$$

where $d_1,d_2,\ldots,d_k$ is the weight hierarchy of C. Then

$$P_{\rm wc}(C,(q-1)/q) = \frac{q^s-1}{q^{d+s-1}}.$$

[1]This work was supported by the Norwegian Research Council

We consider a couple of particular classes of codes.

The first class of codes we consider is the binary simplex codes. For each $m \ge 1$ there is a binary simplex code S_m with parameters $n = 2^m-1$, $k = m$, $d_r = 2^m - 2^{m-r}$ for $1 \le r \le m$.

Theorem 4 *For $m \ge 3$, let*

$$v_0(m) = 1 - (2^m-1)^{-1/(2^{m-1}-1)}.$$

Then

$$P_{\rm wc}(S_m,v) = (2^m-1)v^{2^{m-1}}(1-v)^{2^{m-1}-1}$$

for $0 \le v \le v_0(m)$ and

$$P_{\rm wc}(S_m,v) = v^{2^{m-1}}$$

for $v_0(m) \le v \le 1$.

A similar theorem is true for the first order Reed-Muller codes.

The binary Hamming codes H_m, where $m \ge 1$, have parameters $n = 2^m-1$, $k = 2^m-1-m$, $d = 3$. We conjecture that the following result is true for all m (it is true for $m \le 4$).

Conjecture 1 *Define $g_r(v)$ for $r \ge 2$ by*

$$g_r(v) = \frac{1}{2^r}\Big(1 + (2^r-1)(1-2v)^{2^{r-1}}\Big) - (1-v)^{2^{r-1}}.$$

Let $v_1 = 1$, and for $r \ge 2$ let v_r be the root of the equation $g_r(v) = g_{r+1}(v)$ in the interval $(0,1)$.
Then $v_1 > v_2 > v_3 > v_4 > \cdots$,

$$P_{\rm wc}(H_m,0,v) = g_m(v)$$

for $0 \le v \le v_{m-1}$, and

$$P_{\rm wc}(H_m,0,v) = g_r(v)$$

for $v_r \le v \le v_{r-1}$ and $r = 2,3,4,\ldots,m-1$.

References

[1] V. I. Korzhik and S. D. Dzubanov, "Codes for error detection on local binomial channels", abstract of presentation at The International Workshop on Algebraic Coding Theory, Erevan, Armenia, 1989.

[2] V. I. Korzhik and L. M. Fink, *Noise-Stable Coding of Discrete Messages in Channels with a Random Structure*, (in Russian), Svyaz, Moscow, 1975.

Good error detection codes satisfy the expurgated bound

Takeshi Hashimoto[1]

Dept. Elect. Eng., University of Electro-Communication, Chofu, Tokyo 182, Japan

Abstract — **A q-nary (n,k) linear code is said to be proper if, as an error-detection code, the probability of undetectable error, P_{ud}, satisfies $P_{ud} \leq q^{-(n-k)}$ for completely symmetric channels. In this paper, we show that a proper code, as an error-correction code, satisfies the expurgated bound on the decoding error probability for a class of channels with the associated Bhattacharyya distance begin completely symmetric. Known results on the undetectable error probability then immediately imply that the expurgated exponent is satisfied by many codes which are regarded as good codes.**

I. Introduction

Random coding arguments tell that the most of codes satisfy the random coding bound and that the most of expurgated codes satisfy the expurgated bound asymptotically, but the most of time we can not tell if a specific code satisfies such bound. However, we can show that proper codes or asymptotically proper code satisfy or asymptotically satisfy the expurgated bound.

Before the works of Leung-Yan-Cheong et al.[1, 2], it had bee believed that the probability of undetectable error was upper bounded by $q^{-(n-k)}$ whenever a q-nary (n,k) linear code was used for error detection over a q-nary symmetric channel. They showed some examples of codes which do not satisfy this bound, and called a code which satisfies $q^{-(n-k)}$ bound a proper code. Subsequent works suggest that proper codes are also good as error-correction codes. In fact, it is shown that proper codes satisfy the asymptotic Gilbert-Varshamov bound on the minimu distance[3]. In this paper, we show that proper codes satisfy the expurgated bound.

II. Error probabilities

If we use c as an error-detection code for a DMC Q, then the undetectable error probability when $\mathbf{x}_i \in c$ is sent is written as

$$P_{ud}(\mathbf{x}_i) = \sum_{i \neq j} \exp\left\{ n \sum_{a,a'} V_{i,j}(a,a') \log Q(a'|a) \right\}, \qquad (1)$$

where $V_{i,j}(a,a')$ is the joint type of (a,a') in $(\mathbf{x}_i, \mathbf{x}_j)$.

On the other hand, if we use c as an error-correction code for another DMC P, then, from known arguments for the proof of the expurgated bound, we have a bound

$$P_e^s(\mathbf{x}_i) \leq \sum_{i \neq j} \exp\left\{ n \sum_{a,a'} \log \left[\sum_b \sqrt{P(b|a)P(b|a')} \right]^s \right\}, \quad (2)$$

where s is any non-negative number.

If we compare bounds (1)and (2), then we can notice some similarity. In fact, for a probability mass function $r(a)$, if we let

$$\hat{P}(a'|a) = \frac{r(a') \left[\sum_b \sqrt{P(b|a)P(b|a')} \right]^s}{\sum_{a''} r(a'') \left[\sum_b \sqrt{P(b|a)P(b|a'')} \right]^s}$$

be the channel induced from P, then we have

$$P_e^s(\mathbf{x}_i) \leq \exp\{-nsE_x(1/s, r) + n \log q\} \times \times \sum_{j \neq i} \exp\left\{ n \sum_{a,a'} V_{i,j}(a,a') \log \hat{P}(a'|a) \right\}, \quad (3)$$

where the expurgated exponent is

$$E_{ex}(R) = \sum_{\rho \geq 1} \left[\max_p E_x(\rho, p) - \rho R \right]$$

and the optimal p is used for r. Now, the relationship between (1) and (3) is obvious.

We can show the following theorem:

Theorem 1 *For a given DMC P, suppose that $\hat{P}$ is completely symmetric. Then, the expurgated bound*

$$P_e \leq \exp\{-nE_{ex}(R)\}$$

holds for all proper linear codes.

III. Concluding remark

Up to now, many codes are shown to be proper, and the above theorem then implies that those codes satisfy the expuraged bound. Unfortunately, the expurgated bound is greater than the known error bound for some codes such as the simplex code. Thus, our result does not necessarily solve all the probelm. However, if we note that the error bound is not known for the most of practical codes, our result gives an usefull tool to obtain the first approximation on the error probability.

Acknowledgements

The author thanks Prof. T. Fujiwara for kindly informing several new results on proper codes.

References

[1] S.K. Leung-Yan-Cheong and M.E. Hellman, "Concerning a bound on undetected error probability," *IEEE Trans. Information Theory*, vol. IT-22, pp. 235-237, March 1976.

[2] S.K. Leung-Yan-Cheong, E.R. Barnes, and D.U. Friedman, "On some properties of the undetected error probability of linear codes," *IEEE Trans. Information Theory* vol. IT-25, pp. 110-112, January 1979.

[3] R. Padovani and J.K. Wolf, "Poor error correction code are poor error detection codes", *IEEE Trans. Information Theory*, vol. IT-30, pp. 110-111, January 1984.

[1] E-mail: hasimoto@liszt.ee.uec.ac.jp

On the binomial approximation to the distance distribution of codes

Ilia Krasikov[1] and Simon Litsyn[2]

[1]Tel Aviv University, School of Mathematical Sciences, and Beit-Berl College, Kfar-Sava,
[2]Tel Aviv University, Department of Electrical Engineering – Systems, Ramat-Aviv 69978 Israel

Abstract — **We estimate the range where the distance distribution of a code can be approximated by the binomial distribution.**

I. Introduction

The binomial distribution is a well known approximation to the distance spectra of many classes of codes. For example, it is known to be tight for the weights of BCH codes with fixed minimal distance and of growing length. In general the range where the distance distribution is close to the binomial depends essentially on the dual distance. In the talk we present new bounds [1, 2, 3, 4] for this range for codes with the dual distance about half of the length n of the code, and for codes with the dual distance growing linearly in n.

II. BCH codes

Let the distance distribution of a code C be $\underline{B} = (B_0, \ldots, B_n)$, and $\underline{B}' = (B'_0, \ldots, B'_n)$ stand for the the dual spectrum, that is $\underline{B}'$ is determined by the MacWilliams transform.

Theorem 1 *In the extended BCH code of length $n = 2^m$ and minimum distance $2t + 2 \le 2^{[m+1]/2} + 2$,*

$$B_i = 0 \text{ for } i \text{ odd},$$

$$B_i = \binom{n}{i} n^{-t}(1 + E_i) \text{ for } i \text{ even},$$

$$|E_i| \le \frac{n^t \binom{n}{n/2}\binom{n/2}{i/2}}{\binom{n}{i}\binom{n}{d'}}.$$

Using the theorem we can analyze some particular cases.

Corollary 1 *If $t = o(\sqrt{n})$, and i grows linearly with n, $i = \sigma n$, then*

$$\frac{1}{n}\log_2 |E_{\sigma n}| \le -\frac{1}{2}H(\sigma) + o(1).$$

Corollary 2 *If $t = o(n^{\frac{1}{4}})$, $i = o(\sqrt{n})$, then*

$$|E_i| \le \sqrt{2}\, i^{i/2}\, e^{2(t-1)^2 - i/2}\, n^{t - i/2}\,(1 + o(1)).$$

Now we show that the binomial approximation can not be too tight. Define

$$r_i = B_i - \frac{|C|\binom{n}{i}}{2^n}(1 + (-1)^i B'_n).$$

This is evidently the deviation of the i-th spectrum element from the "expected" value given by the binomial distribution.

Theorem 2 *Let $B'_i = 0$, for $i \in [1, d'_1 - 1] \cup [d'_2 + 1, n - 1]$. Then*

$$\sum_{i=0}^{n} |r_i| \ge$$

$$\frac{2^n - 2|C|}{\sqrt{(n+1)}}\left(\max\left\{\binom{n}{[\frac{d'_2}{2}] - [\frac{d'_1+1}{2}]}\binom{n}{[\frac{d'_2}{2}] + [\frac{d'_1+1}{2}]}, \binom{n}{[\frac{d'_2-1}{2}] - [\frac{d'_1}{2}]}\binom{n}{[\frac{d'_2-1}{2}] + [\frac{d'_1}{2}] + 1}\right\}\right)^{-\frac{1}{2}}$$

For constant t this estimate turns out to be asymptotically tight for BCH codes with distance $d = 2t + 1 < \sqrt{n}$.

Another bound is a corollary of the Parseval identity.

Theorem 3

$$\sum_{i=0}^{n} \frac{r_i^2}{\binom{n}{i}} = \frac{|C|^2}{2^n} \sum_{i=d'_1}^{d'_2} \frac{B'^2_i}{\binom{n}{i}}.$$

III. Codes with linearly growing dual distance

Using an approach similar to linear programming we get the following results.

Theorem 4 *For $j/n \in (1/2 - 1/2\sqrt{\delta'(2-\delta')}, 1/2 + 1/2\sqrt{\delta'(2-\delta')})$,*

$$B_j = O\left(n\frac{\binom{n}{j}}{|C'|}\right).$$

Theorem 5 *For even codes, and $2j/n \in (\frac{(1-2\delta')^2}{2}, 1 - \frac{(1-2\delta')^2}{2})$,*

$$\log B_{2j} = \log \frac{\binom{n}{2j}}{|C'|} + O(\log n).$$

Let C be self-dual. In this case for d asymptotically greater than $0.146447...\, n$ (if such codes exist!) we can guarantee a wider interval of binomiality.

Theorem 6 *If there exists a self-dual code of length n with $d > (1/2 - \sqrt{2}/4)n(1 + o(1))$ then*

$$B_{2j} < \frac{4j}{n}\sqrt{\frac{2\pi(n-j)(n-2j)}{n}}\frac{\binom{n}{2j}}{|C'|}(1 + O(\frac{1}{n})),$$

for $(1/4 - \sqrt{3}/12)n \le j \le (1/4 + \sqrt{3}/12)n$, $j \ne n/2$.

References

[1] I.Krasikov and S.Litsyn, "On spectra of BCH codes", *IEEE IT*, to appear.

[2] I.Krasikov and S.Litsyn, "On the accuracy of binomial approximation to the distance distribution of codes", *IEEE IT*, to appear.

[3] I.Krasikov and S.Litsyn," Bounds on spectra of codes with known dual distance ", submitted.

[4] I.Krasikov and S.Litsyn," Estimates for the range of binomiality in codes' spectra ", submitted.

Extensions of linear codes.

R.Hill and P.Lizak
Department of Mathematics and Computer Science,
University of Salford,
Salford M5 4WT, UK.

One of the first results one meets in coding theory is that a binary linear $[n,k,d]$-code, whose minimum weight is odd, can be extended to an $[n+1,k,d+1]$-code.

This is one of the few elementary results about binary codes which does not obviously generalize to q-ary codes. Although one can readily extend a q-ary code, by adding a further check digit, it is not clear under what circumstances such an extension will increase the minimum distance. The aim of this paper is to give a simple sufficient condition for a q-ary $[n,k,d]$-code to be extendable to an $[n+1,k,d+1]$-code. The result is indeed a generalization of the above result for binary codes. It also generalizes a result for ternary codes due to van Eupen and Lisonek [2], whose proof made use of quadratic forms. Our generalization has an elementary proof.

Theorem 1. Let C be an $[n,k,d]$-code over $GF(q)$ with $\gcd(d,q)=1$ and with all weights congruent to 0 or d (modulo q). Then C can be extended to an $[n+1,k,d+1]$-code, all of whose weights are congruent to 0 or $d+1$ (modulo q).

Proof. Suppose $\underline{x}$ and $\underline{y}$ are two linearly independent vectors of length n over $GF(q)$ and suppose there are exactly z coordinate positions in which $\underline{x}$ and $\underline{y}$ both have a zero entry. Considering the $(q+1) \times n$ matrix whose rows are the vectors in the set $\{\underline{y}, \underline{x}+\alpha\underline{y} : \alpha \in GF(q)\}$, and counting the number of non-zero entries via rows and via columns, gives

$$w(\underline{y}) + \sum_{\alpha \in GF(q)} w(\underline{x}+\alpha\underline{y}) = q(n-z) \equiv 0 (\mathrm{mod}\, q). \qquad (1)$$

Let $C_0 = \{\underline{x} \in C : w(\underline{x}) \equiv 0 \pmod q\}$. If $\underline{x}, \underline{y} \in C_0$ then (1) implies that

$$\sum_{\alpha \in GF(q)\setminus\{0\}} w(\underline{x}+\alpha\underline{y}) \equiv 0 (\mathrm{mod}\, q).$$

By the hypothesis of the theorem, the only possibility is that $w(\underline{x}+\alpha\underline{y}) \equiv 0 \pmod q$ for all α. Hence C_0 is a linear subcode of C.

Furthermore, C_0 has dimension k-1. For otherwise there exists a two-dimensional subcode D of C all of whose non-zero codewords have weight congruent to $d \pmod q$. But then, if $\underline{x}, \underline{y}$ are linearly independent codewords in D, we have

$$w(\underline{y}) + \sum_{\alpha \in GF(q)} w(\underline{x}+\alpha\underline{y}) \equiv (q+1)d \equiv d \neq 0 (\mathrm{mod}\, q),$$

contradicting (1).

Let G be a generator matrix of C of the form

$$\begin{bmatrix} \underline{x} \\ \hline \\ G_0 \\ \\ \end{bmatrix},$$

where G_0 generates C_0. Then the matrix

$$\left[\begin{array}{c|c} \underline{x} & 1 \\ \hline & 0 \\ G_0 & \vdots \\ & 0 \end{array}\right]$$

generates an $[n+1,k,d+1]$-code with the required property. □

Theorem 1 can be useful in classifying codes with given parameters or in showing non-existence. Examples for ternary codes are given in [1] and [2]. We give here two other examples.

Example 1. We will prove the uniqueness of $\left[q^2, 4, q^2-q-1\right]$-codes over $GF(q)$.

It is known that there exists an optimal $\left[q^2+1, 4, q^2-q\right]$-code over $GF(q)$ which meets the Griesmer bound. The code is unique because the columns of a generator matrix form a (q^2+1)-cap in $PG(3,q)$ and hence must be an elliptic quadric [4]. Let C be a $\left[q^2, 4, q^2-q-1\right]$-code. The residual code of C with respect to a codeword of weight q^2-t $(2 \le t \le q\text{-}1)$ is a $[t,3,t\text{-}1]$-code which cannot exist by the Griesmer bound. So the only possible weights of C are q^2-q-1, q^2-q, q^2-1 and q^2. By Theorem 1, C can be extended to a $\left[q^2+1, 4, q^2-q\right]$-code. Finally the uniqueness of the punctured $\left[q^2, 4, q^2-q-1\right]$-code follows from the fact that an elliptic quadric admits a transitive automorphism group.

Remark. Example 1 provides a simple alternative proof of the well known fact that every q^2-cap in $PG(3,q)$ is contained in a (q^2+1)-cap, a result where geometric proof is fairly long (see e.g. [4]).

Example 2. It was shown in [3] that there does not exist a [28,5,19]-code over $GF(4)$. The proof can be simplified by using Theorem 1. It is straightforward to show that such a code has no codewords of weight 21,22,25 or 26 and hence can be extended to a [29,5,20]-code which had already been shown not to exist.

References

[1] M. van Eupen, "Some new results for ternary linear codes of dimension 5 and 6", preprint.

[2] M. van Eupen and P.Lisonek, "Classification of some optimal ternary codes of small length", preprint.

[3] R.Hill, I.Landgev and P.Lizak, "Optimal quaternary codes of dimension 4 and 5", Proceedings of Fourth International Workshop on Algebraic and Combinatorial Coding Theory, Novgorod, Russia, pp. 98-101, 1984.

[4] JWP Hirschfeld, "Finite Projective Spaces of Three Dimensions", Oxford University Press, 1985.

Tabu Search in Coding Theory

Kari J. Nurmela and Patric R. J. Östergård[1]
Department of Computer Science, Helsinki University of Technology,
FIN-02150 Espoo, FINLAND,
E-mail: {Kari.Nurmela,Patric.Ostergard}@hut.fi

Abstract — **Tabu search is a stochastic method for combinatorial optimization. It is shown how this method can be used to construct various record-breaking codes.**

I. Introduction

The problem of designing good codes can be seen as an optimization problem. Unfortunately, many instances of this problem are so hard that methods that provably find best possible codes with respect to given criteria cannot be used in practice. During the last decade, a lot of interest has been focused on stochastic methods for finding optimal and near-optimal solutions of difficult optimization problems. Simulated annealing has turned out to be a very promising such method. In 1987, El Gamal *et al.* [1] showed that simulated annealing can be used in the construction of several types of codes: constant weight codes, source codes, and spherical codes. Since then, simulated annealing and other stochastic methods have successfully been used in many papers to construct codes. For a survey of these results, see [3].

II. Tabu Search

Tabu search [2] is a combinatorial optimization method which in many recent studies has turned out to outperform other stochastic methods, including simulated annealing. One characteristic of tabu search is that it finds good near-optimal solutions early in the optimization run. Tabu search follows the steepest descent heuristic, but has additional features to avoid getting stuck in local optima.

At each step in the optimization process, a set of solutions that slightly differ from the current solution is evaluated. The solutions in this set are said to be *neighbors* of the current solution. In the neighborhood, a new solution that is best with respect to the cost function used is chosen. However, some of the neighbors must not be chosen, namely those obtained by inverses of one of the L most recent moves. The list of these forbidden moves, which has length L, is called the *tabu list.*

III. Constructing Codes Using Tabu Search

Tabu search can be applied to several construction problems in coding theory. In the search for a code with given parameters, the number of codewords is fixed and the problem is formulated as an optimization problem. For example, in the search for coverings, the cost function can be taken as the number of uncovered words in the space; a covering code has then cost value zero. The cost function of error-correcting codes can similarly be taken as the number of words that are covered more than once by Hamming spheres around the codewords; another approach is to consider the mutual distances between the codewords.

A direct search for a large code does not work very well. However, such a code can be found by imposing a structure on it. This can be done by searching for a code that is a union of cosets of a linear code or that has a nontrivial automorphism group.

Said and Palazzo [5] were—to our knowledge—the first to apply tabu search to problems in coding theory. They used the method to construct linear error-correcting codes. Recently, Östergård [4] successfully applied tabu search to the construction of covering codes. We present recent results on the application of tabu search to code constructions. We discuss covering, error-correcting, and spherical codes, and present new codes found by this approach.

[1]The work was supported in part by the Academy of Finland.

References

[1] A. A. El Gamal, L. A. Hemachandra, I. Shperling, and V. K. Wei, "Using simulated annealing to design good codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 116–123, 1987.

[2] F. Glover, "Tabu search—Part I," *ORSA J. Comput.*, vol. 1, pp. 190–206, 1989.

[3] I. S. Honkala and P. R. J. Östergård, "Applications in code design," in *Local Search Algorithms*, E. H. L. Aarts and J. K. Lenstra, Eds., New York: Wiley, to appear.

[4] P. R. J. Östergård, "Constructing covering codes by tabu search," *Discrete Appl. Math.*, submitted for publication.

[5] A. Said and R. Palazzo, Jr., "Using combinatorial optimization to design good unit-memory convolutional codes," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1100–1108, 1993.

Reducing the Complexity of Trellises for Block Codes

L. Enrique Aguado-Bayón and Patrick G. Farrell.
Communications Research Group,
School of Engineering, University of Manchester, M13 9PL, UK.
Tel. 44-161-275-4300/4507, Fax. 44-161-275-4355, e-mail farrell@comms.ee.man.ac.uk

Abstract .- **This contribution presents the results of applying two generic algorithms for reducing the complexity of the trellis of a number of binary linear block codes.**

I. INTRODUCTION.

The type of trellis considered was originally defined by Bahl, et al [1] (the BCJR trellis). Later, both Wolf [2] and Massey [3] showed that this type of trellis is useful because it enables Viterbi Algorithm decoding of linear block codes, which in turn means that soft-decision techniques can be simply applied to improve decoding performance. More recently it has been shown that the BCJR trellis is uniquely "minimal" in a number of ways [4,5]. It is most convenient to construct this minimal trellis from the "trellis oriented" form of the code generator matrix, as originally presented by Forney [6], and later developed by McEliece [4,5], who called it the minimal span generator matrix (MSGM). The span of a row of the generator matrix is the number of symbols in the row enclosed between the leftmost non-zero symbol and the rightmost non-zero symbol. The total span of the generator matrix is the sum of the span of the rows of the matrix. Any generator matrix can be reduced to MSGM form by means of elementary row operations (linear combinations and permutations of row). The span of the MSGM is a useful measure of the complexity of the code trellis.

II. COMPLEXITY REDUCTION.

In determining the MSGM of a given code, column permutations are not allowed. It is easily observed however, that column permutation can lead to a lower total span matrix, corresponding to an equivalent linear block code [3,7,8]. To date there is no known algorithm which guarantees that the "globally minimal" MSGM will be found, we would not even know when we have reached it, so we can just give comparative records. It is, however, possible to determine a lower bound on its span, given by:

$$\sum_{i=1}^{n}(k - p_i - f_{i-1})$$

where n and k are the block length and dimension of the code respectively, p_i is the dimension (or a bound) on the dimension of the best "past" code at depth i in the trellis (i.e., the optimum code with block length i and the same distance as the whole code) and f_{i-1} is the dimension of the best future code at depth $i-1$ (i.e., the optimum code with block length $n-i+1$ and the same distance as the whole code) [4,5,8]. The third column of Table 1 gives this lower bound span, together with lower bounds on the numbers of edges and vertices in the code trellis.

The fourth column in Table 1 gives the parameters of the trellises obtained from the MSGM before column permutations. The MSGM is in turn derived from the standard systematic generator matrix of the code by applying a greedy row operation algorithm.

The first algorithm for reducing the total span of the code MSGM by column permutation is based in one devised by Wei Lin [7,9]. The steps of the algorithm are described in [8], and the results obtained are given in the fifth column of table 1. The second algorithm for column permutation is also described in [8], and the results appear in the sixth column of Table 1. This second algorithm is a modified and extended form of Wei Lin's algorithm, based on a simulated annealing technique, which enables improved results to be obtained even for quite large codes. The details of both algorithms will be outlined during presentation of the paper, together with further results.

III. CONCLUSIONS.

Table 1 indicates the significant reduction in the total span, as well as in the other parameters, which can be obtained by means of the two algorithms. In many cases the total span is quite close to the lower bound. For the (32,16) extended BCH and (24,12) extended Golay codes the bound is achieved. This last one coincides with Forney's generator matrix for the code from the cubing construction [6]. It must be considered that the calculated bounds can not be reached sometimes, as McEliece proves [5]. The relation between span and complexity of the trellis is not direct; we conjecture that despite reaching the bound on the span value does not mean reaching it for the elements of the trellis, in the other way round the relation does apply; i.e., the minimum number of elements in the trellis will only be given for a globally minimal span generator matrix.

REFERENCES.

[1] L. R. Bahl, J. Cocke, F. Jelinek and J, Raviv : "Optimal Decoding of Linear Block Codes for Minimising Symbol Error Rate"; IEEE Trans., Vol. IT-20, pp 284-287,1974.

[2] J. K. Wolf : "Efficient Maximum Likelihood of Linear Block Codes Using a Trellis"; IEEE Trans. on Inform. Theory, Vol. IT-24, pp 76-80, January 1978.

[3] J. L. Massey : "Foundation and Methods of Channel Encoding"; Proc. Int. Conf. Inform. Theory and Systems, NTG-Fachberichte, Berlin, 1978.

[4] R. J. McEliece : "The Viterbi Decoding Complexity of Linear Block Codes"; IEEE Int. Symposium on Inform. Theory, Trondheim, Norway, 1994.

[5] R. J. McEliece : "On the BCJR Trellis for Linear Block Codes"; pre-print, September 1994.

[6] G. D. Forney : "Coset Codes - Part II : Binary Lattices and Related Codes"; IEEE Trans. Inform. Theory, Vol. IT-34, pp 1152-1187, September 1988.

[7] S. Dolinar, L. Ekroot, A. Kiely, W. Lin, R. J. McEliece : "Trellis Complexity of Linear Block Codes"; in preparation.

[8] L. E. Aguado-Bayón : "Fast Trellis Decoding for Block Codes"; Transfer Report, University of Manchester, November 1994.

[9] Wei Lin : private communication, January 1994.

Code	Feature.	Lower Bound	Greedy Algo.	Wei Lin Algo.	Simul. Anneal
(23,12)	Edges	1,790	12,284	4,220	3,452
Golay	Vertices	1,214	8,190	3,134	2,558
Code	Span	124	144	133	129
(24,12)	Edges	2,696	16,380	4,348	3,580
Extend.	Vertices	1,790	12,286	3,262	2,686
Golay	Span	136	156	140	136
(31.16)	Edges	3,198	196,604	42,108	6,268
BCH	Vertices	2,174	131,070	31,550	4,670
Code	Span	186	256	231	195
(32,16)	Edges	4,789	262,140	22,780	6,396
Extend.	Vertices	3,198	196,606	17,086	4,798
BCH	Span	202	272	228	202

Table 1: Trellis features for a few codes

Canonical Representation of Quasi-Cyclic Codes[1]

Morteza Esmaeili[†], T. Aaron Gulliver[‡] and Norman P. Secord[*]

[‡]Dept. of Systems & Computer Eng., Carleton University, 1125 Colonel By Dr., Ottawa, ON, Canada K1S 5B6
[†]Dept. of Mathematics & Statistics, Carleton University, 1125 Colonel By Dr., Ottawa, ON, Canada K1S 5B6
[*]Communications Research Centre, 3701 Carling Ave., Box 11490, Station H, Ottawa, ON, Canada K2H 8S2

A linear block code C of length n is called quasi-cyclic (QC) if it is invariant under a cyclic shift of L positions, T^L, where $L < n$. Any cyclic code can be represented by a unique generator polynomial. In this paper we associate with QC-codes a polynomial generator set which is a natural generalization of the generator polynomial of a cyclic code. A canonical generator matrix of a QC-code which is invariant under T^L is introduced which shows the symmetric structure of the n/L-section minimal trellis diagram (MTD) [1, 2, 3]. The state space dimension is nondecreasing on the left half of this trellis. The canonical generator matrix is important because it provides considerable information about the trellis complexity of QC codes as well as the relation between these codes and convolutional codes.

For a linear block code of length n, the interval $[i, j]$, $1 \leq i \leq j \leq n$, is said to be the support interval of a codeword $c = (c_1, \cdots, c_n)$ if $c_i c_j \neq 0$, and $c_l = 0$ if $l < i$ or $j < l$. $j - i + 1$ is defined as the support length of c, and c is said to start at time index i and end at time index j. c is also said to be active in the interval $[i, j-1]$.

A generator matrix of a linear block code is called a trellis oriented generator matrix (TOGM) if no two rows of the matrix either start or end in the same position [1, 3]. Let M be the TOGM of a linear block code C. Denoting the number of rows of M active at time index i by s_i, we define the state complexity of C to be $s = \max\{s_0, s_1, \cdots, s_n\}$.

Let M be a generator matrix for an (Lm, k) QC-code invariant under T^L. Define the $k \times iL$, $1 \leq i \leq m$, matrices M_i such that the j^{th}, $1 \leq j \leq iL$, column of M_i is the same as that of M. Denote the rank of M_i by p_i.

Definition 1 (Cyclic Form Code) *An (n, k) linear block code C is called a cyclic form code if in M (the TOGM of C) for any i, $1 \leq i \leq k$, precisely one row of M has support interval $[i, n-k+i]$. In this case $n-k+1$ is defined as the effective length of C.*

Definition 2 (Smallest Regular Trellis Diagram) *A trellis diagram G of a linear block code C is called a smallest regular trellis diagram (SRTD) of C if: 1) it has the same state complexity as the MTD of C; 2) the number of vertices of G at time indices i and j are equal, $1 \leq i, j < n$; 3) G has the maximum number of identical parallel sub-trellises among all trellises of G which satisfy conditions 1 and 2.*

The following theorem is used to determine the SRTD of a QC-code.

Theorem 1 ([4]) *The smallest regular trellis diagram of an (n, k) linear cyclic form block code C consists of $\max\{1, 2^{2s-k}\}$ structurally identical parallel sub-trellises, where s is the state complexity of the code.*□

The main result of our work is contained in the following theorem.

Theorem 2 (Canonical Generator Matrix) *Let C be an (n, k) QC-code invariant under T^L, and $n = Lm$. If M is a TOGM of C, then*

$$C = \bigoplus_{i=1}^{p_1} C_i, \tag{1}$$

where C_i is a cyclic form code (if it is considered to be a code of length m with codeword components in F^L [3]), and C has TOGM

$$M' = \begin{bmatrix} M(c^1) \\ M(c^2) \\ \vdots \\ M(c^{p_1}) \end{bmatrix}, \tag{2}$$

where $M(c^i)$ is a TOGM of the cyclic form code C_i with leading codeword denoted by c^i. The C_i's are called the canonical components of C. The number of canonical components of C of dimension w, denoted by x_w, is $2p_w - (p_{w-1} + p_{w+1})$.□

The set of polynomials representing the cyclic form canonical components of C is defined as the *polynomial generator set* of C.

Corollary 1 *The m-section MTD of C consists of $2^{2p_1 - p_2}$ identical parallel sub-trellises.*

Decomposing C into cyclic form sub-codes using Theorems 1 and 2, the SRTD of a QC-code C is given in the following corollary.

Corollary 2 *The m-section SRTD of C consists of $2^{\sum_{i=1}^{p_1} a_i}$ structurally identical parallel sub-trellises, with*

$$a_i = \begin{cases} m - m_i + 1 & if \quad m_i \geq \lceil \frac{m+3}{2} \rceil \\ max\{0, 3(m_i - 1) - m\} & if \quad m_i < \lceil \frac{m+3}{2} \rceil \end{cases} \tag{3}$$

where m_i, $1 \leq i \leq p_1$, is the effective length of the i-th canonical component of C.

This provides a decomposition of a QC-code C into its cyclic form sub-codes which can be used to analyze the trellis structure of the QC-code.

[1]This research was supported in part by the Natural Sciences and Engineering Research Council of Canada and the Telecommunications Research Institute of Ontario.

References

[1] F.R. Kschischang and V. Sorokine, "On the trellis structure of block codes," to appear *IEEE Trans. Infor. Theory.*

[2] G.D. Forney, Jr and M.D. Trott, "The dynamics of group codes: state spaces, trellis diagrams and canonical encoders," *IEEE Trans. Infor. Theory*, vol.39, pp. 1491-1513, Sept. 1993.

[3] M. Esmaeili, T.A. Gulliver and N.P. Secord, Trellis complexity of linear block codes via atomic codewords," submitted to *IEEE Trans. Inf. Theory*, Feb. 1995.

[4] M. Esmaeili, T.A. Gulliver and N.P. Secord, "Quasi-cyclic structure of Reed-Muller codes and the smallest regular trellis diagram," submitted to *IEEE Trans. Inf. Theory*, May 1995.

Newton's identities for minimum codewords of a family of alternant codes

Daniel Augot, Daniel.Augot@inria.fr
INRIA, France

Abstract — **We are able to define minimum weight codewords of some alternant codes in terms of solutions to algebraic equations. Particular attention is given to the case of the classical Goppa codes. Gröbner bases are used to solve the system of algebraic equations.**

I. Words of length n

We consider words of length n over $GF(q)$, n being prime to q. A primitive root α is fixed. The word $c = (c_0, \ldots, c_{n-1})$ is identified with the polynomial $c_0 + c_1 X + \ldots + c_{n-1} X^{n-1}$ mod $X^n - 1$. The *Fourier Transform* of $c \in GF(q')^n$, denoted $\phi(c)$, is $A = (A_0, A_1, \ldots, A_{n-1})$, $A_i = a(\alpha^i)$, $i = 0 \ldots n-1$.

Let $c = (c_0, \ldots, c_{n-1}) \in GF(q')^n$. The *locators* of c are $\{X_1, \ldots, X_w\} = \{\alpha^{i_1}, \ldots, \alpha^{i_w}\}$, where $i_1, \ldots, i_w$ are the indices of non zero coordinates of c. The *elementary symmetric functions* of c, denoted by $\sigma_1, \ldots, \sigma_w$, are $\sigma_i = (-1)^i \sum_{1 \le j_1 < \ldots < j_i \le w} X_{j_1} \cdots X_{j_i}$, $i = 1 \ldots w$. The generalized Newton's identities hold: $\forall i \ge 0$, $A_{i+w} + \sigma_1 A_{i+w-1} + \ldots + \sigma_w A_i = 0$.

We introduce the definition of a spectrally defined code:

Définition 1 *Let C be a code in $GF(q')^n$ (or $GF(q)^n$). If there exists l polynomials in n variables $P_1, \ldots, P_l$, such that, for all $c \in GF(q')^n$ (or $GF(q)^n$), c belongs to C if and only if $P(A_0, \ldots, A_{n-1}) = \ldots = P_l(A_0, \ldots, A_{n-1}) = 0$, where $A = \phi(c)$, then the code has a* spectral definition. *The polynomials $P_1, \ldots, P_l$ are the* code spectral equations.

Our result, which is a generalization of a case of a cyclic code [1], is the following theorem:

Théorème 1 *Let C be a code defined by the spectral equations $P_1, \ldots, P_l$. Let $S_C(w)$ be the following system of equations:*

$$P_1(A_0, \ldots, A_{n-1}) = \cdots = P_l(A_0, \ldots, A_{n-1}) = 0$$
$$A_{i+w} + \sigma_1 A_{i+w-1} + \ldots + \sigma_w A_i = 0, \quad i = 0..n-1$$

with indeterminates $\sigma_1, \ldots, \sigma_w, A_0, \ldots, A_{n-1}$. Let $A = (A_0, \ldots, A_{n-1})$ be a solution to $S_C(w)$ (i.e. there exists $\sigma_1, \ldots, \sigma_w$ such that $(\sigma_1, \ldots, \sigma_w, A)$ is a solution), then A is the Fourier Transform of a codeword of weight $\le w$.

II. "Spectral Definition" of some alternant codes

Let $\underline{\alpha} = (\alpha_0, \ldots, \alpha_{n-1}) \in GF(q')^n$ be distinct elements in $GF(q')$, and let $\underline{v} = (v_0, \ldots, v_{n-1})$ be nonzero elements in $GF(q')$. The generalized Reed Solomon code, $GRS_k(\underline{\alpha}, \underline{v})$, is the code whose codewords are $(v_0 F(\alpha_0), \ldots, v_{n-1} F(\alpha_{n-1}))$, for all $F \in GF(q')[X]$, $\deg F < k$.

The alternant code $\mathcal{A}_k(\underline{\alpha}, \underline{v})$ is the $GF(q)$-subfield subcode of $GRS_k(\underline{\alpha}, \underline{v})$. Let $\underline{\alpha} = (\alpha_0, \ldots, \alpha_{n-1}) \in GF(q')^n$ be distinct elements in $GF(q')$, and let $\underline{v} = (v_0, \ldots, v_{n-1})$ be nonzero elements in $GF(q')$. The generalized Reed Solomon code, $GRS_k(\underline{\alpha}, \underline{v})$, is the code whose codewords are $(v_0 F(\alpha_0), \ldots, v_{n-1} F(\alpha_{n-1}))$, for all $F \in GF(q')[X]$, $\deg F < k$. The alternant code $\mathcal{A}_k(\underline{\alpha}, \underline{v})$ is the $GF(q)$-subfield sub-code of $GRS_k(\underline{\alpha}, \underline{v})$.

We consider a partial class of alternant codes, the alternant codes $\Gamma(L, G)$ where $L = \{1, \alpha, \ldots, \alpha^{n-1}\}$, the set of all n-th roots of unity. We denote these codes $\Gamma(\alpha, \underline{v})$. We get that the code spectral equations of $\mathcal{A}_k(\alpha, \underline{v})$ are

$$\begin{cases} \sum_{i+j=t \bmod n} A_i H_j = 0, \ t = 0 \ldots n-k-1 \\ A_{iq \bmod n} = A_i^q, \ i = 0 \ldots n-1 \end{cases}$$

where H is the Fourier Transform of h defining the dual of the $GRS_k(\underline{v})$.

III. A short Goppa code

Since classical Goppa codes with support $L = \{\alpha^i, \ i = 0 \ldots n-1\}$ are alternant codes, we are also able to construct spectral equations for these codes. As an example we study the Goppa code of length 32, with defining polynomial $g(x) = x^3 + x + 1$. We index codewords c in the following way: $c = (c_\infty, c_0, \ldots, c_{30})$, where the defining set of the Goppa code is $L = \{0, 1, \alpha, \ldots, \alpha^{30}\}$. Since our result works for a support of length n prime to 2, we first consider the sub-code C_{31} of C which is the shortened code with respect to the coordinate c_∞. This code is also a Goppa code with support $L_{31} = \{1, \alpha, \ldots, \alpha^{30}\}$ and defining polynomial $g(X)$. Thus writing the system $S_{C_{31}}(7)$, we get equations for codewords such that $c_\infty = 0$. Computing a Gröbner basis of the system, we get 105 solutions. Next, we want to study minimum weight codewords such that $c_\infty \neq 0$. The parity check matrix for C is

$$G = \begin{bmatrix} 1 & g(\alpha^0)^{-1} & \ldots & g(\alpha^{30})^{-1} \\ 0 & \alpha^0 g(\alpha^0)^{-1} & \ldots & \alpha^{30} g(\alpha^{30})^{-1} \\ 0 & (\alpha^0)^2 g(\alpha^0)^{-1} & \ldots & (\alpha^{30})^2 g(\alpha^{30})^{-1} \end{bmatrix}.$$

We search for words $c_0, \ldots, c_{30}$ of weight 6, of length 31 such that $G' c^t = (1, 0, \ldots, 0)^t$. where G' is the parity check matrix for C_{31}. Thus the spectral equations for these codewords are:

$$\begin{cases} \sum_{i+j=0 \bmod 31} A_i H_j = 1 \\ \sum_{i+j=t \bmod 31} A_i H_j = 0, \ t = 1, 2 \\ A_{2i \bmod 31} = A_i^2, \ i = 0 \ldots 30 \end{cases}$$

These equations, plus the Newton's identities for the weight 6, gives equations for codewords of C of weight 7 whose support is not included in $[0, 30]$. The Gröbner basis gives 23 solutions, thus 128 codewords of weight 7 for the whole code C, as in the table of [2, p344].

References

[1] D. Augot. Algebraic characterization of minimum weight codewords of cyclic codes. In *Proceedings IEEE, ISIT'94*, Trondheim, Norway, June 1994.

[2] F.J. Mac Williams and N.J.A. Sloane. *The Theory of Error Correcting Codes*. North-Holland, 1986.

Information Theoretical Lower Bounds for Unconditionally Secure Group Authentication

Christian Gehrmann[1]
Department of Information Theory, Lund University
Box 118, S-221 00 Lund, Sweden

Abstract — **The single sender single receiver authentication model was extended by Desmedt and Frankel [1] to the case where certain groups of persons are able to sign a message. The problem is further developed and discussed in [2]. The unconditionally secure group authentication problem was formulated and investigated using the generalized vector space construction in [3]. We give information theoretic bounds on the security of a group authentication scheme and propose a construction based on the Shamir secret sharing scheme and maximum rank distance codes (MRD-codes).**

I. Summary

Let a secret key K be shared among a set of participants P such that certain subsets of participants are able to compute the authentication tag $Z = F(M, K)$ of the message M , where F is the authentication function. Denote by $\mathcal{M}$, the set of messages and by $\mathcal{K}$ the set of secret keys. The receiver is also assumed to be the dealer of the secret key. To share a secret key $K \in \mathcal{K}$ he uses a secret sharing scheme to give each participant the share K_i. Denote by Γ a monotone access structure, i.e., the set of qualified groups with monotonic properties. To authenticate a message each participant i in a qualified group, $X \in \Gamma$ first calculates

$$F_i^X(M, K_i)$$

and sends this to a (not necessarily trustable) combiner who evaluates $Z = C_X(F_i^X(M, K_i); i \in X$, which equals $F(M, K)$. The output Z of the combiner is the authentication tag, which together with the message, is sent to the receiver, who can check the correctness of a message by calculating $F(M, K)$ directly.

As in ordinary single authentication schemes we measure the security of a scheme by the probabilities of successful impersonation and substitution. Denote by P_I the worst case probability of finding a correct authentication tag given the knowledge of the shares of any non-qualified group. The probability of successful substitution attack is denoted by P_S and is defined as the worst case probability of finding a correct authentication tag given the knowledge of the shares from a non-qualified group.

As in ordinary secret sharing we call a scheme perfect if $Y \notin \Gamma, H(K|Y) = H(K)$. Using results on secret sharing schemes [4] we are able to prove the following theorem on P_I and P_S.

Theorem 1 *Let $Y \notin \Gamma$ and $K_i \cup Y \in \Gamma$. For a prefect scheme*

$$P_I \geq \max_{K_i, Y} 2^{-I(K_i; Z|Y) + H(K_i|YK)}, \quad (1)$$

$$P_S \geq \max_{K_i, Y} 2^{-H(K_i|YZ)}. \quad (2)$$

We especially consider the situation where the combiner just adds the partial authentication tags and where the authentication function F is linear. Let the message M be represented as an $r \times n$ matrix over $\mathbb{F}_q$ and let the secret key k be a vector of length n over $\mathbb{F}_q$. Thus,

$$F(M, k) = Mk.$$

Furthermore, assume that the dealer uses the Shamir scheme [5] to give each participant a share $k_i \in \mathbb{F}_{q^n}$. The secret key k may then be calculated as a linear combination

$$k = \sum_{i \in X} \beta_i k_i$$

of t shares from a qualified group X, i.e., at least t participants. By restricting the message matrix to matrices of the form $[I_r M]$, where I_r is the $r \times r$ identity matrix and thus is M an $r \times (n-r)$ matrix we translate our scheme to one equivalent to an authentication function in the well-known form $k_0 + g_M(k_1)$, where k_0 is the vector consisting of the r first elements and k_1 a vector consisting of the $n-r$ last elements of k. Furthermore, $g_M : \mathbb{F}_{q^{n-r}} \mapsto \mathbb{F}_{q^r}$ belongs to a set of linear functions. For this situation we are able to prove the following theorem:

Theorem 2 *Denote by $\mathcal{A}$ the set of matrices*

$$\mathcal{A} = \{M - \hat{M}; M \in \mathcal{M}, M \neq \hat{M} \in \mathcal{M}\}.$$

Then for the group authentication scheme described above

$$P_I = q^{-r}, \quad (3)$$

$$P_S = q^{-d}, \quad (4)$$

where $d = \min_{A \in \mathcal{A}} \text{rank} A$.

This relates the problem to codes for the rank metric [6] and construction for A^2-codes made by Johansson [7]. The above result can also be obtained with the general technique in [3].

[1] This work was supported by the TFR Grant 94-457

References

[1] Y. Desmedt and Y. Frankel, "Shared generation of authentication signatures", *Proceedings of Crypto '91*, 1991, pp. 457-469.

[2] Y. Desmedt, "Threshold Cryptography", *European Trans. on Telecommunication*, Vol. 5, 1994, pp. 449-457.

[3] M. van Dijk, C. Gehrmann and B. Smeets, "Unconditionally Secure Group Authentication", submitted to Eurocrypt '95.

[4] R. M. Capocelli, A. De Santis, L. Gargano and Vaccaro U, "On the size of Shares for Secret Sharing Schemes". *Journal of Cryptology*, 6:157-167, 1993.

[5] A. Shamir, "How to share a secret", *Commun. ACM*, Vol. 22, 1979, pp. 612-613.

[6] E. M. Gabidulin, "Theory of Codes with Maximum Rank Distance", *Problems of Information Transmission*, Vol. 21, no. 1, pp. 1-12, July 1985, (Russian Original, January-March, 1985).

[7] T. Johansson, *Contributions to Unconditionally Secure Authentication*, PhD thesis, Lund Dec. 1994.

Spectral Properties and Information Leakage of Multi-Output Boolean Functions

A. M. Youssef and S. E. Tavares
Department Of Electrical and Computer Engineering
Queen's University
Kingston, Ontario, Canada, K7L 3N6

Abstract — In this paper, we extend the concept of information leakage in [1], [2] to the case of multi-output boolean functions. A spectral characterization of multi-output boolean function is given. This result is used to express different forms of information leakage of multi-output boolean function in terms of the Walsh transform of every linear combination of its output coordinates. Conditions on the Walsh transform of the multi-output boolean function are given, which imply that the function satisfies certain cryptographic properties of interest such as balance, correlation immunity, Strict Avalanche Criterion (SAC) , higher order SAC, Propagation Criterion (PC), higher order PC, and Perfect non-linearity.

Definitions:

Throughout this paper, let Y be the output of a boolean function $f(X) : \mathrm{Z}_2^n \rightarrow \mathrm{Z}_2^m$, then

• The Walsh transform of the linear combination of its output coordinates $c.f(X)$ is defined as[1]

$$F_c(w) = \frac{1}{2^{n/2}} \sum_{X \in \mathrm{Z}_2^n} (-1)^{c.f(X)-w.X}.$$

• The static information leakage of Y, given input subvector X_k, is defined by:

$$SL(Y; X_k) = m - H(Y|X_k).$$

Similarly the dynamic information leakage of $\triangle Y$, given the input change vector $\triangle X$ is defined by:

$$DL(\triangle Y; \triangle X) = m - H(\triangle Y|\triangle X)$$

where $\triangle Y = Y(X) \bigoplus Y(X \bigoplus \triangle X)$.

• The self static/dynamic information leakage of Y is defined as:

$$SSL(Y) = m - H(Y)$$
$$SDL(Y) = m - H(\triangle Y) .$$

Results:

Let Y be the output of a boolean function $f(X)$ then for $N_y = \#\{X \in \mathrm{Z}_2^n | f(X) = y\}$, $N_{\dot{x}y} = \#\{X \in Z_2^n \mid X_k = \dot{x}, Y = y\}$ and $N_{\triangle x \triangle y} = \#\{X \in \mathrm{Z}_2^n \mid f(X \bigoplus \triangle x) \bigoplus f(X) = \triangle y\}$, $\dot{x} \in \mathrm{Z}_2^k$, $\triangle x \in \mathrm{Z}_2^n$, $y \in \mathrm{Z}_2^m$, $\triangle y \in \mathrm{Z}_2^m$, then one can prove that :

$$N_y = 2^{n/2-m} \sum_{c \in \mathrm{Z}_2^m} F_c(0)\,(-1)^{c.y},$$

$$N_{\dot{x}y} = 2^{n/2-m-k} \sum_{\substack{\dot{w} \in \mathrm{Z}_2^k \\ c \in \mathrm{Z}_2^m}} F_c(w^o)(-1)^{\dot{w}.\dot{x}+c.y},$$

$$N_{\triangle x \triangle y} = \frac{1}{2^m} \sum_{\substack{c \in \mathrm{Z}_2^m \\ w \in \mathrm{Z}_2^n}} F_c^2(w)(-1)^{\triangle x.w+\triangle y.c},$$

where w^o denotes the n-dimensional vector obtained by completing the k-dimensional subvector $\dot{w}$ with zeros. For example if $n = 6$, $\dot{x} = \{x_0, x_2, x_5\}$ then $w^o = \{\dot{w}_0, 0, \dot{w}_2, 0, 0, \dot{w}_5\}$. Using the above results we get:

Theorem 1:

Let Y be the output of a boolean function $f(X)$ then the different forms of information leakage of Y can be expressed as:

$$SSL(Y) = m - \sum_{y \in \mathrm{Z}_2^m} \frac{N_y}{2^n} log_2\left(\frac{2^n}{N_y}\right)$$

$$SL(Y; X_k) = m - 2^{-k} \sum_{\substack{y \in \mathrm{Z}_2^m \\ \dot{x} \in \mathrm{Z}_2^k}} \frac{N_{\dot{x}y}}{2^{n-k}} log_2\left(\frac{2^{n-k}}{N_{\dot{x}y}}\right)$$

$$DL(\triangle Y; \triangle X) = m - 2^{-n} \sum_{\substack{\triangle x \in \mathrm{Z}_2^n \\ \triangle y \in \mathrm{Z}_2^m}} \left(\frac{N_{\triangle x \triangle y}}{2^n}\right) log_2\left(\frac{2^n}{N_{\triangle x \triangle y}}\right)$$

where $N_y, N_{\dot{x}y}, N_{\triangle x \triangle y}$ are given by the equations above.

Let criterion "C" be any of the following : balance, correlation immunity , Strict Avalanche Criterion (SAC) , higher order SAC , Propagation Criterion (PC), higher order PC , or perfect nonlinearity.

Theorem 2:

If Y is the output of a multi-output boolean function then Y satisfies criterion C if and only if every non zero linear combination of its output coordinates satisfies criterion C.

References

[1] R. Forré. Methods and instruments for designing S-boxes. *J. of Cryptology, Vol .2, No.3 pp. 115–130*, 1990.

[2] M. Zhang, S.E. Tavares, and L.L. Campbell. Information leakage of boolean functions and its relationship to other cryptographic criteria. *Proc. 2nd ACM Conf. on Computer and Commun. Security, Fairfax, Virgina, pp. 156-165.*, 1994.

[1] To be precise, this is the Walsh Transform of the function $(-1)^{c.f(X)}$

Cryptographic Redundancy and Mixing Functions

Oliver Collins
University of Notre Dame
South Bend, IN 46556

This talk studies the application of structures based on error correcting codes to systems where the major requirement is not error control but secrecy. In many cases the same code can achieve both error control and secrecy. The first section of the talk describes an optimal construction for combining multiple semi-secure channels, e.g., a bundle of fiber-optic cables or wires running through individual conduits, into a single channel with much higher security. Usually the security of a communications channel cannot be guaranteed, only promised with a high degree of probability. The first section shows how to combine semi-secure channels in such a way that any predetermined number may be compromised before information is revealed.

Semi-secure channels can take on many forms. Any conventional or public key cryptosystem used over a public channel is only semi-secure, since there is currently no method of proving that any particular system purporting to have computational security is genuinely hard to break. Other examples of semi-secure channels are copper wires running through separate conduits pressurized with gas to make tampering easy to detect and fiber optic cables, which are intrinsically fairly difficult to tap.

Clearly, the maximum possible secure capacity of a set of semi-secure channels is just the sum of the capacities of those channels that are in fact secure. The first theorem in this talk states that this bound on total secure capacity is, in fact, achievable.

Theorem 1: Given a set of N channels, each with capacity C, any K of which can be intercepted by the enemy, it is possible to form a composite channel of capacity (N-K)C which is completely secure, even if neither the sender nor the receiver knows which channels have been intercepted.

The very simple constructive proof uses an (N,K) maximum distance separable (MDS) code which can, by definition, correct N-K erasures. The K inputs to the encoder come from a source of perfect randomness, e.g., a thermal noise source followed by a hard limiter. The symbols sent over the first K channels are the first K symbols produced by the encoder. If the encoder is systematic, then these may be just the random input symbols themselves. The symbols sent over the remaining N-K channels are formed by adding one symbol of information to be transmitted to each of the remaining symbols in the encoder output and then sending each of these sums over one of the remaining channels.

The concept of a mixing function was introduced in Reference 1 to improve the security against ciphertext-only attack of a single cryptosystem operating over a single channel by destroying the local statistics which are essential to assaults based on letter or word frequency. The idea is to create a function which mixes text so that small groups of letters appear totally random, i.e., have maximum entropy. The talk proceeds to show how mixing and scrambling functions formed from error correcting codes can be used to enhance the security of trunked communications circuits and conventional cryptographic systems which depend, for their security, on unproved assertions about computational difficulty.

The last segment of the talk presents a concept for applying information theoretic security to spread spectrum communications and ranging systems so that even an intended recipient of the message will not be able to jam the signal. Airplane instrument landing systems and other navigation signals are an obvious potential application of this idea.

1) C.E. Shannon, "Communication Theory of Secrecy Systems" Bell System Technical Journal, vol. 28 October 1949 pp.711-715

Design of The Extended-DES Cryptography [1]

Haeng-Soo Oh* Seung-Jo Han**
* Dept. Elect. Eng., Dongshin Junior col.
Kwangju City, SOUTH KOREA.
** Dept. Elect. Eng., Chosun University.
Kwangju City, SOUTH KOREA.

Abstract - **In order to solve the problem that the DES may be attacked by the differential cryptanalysis, this paper aims to design the Extended-DES, first by breaking a block that is composed of 96 bits into 3 sub-blocks, then performing different f functions on each of the 3 sub-blocks, and finally increasing the S_1-S_8 of the S-box to S_1-S_{16} which makes it less vulnerable to attack by differential cryptanalysis.**

I. SUMMARY

In order to increase the cryptographic security of the DES, this paper offers some suggestions as follows.

The 128 key bits that are inputed to increase the key from 56 bits to 112 bits, are each divided into 64 bits, K_1,K_2. According to the key schedule of the DES, there are 64 bits in K_1 on the left, and then after removing 8 parity bits, through the Permuted Choice 1(PC-1), 56 bits are outputed. The 56 bits are then divided into 28 bits on the left and on the right. Then the sub-key is shifted, according to the number of times of the left shift of the key schedule in each round. They produce the $K_{1,1}$-$K_{1,16}$ that are the sub-keys of 48 bits through the Permuted Choic 2(PC-2). K_2, the 64 bit on the right, and $K_{2,1}$-$K_{2,16}$, the sub-keys of 48 bits, are also produced by the key schedule. As a result, when applying $K_{1,i}$ and $K_{2,i}$ to the f functions on the left and the right, the following encryption and decryption formula is derived:

Encryption : $A_i=B_{i-1}$
$B_i=C_{i-1}\oplus f(B_{i-1},K_{2,i})$
$C_i=A_{i-1}\oplus f(B_{i-1},K_{1,i})$

Decryption : $A_{i-1}=C_i\oplus f(A_i,K_{2,i})$
$B_{i-1}=A_i$
$C_{i-1}=B_i\oplus f(A_i,K_{1,i})$

As in the figure, the A_{16} and B_{16} of the last round during the encryption process should be interchanged while the decryption process remains the same, except that A_1 and B_1 should be exchanged and inputed into the sub-block. The key has to be inputed in the reverse order of $K_{1.16}$,$K_{1.15}$,$K_{1.14}$,....,$K_{1.1}$ with $K_{1.1}$ on the left and $K_{2.1}$ on the right. S_1-S_8 on the left and S_9-S_{16} on the right should also be interchanged. Against an attack by differential cryptanalysis, the iteration number of the f function performed in each sub-block during 16 rounds should be different, which creates a decreased probability of having the feature of N round. During the DES, the f function performed in the sub-block is repeated 8 times; in the Extended-DES, it is repeated 11 times during the performance of A_0 to B_{16}, 10 times during from B_0 to A_{16}, and 11 times during C_0 to B_{16}. Additionally, it is known that the f function is repeated differently according to each of the sub-blocks. Therefore, as the DES has the same iteration number of f functions according to each sub-block, it can easily be attacked by the differential cryptanalysis; but because the Extended-DES has a different iteration number of f functions in each sub-block, it can be said that it resists differential cryptanalysis. Also, in the Extended-DES, the S_1-S_8 of the S-box is enlarged to S_1-S_{16} and the S-box is chosen when each entry is suitable both for the SAC, and the correlation coefficient condition.

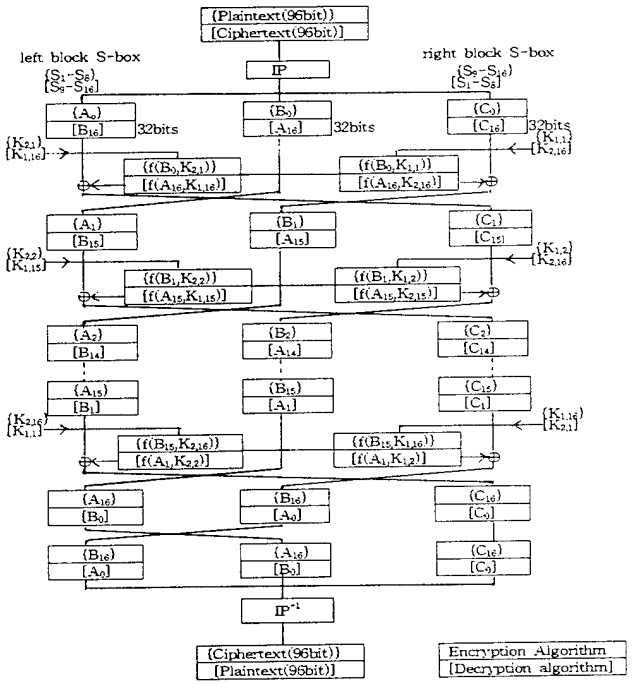


Fig. Algorithm of the Extended-DES

To improve the cryptographic security in the Extended-DES design, each entry in the S-box is arranged randomly, so that S-box, which agrees with the condition as well as the correlation coefficient is increased to S_1-S_{16}. The condition that the probability of the output bit, j^{th}, being changed is $P_{i,j}=X_j/2^n$ when the input bit of i^{th} is complemented. The nearer $P_{i,j}$ approaches to 0.5, the closer the S-box is to the condition of SAC. The result of the simulation shows that the Extended-DES agrees with the condition of SAC better than the DES in that the $P_{i,j}$ of the S-box in the Extended-DES approaches nearer to 0.5. Then, the correlation coefficient between each bit of the S-box output must be independent, and is considered to be the better design when the correlation coefficient($-1 \leqq | \rho_{i,j}(k) | \leqq 1$) approaches zero. In this paper, the $\rho_{i,j}$ of the S-box in the Extended-DES approaches zero nearer than the $\rho_{i,j}$ of the DES. Consequently, it is known that when designing the S-box, SAC and the correlation coefficient, the S-box of the Extended-DES is better than the DES's. Therefore, the Extended-DES developed in this paper has been implemented into software and it has been verified that it's cryptographic security is superior to that of the DES.

REFERENCES

[1] E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystem," Weizmann Institute of Science, Technical Report, Rehovot, Israel, 19 July 1990.

[2] R. Forre, "The Strict Avalanche Criterion: Special Properties of Boolean Function and an Extended Definition," Proc. of Crypto'88, Springer-Verlag, pp.167-173, 1989.

[1]This work was supported by a Chosun University Grant.

Constructions of asymmetric authentication systems

Thomas Johansson[1]

Department of Information Theory, Lund University, Box 118, S-221 00 Lund, Sweden.

Abstract — **Constructions of asymmetric authentication systems based on families of mappings with the vector space property are considered.**

I. Introduction

Simmons [1] introduced *asymmetric authentication systems* when he extended conventional authentication codes to codes with arbitration, called A^2-codes. It is now the notion for any authentication system where the participants possess different keys which, in some way, are dependent. Several different systems of this kind have been considered [2], [3], [4].

II. A^2-codes and vector spaces of mappings

Let $\mathcal{F} = \{f_i\}$ be a set of functions $f_i : \mathcal{S} \to R$, where R is a ring. Let $\mathcal{F}$ have the vector space property, i.e., $c_1 f_i + c_2 f_j \in \mathcal{F}$ for any $c_1, c_2 \in R$ and any $f_i, f_j \in \mathcal{F}$, $i \neq j$. We randomly choose $f, f_1, f_2 \in \mathcal{F}$ and $z \in R$ in such a way that $f = f_1 + z f_2$. The A^2-code is now given as follows. The transmitter has as his key E_T the pair (f_1, f_2) and the receiver has as his key E_R the pair (f, z). To send the source state $s \in \mathcal{S}$ the transmitter generates the message $m = (s, f_1(s), f_2(s))$. The receiver receives $m = (s, m_2, m_3)$ and checks that $f(s) = m_2 + z m_3$. In a correct transmission, $m_2 = f_1(s), m_3 = f_2(s)$, and thus $f(s) = f_1(s) + z f_2(s)$.

III. Broadcast authentication systems

The idea of broadcast authentication systems was first introduced by Desmedt and Yung [2]. We generalize their ideas to include any specified attack. The set of participants $\mathcal{P}$ consists of a transmitter T, a set of receivers $\mathcal{R} = \{R_i\}$, and possibly a set of other participants $\mathcal{O} = \{O_i\}$. The transmitter T will generate a message m, and it can be addressed to any $R_i \in \mathcal{R}$, or to some specified subset of $\mathcal{R}$. The address is contained in the source state s, and changing it implies a substitution attack. We also specify: how disputes are to be solved; collaboration sets $\mathcal{C} = \{\mathcal{C}_x\}$ (which collusions of cheating participants exist against participant x); verification sets $\mathcal{V}_x$ (which participants must be able to verify messages to a certain receiver x);

We describe the existing attacks. There are two classes of attacks. The first class of attacks is some subset of participants trying to get a fraudulent message accepted by some receiver, i.e., trying to cheat a receiver. We separate into two cases, depending on whether the transmitter is included in the cheating subset or not. We denote the probability of success as $P_I(\mathcal{C})$ for the impersonation, and $P_S(\mathcal{C})$ for the substitution attack, when the transmitter is not included in the cheating subset. If the transmitter is included, we denote the probability of success as $P_T(\mathcal{C})$.

The second class of attacks is a subset collaborating, claiming to have received a message that was never sent and thus trying to frame the transmitter. Here we have both the impersonation case and the substitution case. We denote the probability of success as $P_{R_0}(\mathcal{C})$ and $P_{R_1}(\mathcal{C})$, respectively.

[1]This work was supported in part by the Swedish Research Council for Engineering Sciences under Grant 94-457

Let $\mathcal{M}(e_t)$ be the set of messages that the transmitter can generate when he is in possession of the key e_t, and let $e(\mathcal{L})$ be the set of keys for a subset $\mathcal{L}$ of participants. The definitions of the probabilities of success in the different attacks are:

$$P_I(\mathcal{C}) = \max_{R_i} \max_{\substack{\mathcal{L}\in\mathcal{C}_{R_i} \\ T\notin\mathcal{L}}} \max_{e(\mathcal{L}),m} P(m \text{ accepted by } R_i | e(\mathcal{L})),$$

$$P_S(\mathcal{C}) = \max_{R_i} \max_{\substack{\mathcal{L}\in\mathcal{C}_{R_i} \\ T\notin\mathcal{L}}} \max_{\substack{e(\mathcal{L}),m,m' \\ m\neq m'}} P(m' \text{ accepted by } R_i | m, e(\mathcal{L})),$$

$$P_T(\mathcal{C}) = \max_{R_i} \max_{\substack{\mathcal{L}\in\mathcal{C}_{R_i} \\ T\in\mathcal{L}}} \max_{\substack{e(\mathcal{L}),m \\ m\notin\mathcal{M}(e_t)}} P(m \text{ accepted by } R_i | e(\mathcal{L})),$$

$$P_{R_0}(\mathcal{C}) = \max_{\mathcal{L}\in\mathcal{C}_T} \max_{e(\mathcal{L}),m} P(m \in \mathcal{M}(e_t) | e(\mathcal{L})),$$

$$P_{R_1}(\mathcal{C}) = \max_{\mathcal{L}\in\mathcal{C}_T} \max_{\substack{e(\mathcal{L}),m,m' \\ m\neq m'}} P(m' \in \mathcal{M}(e_t) | m \in \mathcal{M}(e_t), e(\mathcal{L})).$$

An important class of broadcast authentication systems is a system where $\mathcal{P} = \{T, R_1, R_2, \ldots, R_n, A\}$, and such that: an honest arbiter A makes decisions in case of a dispute; all attacks from any subset of at most k participants (excluding the arbiter) exist; the verification set is $\mathcal{V}_i = \{R_1, R_2, \ldots, R_n, A\}$, $\forall i = 1, \ldots, n$. We call such a system an (n, k)-threshold USDS [4]. Such system can be constructed by choosing $f^{(i)}, f_j \in \mathcal{F}$ and $z_j^{(i)} \in R$ such that

$$\begin{pmatrix} 1 & z_2^{(1)} & \ldots & z_{k+1}^{(1)} \\ 1 & z_2^{(2)} & \ldots & z_{k+1}^{(2)} \\ & \vdots & & \\ 1 & z_2^{(n)} & \ldots & z_{k+1}^{(n)} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{k+1} \end{pmatrix} = \begin{pmatrix} f^{(1)} \\ f^{(2)} \\ \vdots \\ f^{(n)} \end{pmatrix}, \quad (1)$$

where any $(k+1)$ rows are linearly independent. The transmitter's key is $E_T = (f_1, \ldots, f_{k+1})$ and receiver i has key $E_{R_i} = (f^{(i)}, z_2^{(i)}, \ldots, z_{k+1}^{(i)})$. The transmitter sends the message $m = (s, f_1(s), \ldots, f_{k+1}(s))$ and receiver i checks that

$$f_1(s) + z_2^{(i)} f_2(s) + \cdots + z_{k+1}^{(i)} f_{k+1}(s) = f^{(i)}(s).$$

An example of performance is given by the following theorem.

Theorem 1: *Let $\mathcal{F} = \{f(s); f(s) = as + b, \forall a, b \in \mathbb{F}_q\}$. Then (1) is an (n, k)-threshold USDS, where*

$$P_I(\mathcal{C}) = P_S(\mathcal{C}) = P_T(\mathcal{C}) = P_{R_0}(\mathcal{C}) = P_{R_1}(\mathcal{C}) = 1/q.$$

We further consider collaboration sets which are not of threshold type.

References

[1] G.J. Simmons, "A Cartesian Product Construction for Unconditionally Secure Authentication Codes that Permit Arbitration", *Journal of Cryptology*, vol. 2, no. 2, 1990, pp. 77–104.

[2] Y. Desmedt, M. Yung, "Arbitrated unconditionally secure authentication can be unconditionally protected against arbiter's attack", *Proceedings of Crypto'90*, LNCS 537, pp. 177–188.

[3] Y. Desmedt, M. Yung, "Multi-receiver/Multi-Sender network security: Efficient authenticated multicast/feedback", Infocom, May 1992.

[4] T. Johansson, "Contributions to unconditionally secure authentication", Ph.D. Thesis, Lund 1994.

Two Simple Schemes for Access Control

Man Yiu Chan and Raymond W. Yeung
Department of Information Engineering, The Chinese University of Hong Kong, N.T., Hong Kong

I. Introduction and Summary

In an *open* information system, all the information in the system are known by the public. For such a system, the users are responsible to encipher their own information in a way that they can be accessed by authorized users only. Let $U = \{U_1, \ldots, U_n\}$ be the set of users in the system. Associated with each U_i is an *authorization list* $A_i \subset U$ such that U_i can access the information of U_j if and only if $U_i \in A_j$.

In this paper, we propose two simple access control schemes. For the first scheme (Scheme 1), for any $1 \leq i, j, k \leq n$,

$$(U_k \in A_j \text{ and } U_j \in A_i) \Rightarrow U_k \in A_i. \quad (1)$$

In other words, if U_k can access the information of U_j and U_j can access the information of U_i, then U_k can accessed the information of U_i. This is called *hierarchical* accessibility. For the second scheme (Scheme 2), there is no constraint on the authorization lists. To our knowledge, this is the first scheme in the literature that supports arbitrary accessibility.

II. Scheme 1: Hierarchical Accessibility

If $A_i, i = 1, \ldots, n$ satisfy (1), the elements in U has a partial order, with $U_i \geq U_j$ signifies that U_i can access the information of U_j. U_i is called a *predecessor* of U_j, and U_j is called a *successor* of U_i. The scheme we propose is as follows. Each U_i has an encryption algorithm E_i and a decryption algorithm D_i which are parametrized by e_i and d_i, respectively, where e_i is publicly revealed and d_i is kept secret to U_i. (It is assumed that the class of encryption/decryption algorithms that E_i and D_i belong to is publicly known, and E_i and D_i are completely characterized by e_i and d_i, respectively.) Further, the encryption/decryption pair (E_i, D_i) forms a public key cryptosystem, i.e.,

(PK1) For each message m, $D_i(E_i(m)) = m$.

(PK2) E_i and D_i are easy to compute.

(PK3) It is practically impossible to find a decryption algorithm D_i' from E_i such that $D_i'(E_i(m)) = m$ for all m.

The scheme works as follows. Let m_i be the information of U_i. Each U_i enciphers m_i as $E_i(m_i)$ and reveals it publicly. Let U_j be an immediate predecessor of U_i (U_i can have more than one immediate predecessor). In order that U_j can access the information of U_i, U_i enciphers d_i as $E_j(d_i)$ and reveals it publicly. Then U_j can recover d_i as $D_j(E_j(d_i))$ and then recover m_i as $D_i(E_i(m_i))$.

Now suppose U_j is an immediate successor of U_k and U_i is an immediate successor of U_j. Then U_k can recover d_j as described above. With d_j, U_k can also recover d_i as $D_j(E_j(d_i))$, since $E_j(d_i)$ is publicly known. With d_i, U_k can then recover m_i. Likewise, U_k can access the information of any of its successors.

Different hierarchical access schemes have been proposed in the literature ([1,2], [4]-[7]). All these schemes have the common property that key management in the system is performed by a central authority. By contrast, our scheme is completely decentralized and does not need a central authority. In addition, our scheme has the following advantages:

1. Users are allowed to choose their own keys.
2. It is not necessary to deliver keys to the users in a secure way (cf. for example [1,2]).
3. The amount of storage required is proportional to the total number of immediate successors in the system.
4. Insertion and deletion of users are simple, and do not affect the encryption and decryption procedures of existing users.
5. Update of encryption and decryption keys is simple.

III. Scheme 2: Arbitrary Accessibility

We assume that $A_i, i = 1, \ldots, n$ are arbitrary. In this scheme, each U_i has two encipher algorithms E_{1i} and E_{2i}, and two decipher algorithms D_{1i} and D_{2i}, which are parameterized by e_{1i}, e_{2i}, d_{1i} and d_{2i}, respectively. e_{2i} are revealed publicly, while e_{1i}, d_{1i} (called the *file decryption key*) and d_{2i} are kept secret to U_i. (E_{2i}, D_{2i}) forms a public key cryptosystem, while (E_{1i}, D_{1i}) forms a conventional cryptosystem.

The scheme works as follows. Each U_i enciphers its information m_i as $E_{1i}(m_i)$ and reveals it publicly. For each $U_j \in A_i$, U_i enciphers d_{1i} as $E_{2j}(d_{1i})$ and reveals it publicly. Then U_j can recover d_{1i} as $D_{2j}(E_{2j}(d_{1i}))$, and then recover m_i as $D_{1i}(E_{1i}(m_i))$. It is easy to see that if $U_j \notin A_i$, then U_j cannot access the information of U_i.

In Scheme 1, a user uses the same encryption/decryption pair for both its own information and the decryption keys of its immediate successors. For this reason, a user can access the information of all its successors. In Scheme 2, however, two different encryption/decryption pairs are used for its own information and the file decryption keys of those users whose information it can access. This arrangement breaks up the hierarchical structure of the scheme.

To our knowledge, this is the first scheme in the literature that supports arbitrary accessibility. This scheme enjoys all the advantages of Scheme 1 except that the amount of storage required is proportional to $\sum_{i=1}^{n} |A_i|$, which is upper bounded by n^2.

References

[1] S. G. Akl and P. D. Taylor, "Cryptographic solution to a problem of access control in a hierarchy," *ACM Trans. Comput. Syst.*, vol. 1, 239-248, 1983.

[2] C.-C. Chang, R.-J. Hwang, and T.-C. Wu, "Cryptographic key assignment scheme for access control in a hierarchy," *Inform. Syst.*, vol. 17, 243-247, 1992.

[3] N.-Y. Lee and T. Hwang, "A pseudo-key scheme to dynamic access control in a hierarchy," 1993 International Symposium on Communications, Hsinchu, Taiwan, R.O.C., 7-10 Dec 1993.

[4] S. J. Mackinon, P. D. Taylor, H. Meijer, S. G. Akl, "An optimal algorithm for assigning cryptographic keys to access control in a hierarchy," *IEEE Trans. Comput.*, vol. c-34, 707-802, 1985.

[5] R. S. Sandhu, "Cryptographic implementation of a tree hierarchy of access control," *Inform. Proc. Lett.*, vol. 27, 95-98, 1988.

[6] Y. Zheng, T. Hardjono, and J. Pieprsyk, "Sibling intractable function families and their applications," *Proc. AsiaCrypto - 91*, 67-74, 1992.

Attacks on Tanaka's Non-interactive Key Sharing Scheme

Sangjoon Park Yongdae Kim Sangjin Lee Kwangjo Kim
Electronics and Telecommunications Research Institute,
P.O.Box 106, Yusong, Taejon, 305-600, KOREA

Abstract — **We propose two different methods for attacking Tanaka's IDNIKS presented in SCIS'94. One is to find the secret informations using public parameters, and the other is to find the center's secret keys by collusion.**

I. Introduction

IDentity-based Non-Interactive Key Sharing(IDNIKS) scheme was first proposed by Blom[1]. Since then, there have been many works for IDNIKS [2], [3], but, many of them were found to be breakable by collusion attacks [4], [6].

In SCIS'94, H.Tanaka[5] proposed the new IDNIKS which could be easily implemented. We propose two methods for attacking Tanaka's IDNIKS. The first method is to find the secret informations using public parameters of the center, and the other is to find the center's secret keys by 8 collaborators.

II. Tanaka's IDNIKS

At first, a center chooses RSA modulus $N(=PQ)$, one-way hash function f, random number e, e_1, e_2 satisfying $\gcd(e_1, e_2) = \gcd(e_1, c) = 1$. And choose x, y, d, r_1, r_2 such that $x = (r_1 L)/\gcd(e_1^4(ee_2+e_1)r_1, L)$, $y = (r_2 L)/\gcd(e_2^4(ee_2+e_1)r_2, L)$, $d = L/\gcd(ee_2 - e_1, L)$, and keep them secret. Center selects a random number r_A for entity A, and calculates the secret keys g_{A1} and g_{A2} such that $g_{A1} = r_A^{-d} g^{xI_{A1}^2}$, $g_{A2} = r_A^{e^2 d} g^{yI_{A2}^2}$ where $I_{A1} = e_1 f(ID_A) + e = e_1 f_A + e, I_{A2} = e_2 f_A + 1$.

The common key K_{AB} between A and B can be obtained as follows : $K_{AB}^{(A)} = g_{A1}^{I_{B1}^2} g_{A2}^{I_{B2}^2} = g_{B1}^{I_{A1}^2} g_{B2}^{I_{A2}^2} = K_{AB}^{(B)}$.

III. Attacking Methods

Method 1 : Assume that $\gcd(e_1^4(ee_2+e_1), L) = \gcd(e_1(ee_2+e_1), L), \gcd(e_2^4(ee_2+e_1), L) = \gcd(e_2(ee_2+e_1), L)$. In RSA type modulus N, the previous equations hold with high probability. Then, the following relations hold : $xe_1 m = ye_2 m = 0 \pmod L \Longrightarrow K_{AB}^m = g^{m(xe^4+y)} \pmod N$ where m denotes $\gcd(ec_2 + e_1, L)$. Thus the *m-th* power of common keys between any entities have the same value. If m is small enough, then the common keys between entities will be the same value with high probability.

Method 2 : Now we consider collusion attack. First, an entity A builds the following equations for the common key K_{AB} between two entities A and B:

$$X_1 = g^{xe_1^4+ye_2^4}, \quad X_2 = g^{2xee_1^3+2ye_2^3}, \quad X_3 = g^{xe^2e_1^2+ye_2^2},$$
$$X_4 = g^{4xe^2e_1^2+4ye_2^2}, \quad X_5 = g^{2xe^3e_1+2ye_2}, \quad X_6 = g^{xe^4+y},$$

$$Y_{A1} = X_1^{f_A^2} X_2^{f_A} X_3, Y_{A2} = X_2^{f_A^2} X_4^{f_A} X_5, Y_{A3} = X_3^{f_A^2} X_5^{f_A} X_6,$$

$$\begin{aligned} K_{AB} &= (X_1^{f_A^2} X_2^{f_A} X_3)^{f_B^2} (X_2^{f_A^2} X_4^{f_A} X_5)^{f_B} (X_3^{f_A^2} X_5^{f_A} X_6) \\ &= Y_{A1}^{f_B^2} Y_{A2}^{f_B} Y_{A3} \pmod n. \end{aligned}$$

The attacking procedure consists of the following 3 steps :

step 1. As above, any entity A can obtain K_{AB}, K_{AC}, K_{AD}, K_{AE}, K_{AF}, K_{AG}, K_{AH} and K_{AI}. Using them, $Y_{A1}^2, Y_{A2}^2, Y_{A3}$ can be easily obtained.

step 2. By collaboration, we obtain X_1^4, X_2^4, X_3^2, X_4^4, X_5^2, and X_6.

step 3. Then a common key K_{UV} for any entities U and V can be expressed :

$$K_{UV} = X_1^{4c_1+i_1} X_2^{4c_2+i_2} X_3^{2c_3+i_3} X_4^{4c_4+i_4} X_5^{2c_5+i_5} X_6,$$

where $0 \le i_1, i_2, i_4 < 4$ and $0 \le i_3, i_5 < 2$. Let $a = f_U \bmod 4$ and $b = f_V \bmod 4$. Then in case $(a, b) = (0, 0), (0, 2), (2, 0)$ or $(2, 2)$, $i_1, \ldots, i_5$ are all zeros, and so K_{UV} can be calculated.

All the other cases are classified as follows:

(i) (0, 1), (0, 3), (1, 0), (3, 0) : $X_3 X_5$

(ii) (1, 1), (3, 3) : $X_1 X_2^2 X_4$

(iii) (1, 2), (2, 1), (2, 3), (3, 2) : $X_2^2 X_3 X_4^2 X_5$

(iv) (1, 3), (3, 1) : $X_1 X_4^3$

Hence Tanaka's scheme can be broken if 8(A, B, C, D, E, F, G, and H) entities collude, whose hashed values are : $f_A \bmod 4 = 0$, $f_B \bmod 4 = 1, f_C \bmod 4 = 2$, $f_D \bmod 4 = 3$, $f_E \bmod 4 = 0$, $f_F \bmod 4 = 1, f_G \bmod 4 = 2$, $f_H \bmod 4 = 3$, and $\gcd(R, S) = 2$ where $R = (f_A - f_B)(f_C - f_D)(f_A + f_B - f_C - f_D)$ and $S = (f_E - f_F)(f_G - f_H)(f_E + f_F - f_G - f_H)$.

IV. Conclusion

In this paper, we introduced two different ways for attacking Tanaka's IDNIKS. First, we have shown how to get the secret information from the public parameters of the center. Second, even if it is impossible to get the secret information of the public parameters, we have shown that Tanaka's scheme can be broken by 8 collaborators.

Acknowledgements

The authors would like to express great thanks to Prof. T. Matsumoto for his valuable comments on this paper.

References

[1] Blom, R., "Non-Public Key Distribution", *Advances in Crytology - CRYPTO'82*, pp.231 - 236, Plenum Press, 1983.

[2] Matsumoto, T. and Imai, H., "On the Key Predistribution System: A Practical Solutions to the key distribution problem", *Advances in Crytology - CRYPTO'87*, Lecture Notes in Computer Science, no. 293, pp.185 - 193, Springer-Verlag, 1988.

[3] Maurer, U.M. and Yacobi, Y., "Non-Interactive Public Key Cryptography", *Advances in Crytology - EUROCRYPT'91*, Lecture Notes in Computer Science, no. 547, pp. 498 - 507, Springer-Verlag, 1991.

[4] Tanaka, H., "Identity-Based Non-Interactive Key Sharing", *Proc. SCIS'93* (Shuzenji), SCIS93-17C, Jan. 30, 1993.

[5] Tanaka, H., "New Schemes of Noninteractive ID-Based Key Sharing Schemes", *Proc. SCIS'94* (Lake Biwa), SCIS94-3C, Jan. 27, 1994.

[6] Tsujii, S., Nisio, G. and Chao, J., "An ID-Based Cryptosystem Based on the Discrete Logarithm Problem", *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 4, 1989.

Simply Implemented Identity-Based Non-Interactive Key Sharing

Hatsukazu TANAKA
Department of Electrical & Electronics Engineering
Faculty of Engineering, Kobe University
Rokko, Nada, Kobe, Japan 657

Abstract — **A new simply implemented identity-based non-interactive key sharing scheme(IDNIKS) has been proposed. The center algorithm is very simple and easily implemented. The security depends on the difficulty of factoring. The proposed IDNIKS can be certified to be secure for the considerable attacks involving user's collusion.**

I. Introduction

In this paper a new identity-based non-interactive key sharing scheme(IDNIKS) has been proposed in order to realize the Shamir's original concept of identity-based cryptosystem[1]. The algorithm is very simple and easily implemented. The security depends on the difficulty of factoring and it can be certified to be secure for user's collusion.

II. Basic Center Algorithm

Let P and Q be two large primes and their product be $N = PQ$. Then the Carmichael function of N can be given by L=LCM{P-1,Q-1}. Let g be a primitive element in GF(P) and GF(Q), and let n and $w(\simeq n/2)$ be two positive integers which satisfy gcd{w, L}=1. We assume here that the identity information of each user l (l=A,B,C,...) is given by ID_l, and introduce a one-way function f which satisfies $0 \le I_l = f(ID_l) \le nCw$ - 1. Then using the Schalkwijk algorithm[2], we obtain a constant weight binary vector $\mathbf{v}_l = (a_{l,0}, a_{l,1}, \cdots, a_{l,n-1})$, $a_{l,i} \in$ GF(2) from I_l, where the Hamming weight of $\mathbf{v}_l$ is w. We assume here that any n-vectors of $\mathbf{v}_l$ are linearly independent. From the vector $\mathbf{v}_l$ an index set can be defined as $J_l = \{j \mid a_{l,j} = 1,\ 0 \le j \le n-1\}$. Here we introduce a set of random numbers $X = \{x_0, x_1, \ldots, x_{n-1}\}$, and calculate the following equations.

$$S_l = \sum_{j \in J_l} x_j \,(\mathrm{mod}L) \tag{1}$$

$$g_l(j) = g^{S_l x_j} \,(\mathrm{mod}N) \quad (0 \le j \le n-1) \tag{2}$$

Finally the trusted center publishes $\{N, n, w, f, ID_l(l$ = A,B, C,..)} and delivers $G_l = \{g_l(j);\ 0 \le j \le n-1\}$ to each user l though a secure channel or by an IC card.

III. Non-Interactive Key Sharing

We assume here that two users A and B want to share a common-key K_{AB} between them non-interactively. First A calculates J_B from ID_B using f and the Schalkwijk algorithm, and then performs the following simple calculation to share a common-key K_{AB} with B.

$$\begin{aligned} K_{AB}^{(A)} &= \prod_{j \in J_B} g_A(j) \quad (\mathrm{mod}N) \\ &= g^{S_A\left(\sum_{j \in J_B} x_j\right)} \quad (\mathrm{mod}N) \\ &= g^{S_A S_B} \quad (\mathrm{mod}N) \end{aligned} \tag{3}$$

Similarly B calculates

$$\begin{aligned} K_{AB}^{(B)} &= \prod_{j \in J_A} g_B(j) \quad (\mathrm{mod}N) \\ &= g^{S_B S_A} \quad (\mathrm{mod}N). \end{aligned} \tag{4}$$

Then their shared common-key is given by

$$K_{AB} = g^{S_A S_B} \quad (\mathrm{mod}N). \tag{5}$$

IV. Considerations on the Security

In order to certify the security of our proposed IDNIKS we must show that a common-key between any third parties can not be forged under the following assumptions even if the collusion among users would be allowed.

Assumptions

A1. Factoring of $N = PQ$ is too difficult to execute.

A2. Any set of less than or equal to n vectors is linearly independent.

The possible strategies to forge a common-key between any third parties X and Y are only the following two attacks.

Attack 1: to solve in $z_{ij} = g^{x_i x_j} \,(\mathrm{mod}\ N)$ $(0 \le i, j \le n-1)$ the simultaneous equations given by G_l or K_{lt} gathered by user's collusion, and to construct a desired common-key between X and Y using their index sets J_X and J_Y.

Attack 2: to gather many $g_l(j)$ or K_{lY} and forge a user X's secret $g_X(j)$ or a common-key K_{XY} by replacing the exponent part using a linear combination $\mathbf{v}_X = a\mathbf{v}_A + b\mathbf{v}_B + c\mathbf{v}_C + \cdots \,(\mathrm{mod}\ L)$, where $a, b, c, \ldots$ are integer coefficients. For example, K_{XY} seems to be forged by

$$K_{XY} = K_{AY}^{a} K_{BY}^{b} K_{CY}^{c} \cdots \quad (\mathrm{mod}N). \tag{6}$$

In the process of executing the attack 1, we are inevitably confronted with solving an equation

$$z_{ij}^{w} = C \,(\mathrm{mod}N) \quad or \quad z_{ij}^{w^2} = C \,(\mathrm{mod}N), \tag{7}$$

where C is some known factor. However, we can not calculate w-th root of C because the inverse element (mod L) of w is unknown under the assumption A1. Such a situation is the same as that of RSA public-key cryptosystem.

In order to obtain the integer coefficients $a, b, c, \ldots$ for execution of the attack 2, we must solve an equation $(a, b, c, ..)V = \mathbf{v}_X \,(\mathrm{mod}\ L)$, where V is an $n \times n$ matrix of which each row vector is $\mathbf{v}_l$. However, it is impossible to solve it in $(a, b, c, \ldots)$ because, from the assumption A2, $|V|$ is always equal to w or -w, of which inverse element (mod L) can not be obtained because of the assumption A1. Such a situation is the same as that of RSA public-key cryptosystem.

References

[1] A. Shamir, "Identity-based cryptosystems and signature schemes," *Proceedings of Crypto'84*, pp. 47–53, 1985.

[2] J.P.Schalkwijk, "An algorithm for source coding," *IEEE Trans. Inform. Theory*, vol. IT-18, No.3, pp. 395–399, 1972.

The Degeneration and Linear Structures of Multi-Valued Logical Functions

Deng-Guo Feng, Bin Liu, and Guo-Zhen Xiao

P. O. BOX 304, Xidian University,

Xi'an, 710071, P. R. China

It is important in cryptology to study the degeneration of multi-valued logical functions (MVLF). The main purpose of this article is, with the help of Chrestenson spectrum, to reveal the relationship between the degeneration of MVLF and its linear structures, and to characterize the property of these linear structures. The discussion here after is restrained to the prime field $GF^*(p)$.

Definition 1: Assume MVLF f: $GF^*(p) \to GF(p)$, *the Chrestenson transform* and *reverse transform* are

$$S_f(\omega) = p^{-n} \sum_{x \in GF^*(p)} u^{f(x)} \overline{u^{\langle \omega, x \rangle}}$$

and

$$f(x) = \log_u \Big(\sum_{\omega \in GF^*(p)} S_f(\omega) u^{\langle \omega, x \rangle} \Big)$$

respectively, where $u = \exp\left(\frac{2\pi \sqrt{-1}}{p} \right)$, $\langle \omega, x \rangle$ denotes the inner production of vetor ω and x, and $\overline{u^{\langle \omega, x \rangle}}$ the conjugate of $u^{\langle \omega, x \rangle}$.

In the following description, $\oplus$ means module p addition, and $+$ the ordinary addition. And f is a MVLF such that $GF^*(p) \to GF(p)$.

Definition 2: $f(x)$ is said to be *degenerate* if there exists a $k \times n (k < n)$ matrix D and MVLF $g(y)$ over $GF^k(p)$ such that $f(x) = g(Dx) = g(y)$, $\forall\, x \in GF^*(p)$, where $y = Dx$.

Definition 3: Let $W_i = \{x \in GF^*(p) \mid f(x) = i\}$, $0 \leqslant i \leqslant p-1$, and $|W_i|$ be the number of the elements in W_i. $f(x)$ is said to be *balanced* if $|W_0| = |W_1| = \cdots = |W_{p-1}|$.

Definition 4: $\alpha \in GF^*(p)$ is refered to as a *linear structure* of f if $f(x \oplus \alpha) - f(x) = constant (= f(\alpha) - f(0))$, $\forall\, x \in GF^*(p)$.

Let U_f be the set of all linear sturctures. An immediate conclusion from the definition is that U_f is a linear subspace of $GF^*(p)$. Let $U_i = \{\alpha \in GF^*(p) \mid f(x \oplus \alpha) - f(x) = i, \forall\, x \in GF^*(p)\}$, $0 \leqslant i \leqslant p-1$. The elements in U_i are refered to as the *ith* class of linear structure. Obviously, the difference between any two points in $U_i (1 \leqslant i \leqslant p-1)$ belongs to U_0. Hence, if $U_i \neq \phi$, then $U_i = \beta + U_0$. So U_f is the union of U_0 and some of its cosets.

Theorem 1: Let $V = \langle \{\omega \mid S_f(\omega) \neq 0\} \rangle$, dim $V = k$, and $H = [h_1, h_2, \cdots, h_k]^T$, Where $h_1, h_2, \cdots, h_k$ be group of bases of V. Then there exist functions with variables $g(y)$: $GF^k(p) \to GF(p)$, such that $g(y) = g(Hx) = f(x)$.

Theorem 2: $dim \langle \{\omega \mid S_f(\omega) \neq 0\} \rangle = k$ if and only if $f(x)$ degenerates into a function with at most k variables.

Theorem 3: $\alpha \in U_i$ if and only if $S_f(\omega) = 0, \forall\, \omega \in GF^*(p)$, $\langle \omega, \alpha \rangle \neq i$, where U_i is set of the *ith* class linear structure of $f(x)$.

Theorem 4: $U_0 = \{\alpha \mid H\alpha = 0\} = \langle \{\omega \mid S_f(\omega) \neq 0\} \rangle^{\perp}$.

By Theorem 2 and Theorem 4 it is known that the *0th* class linear structure virtually characterizes the degree of degeneration of function $f(x)$. The function is degenerated whenever $U_0 \neq \{0\}$. Meanwhile, it is pointed out that $U_0 = \langle \{\omega \mid S_f(\omega) \neq 0\} \rangle^{\perp}$.

Corollary: dim $U_f = n$ if and only if f is a linear function, i.e., $f(x) = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n + c_0$ where $c_i \in GF(p)$.

Theorem 5: If $f(x)$ has *ith* $(i \neq 0)$ class linear sturcture then (1) $f(x)$ must has other classes linear structures; (2) $f(x)$ is balanced.

By Theorem 5, if $U_f \neq U_0$, then all classes of linear sturcures exist. Once a 1*st* class linear structure α is found, $k\alpha (0 \leqslant k \leqslant p-1)$ is a *kth* class linear sturcture Thus, if U_0 and one $\alpha \in U_1$ is determined, all U_f can be determined.

References

[1] M. G. Karpovsky, "*Finite Orthogonal Series In the Design of Digital Devices*", John Wiley and Sons New York, 1976.

[2] K. Nyberg, "On the construction of Highly Nonlinear Permutations", *Advances in Cryptology*, Proceedings of Eurocrypt'92, Spring-Verlag, pp. 92-98, 1993.

[3] Wu Chuankun, "Spectral analysis of some independences of multiple-valued logical functions on their variables", Journal of Electronics, Vol. 10, No. 3 pp. 217-226, 1993.

A Fast Identification Scheme

P. Véron[1]

G.E.C.T., Université de Toulon et du Var, B.P. 132, 83957 La Garde Cedex, France

Abstract — **Many cryptographic protocols depend on one and only problem, the one of factoring. This paper presents a new identification scheme whose security depends on an NP-complete problem from the theory of error correcting codes : the syndrome decoding problem. The computation complexity of the proposed scheme is smaller than those of the other schemes based on SD problem. Moreover the amount of memory needed by the prover is very small.**

I. Introduction

We define, in this paper, a new identification scheme based on the syndrome decoding problem *[1]*. The decision problem of the SD problem (stated in terms of generator matrix) is the following :

Let $G(k, n)$ be a generator matrix of a random linear binary code. Let p be an integer and x be a random binary vector of length n. Does there exist a word e of length n and weight p such that $x+e$ belongs to the code generated by G. Thus the problem is to know if there exists a couple (m, e) such that $x = mG + e$ where e is a word of weight p.

We will use the following definitions :

. Let $\beta = \{\beta_1, \ldots, \beta_k\}$ be a basis of F_{2^k} and $\gamma = \sum_{i=1}^{k} \gamma_i \beta_i$ be an arbitrary element of F_{2^k}, the β-weight of γ, $\omega_\beta(\gamma)$, is defined as the Hamming weight of $(\gamma_1, \ldots, \gamma_k)$.

. Let γ be an arbitrary element of F_{2^k}. The β-product matrix of γ, $[\gamma]_\beta$, is defined as the Kronecker product $\beta' \otimes \gamma$ where, for $1 \leq i \leq k$, $(\beta' \otimes \gamma)_i$ is considered as a row of k elements over F_2, and β' denotes the transpose of the row vector β. Thus $[\gamma]_\beta$ is a $k \times k$ invertible binary matrix.

. Let γ be a fixed element of F_{2^k}. Let $\rho = \sum_{i=1}^{k} \rho_i \beta_i$ be an arbitrary element of F_{2^k} then : $\rho\gamma = (\rho_1, \ldots, \rho_k)[\gamma]_\beta$.

II. The Identification Scheme

Notations

. From now on, a binary vector of length k will be considered as an element of F_{2^k} if needed (and vice versa),

. Let y be a vector of length n and σ be a permutation over $\{1, \ldots, n\}$, then $y\sigma$ is defined as the vector z such that $z_j = y_{\sigma(j)}$. Likewise, if M is an $m \times n$ matrix then $M\sigma = (m_{i,\sigma(j)})$,

. $< x >$ denotes the action of a hash function over the string x,

. A vector x of length $2k$ will be represented by the couple (x_1, x_2) where x_1 and x_2 are vectors of length k. ◇

Let $\beta_1 = \{1, \alpha, \ldots, \alpha^{k-1}\}$ be a basis of F_{2^k}, β_2 be a basis of $F_{2^{2k}}$ and β_2^* be its dual trace basis. A certification center C, having the confidence of all users, computes two random elements γ_1 and γ_2 of F_{2^k}. Let S be equal to $\begin{pmatrix} [\gamma_1]_\beta \\ [\gamma_2]_\beta \end{pmatrix}$ and G' be a random $k \times 2k$ binary matrix of rank k. C computes $G = SG'$. This matrix is common to all users. S and G' are no longer needed and are unknown to all users. Finally, C computes for each user : a random binary vector $u = (u_1, u_2)$ of length $2k$, which verifies $uS = 0$, and the matrix $\hat{G} = [\rho]_{\beta_2^*}(G \mid Q)\pi_2$, where π_2 is a random permutation of $\{1, \ldots, 4k\}$, and Q is a random $(2k, 2k)$ matrix.

Secret quantities of each user are : ${\pi_2}^{-1}$, m a binary word of length $2k$, e a binary word of length $2k$ and weight p, u, mG and ρ^{-1}.

Public data of each user are : α^k, $\hat{G}$, $x = mG + e$ and p.

Suppose that A wants to prove its identity to B. The protocol includes r rounds, each of these being performed as follows :

- A computes a random element η of F_{2^k} and v a random vector of length $2k$. Let $w = u + v$, A computes $y = (\eta v_1, \eta v_2)$ and sends to B the quantity $y\rho^{-1}$,
- B sends back $z = (y\rho^{-1})\hat{G}$,
- A randomly computes : a permutation σ of $\{1, \ldots, 2k\}$ and $z\pi_2{}^{-1}$. The first $2k$ bits of this vector are equal to $(\eta w_1, \eta w_2)G$ (since $uS = 0$). Let $\tau = (\eta w_1, \eta w_2)$, A sends to B : $c_1 =< \sigma >$, $c_2 =< (\tau+m)G\sigma >$, $c_3 =< (\tau G + x)\sigma >$
- B sends a random element e of $\{0, 1, 2\}$,
- If e is 0, A discloses $\tau + m$ and σ. B checks the validity of c_1 and c_2,
- If e is 1, A discloses $(\tau + m)G\sigma$ and $e\sigma$. B checks the validity of c_2 and c_3 and verifies that $\omega(e\sigma) = p$,
- If e is 2, A discloses τ and σ. B checks the validity of c_1 and c_3.

The security of the scheme is linked to the values of the parameters k, p, $\omega_{\beta_2^*}(\rho)$ and r.

III. Security of the Scheme

According to *[2]*, minimal parameters which guarantee the security of the scheme are : $k = 255$, $p = 56$, $\omega_{\beta_2^*}(\rho) = 20$ and $r = 35$. The complexity of the various attacks is then, at least, 2^{70}, and the probability of success of the different frauds is about 10^{-6}.

It can be shown that repetition of the protocol is a "proof of knowledge" of a solution of the system $x = mG + e$, $\omega(e) = p$. Moreover we believe that this scheme is computationally zero-knowledge.

IV. Performances of the Scheme

The prover do not have to store the matrix G or the matrix $\hat{G}$ since he doesn't execute any computation with these matrices. Thus, the latter needs a very little amount of memory. Using the basis β_2^* allows the prover to compute $y\rho^{-1}$ without storing ρ^{-1}. Moreover the complexity of the computations done by the prover can be reduced by using an irreducible trinomial *[5]* so as to generate F_{2^k}. To show the efficiency of the scheme, we have compared it with Stern's scheme *[4]*, which among the schemes based on SD problem is the most practical. Here are results :

	Stern's scheme	Our scheme
Global transmission rate	$\simeq$ 40133 bits	$\simeq$ 75740 bits
ROM	66048 bits	2415 bits
Prover's workfactor	$\simeq 2^{22.13}$	$\leq 2^{19.9}$

References

[1] E.R. Berlekamp, R.J. Mc Eliece & H.C.A. Van Tilborg, *"On the inherent intractability of certain coding problems"*, IEEE Trans. Inform. Theory, 978, 84-86.

[2] A. Canteaut & F. Chabaud, *"Improvements of the attacks on cryptosystems based on error-correcting codes"*, private communication.

[3] U. Feige, A. Fiat & A. Shamir, *"Zero-knowledge proofs of identity"*, Proc. 19th ACM Symp. Theory of Computing, 10-17, 987.

[4] J. Stern, *"A new identification scheme based on syndrome decoding"*, Crypto'93, Lecture Notes in Computer Science 773, 3-1, Springer-Verlag, 994.

[5] N. Zierler, *"On the Theorem of Gleason and Marsh"*, Proc. Am. Math. Soc., 9 : 36-37, Math. Rev., 20 : 51, 958.

[1] veron@marie.polytechnique.fr

Transmission of Two-tone Images over Noisy Channels with Memory

Philippe Burlina[†], *Fady Alajaji*[‡] *and Rama Chellappa*[†]

[†] Center for Automation Research, University of Maryland, College Park, MD 20742, USA

[‡] Department of Mathematics and Satistics, Queen's University, Kingston, Ontario K7L 3N6, Canada

Extended Abstract

We consider an alternate approach to coding information bearing data for the reliable transmission of two-tone images over noisy communication channels with memory. This consists of jointly designing the source and channel codes (a technique referred to as joint source-channel coding).

Source and channel coding are two problems that have traditionally been implemented separately, forming what is known as a tandem source-channel coding system. The separation of channel and source coding is only optimal in an asymptotic sense, i.e., when no constraints exist on the coding block lengths (delay) and on the complexity of the encoder/decoder [1]. Joint source-channel coding, however, has recently received increased attention. It has been shown that if delay and complexity are constrained, performance can be increased if the source and channel codes are jointly designed, as opposed to being treated independently [2, 3].

In this work, we propose joint source-channel coding schemes for the reliable transmission of two-tone images over a binary channel with additive Markov noise. Applications of this work are in the transmission of facsimile documents over land mobile radio channels.

We model the image as a one-dimensional non-uniform binary iid, a Markov process or as a two-dimensional causal Markov process. We then investigate the problem of the maximum a posteriori probability (MAP) detection of binary images directly transmitted over the Markov channel. The objective is to design a MAP detector that fully exploits the redundancy of binary images to combat channel noise. It will also exploit the larger capacity of the channel with memory as opposed to the interleaved (memoryless) channel. Since this is a model-based decoding algorithm, we assume that the image parameters are provided to the decoder (this can be achieved by transmitting them over the channel using a forward error-control code). We next address the problem of MAP detection of compressed binary images directly transmitted over the Markov channel. Comparisons of the performance of the above coding schemes with traditional tandem schemes (that use Run-length and Huffman coding for source coding, and convolutional codes and interleaving for channel coding), are also presented.

Simulation results for the transmission and detection of an uncompressed two-tone image of Lena are displayed in Figures 1 and 2. In this experiment, the Markov channel bit error rate is $Pr(Z_n = 1) = \epsilon = 0.1$ and the noise correlation parameter is $\delta = 10.0$ (the corresponding noise correlation coefficient is $\frac{\delta}{1+\delta}$). These parameters correspond to a very noisy channel with high noise correlation. The resulting average decoding bit error probability is 0.02. This result is very promising given the low complexity of the system (which primarily resides in the MAP decoder). The decoder is implemented using a modified version of the Viterbi algorithm.

Figure 1. Received two-tone Lena

Figure 2. Decoded two-tone Lena

REFERENCES

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, July 1948.

[2] E. Ayanoglu and R. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Transactions on Information Theory*, vol. 33, pp. 885-865, November 1987.

[3] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of Binary Markov Sources Over Channels with Additive Markov Noise," submitted to *IEEE Transactions on Information Theory*, April 1994.

Multilevel Resolution of Digital Binary Images

Jack Koplowitz
ECE Department
Clarkson University
Potsdam, NY 13699-5720

Joseph DeLeone
Corning Community College
Corning, NY 14830-3297

Abstract - A hierarchical scheme for chain encoded digital contours is introduced. If the contour represents the boundary of a binary image, a true digitization of the image is realized as the pyramid structure goes from the fine to coarser resolution levels. A progressive transmission system is designed to go from a coarse resolutin level to finer levels which uses essentially the same number of bits as transmission of the contour at the finest resolution.

SUMMARY

We consider a pyramidal structure for digital binary images which lends itself to efficient multiresolution transmission. Binary images of objects are digitized by coloring a pixel black if the center point of the pixel cell is within the object. Otherwise, it is white.

Quadtrees are commonly used to create multiresolution structures. In this method the pixel cells which intersect the boundary of the object are designated as gray cells. It is only these cells that need to be subdivided to obtain a finer level of representation. Grey cells (or nodes) of the quadtree are designated as either gray-colored black or gray-colored white. Consequently, node designations are of four types, white, black, gray-white and gray-black. Thus, for a heirarchial representation using quadtrees, 8 bits are used to describe the four higher resolution children cells of each coarse resolution gray cell.

A second approach for the encoding of binary images is to use chain codes to follow the boundary of the digitized object. This consists of using 4-directional links that follow the edges or "cracks" of the pixel cells and hence is sometimes referred to as crack codes.

The crack code requires 2 bits per link and on average the number of links equals the number of gray cells for quadtrees. For contour following codes the number of links double with each level of increased resolution. Thus the "brute force" method of simply transmitting the full crack code at each level of resolution uses 4 bits per coarse link, half that for quadtrees.

A pyramidal structure for the crack code which makes use of the coarse information while transmitting information for the finer resolution could give further improvement. However, one finds that for digital binary images the content in the fine resolution is not sufficient to determine the coarse resolution image. This is due to the fact that as 4 small adjacent pixel cells coalesce into a coarser cell, knowledge of whether the center point of the coarser cell is within the object is not given by the colors of the smaller cells.

Conversely, this implies that when going from coarse to fine, a portion of the coarse information is not relevent when providing additional information for the fine resolution image. Thus for an efficient multiresolution system one should search for a structure whereby the course information is contained by the finer. This can be achieved by shifting (in each direction) the pixel cells, or equivalently the image, by 1/2 the value of the side of a cell. In this way the center point of a coarse cell coincides with that of a smaller cell.

We construct such a multiresolution structure for contour following chain codes. The total number of bits required for transmission if sent progressively is shown to be essentially the same as that for transmission only at the final level of resolution. Thus, in a sense, no coding inefficiency is created by the multiresolution structure and the scheme makes full use of the coarse data. Each level of resolution requires 2 bits per coarse link or coarse cell as opposed to 8 bits for quadtrees.

REFERENCES

[1] Samet, H., "The quadtree and related hierarchical data structures," *Comput. Surveys,* vol. 16, pp. 187-260, 1984.

[2] Renade, S., Rosenfeld, A., and Samet, H., "Shape approximation using quadrees,' *Pattern Recognition,* vol. 15, pp. 31-40, 1982.

[3] Rosenfeld, A., Samet, H., Shaffer, C., and Webber, R.E., "Application of hierarchical data structures to geographical information systems," RT-1197, Computer Science Dept., University of Maryland, College Park, MD, June 1982.

[4] Meer, P., Sher, C.A., and Rosenfeld, A., 'The chain pyramid: hierarchical contour processing," *PAMI,* vol. 12, pp. 363-376, 1990.

[5] Samet, H., "Data structures for quadtree approximation and compression,' *Commun. ACM,* vol. 28, pp. 973-993, 1985.

[6] Knowlton, K., "Progressive image transmission of grey-scale and binary pictures by simple, efficient, and lossless encoding schemes," *Proc. IEEE,* vol. 68, pp. 885-896, 1980.

A New Efficient Coding Method of a Still Image using Three-Dimensional DCT

Hiroshi KONDO, Suharno AGUS, and Syozo KOMORI
Dept. Elect. Eng., Kyushu Institute of Technology,
Kitakyushu 804, Japan

Abstract — **An efficient coding method using a three-dimensional discrete cosine transform (DCT) for still images is presented. This is an extended version of the traditional DCT coding method. The adaptive application of the three-dimensional DCT to each sub-block makes the coding more efficient than the other DCT methods.**

I. Introduction

In Near future an information-superhighway will be working all over the world. In such a situation an image coding method was standardized by the international body called Joint Photographic Expert Group (JPEG) [1]. And still now an efficient and fine image coding technique is required as urgently as ever. In this work using three-dimensional DCT we demonstrate that a more attractive image coding method for still images can be made.

II. Image Coding

Traditional DCT image coding is done by sectioning the full picture into tiny sub-blocks separately. The block size is usually taken an 8~32 pels. In such a tiny size there exist strong correlations between the block and its neighbor block. Hence it is considered that there exists redundancy in taking a transform coding for each sub-block independently. To remove such redundancy, we adopt a three-dimensional DCT for the difference between sub-blocks. First we take nine sub-blocks (3x3 blocks) as a unit as shown in Fig. 1. Each sub-block is square with 8 pels. The nine sub-blocks are ordered as in Fig. 1. After each sub-block is transformed by two dimensional DCT the DCT coefficients of the 0th sub-block are subtracted from those of other sub-blocks.

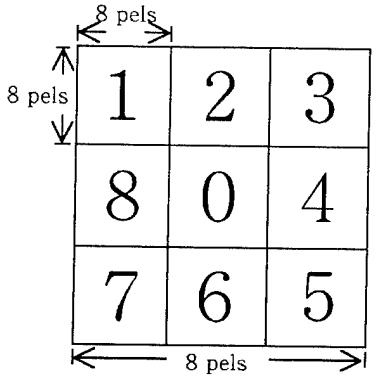


Fig. 1: Sub-block and its ordering

That is to say, with $D_l(i,j)$ the (i,j)th DCT coefficient of the lth sub-block we calculate differences $D_l(i,j) - D_0(i,j)(i,j = 1,2,...8)$ for any $l(l = 1,2,...8)$ and set these differences in the lth sub-block. Next using these sub-blocks we make a cubic structure as in Fig. 2. And the one-dimensional DCT is utilized again in the depth direction for the cubic structure.

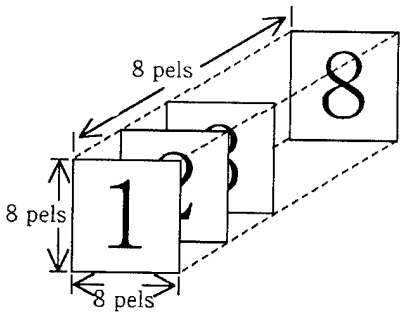


Fig. 2: Cubic structure

If the unit includes the clear edges of the image then the traditional coding method (JPEG method) is applied for each sub-block of the unit because the separate processing for each sub-block is preferable in such a case. Hence, the proposed method is a hybrid type of two-dimensional and three-dimensional DCTs. The total bit rate is determined from the rate distortion function and its quantization level from the Max's theory. For simplicity, however, using a constant reduction area and the equidistant quantization levels we can make an easier coding method which is still effective.

III. Simulation

The reconstructed image by using this method in 1 bit/pel have been of the SNR 29dB~33dB. The image qualities are all good and the block noise does not appear in these images.

IV. Conclusions

Using three-dimensional DCT we have constructed an efficient transform image coding method. It is considered that this method can play an important role for quick transmitting of still images.

Acknowledgements

This work was supported by Telecommunication Advancement Foundation (TAF).

References

[1] "Digital Compression and Coding of Continuous-Tone Still Images", ISO/IEC Draft International Standard 10918-1, January 10, 1992.

Nonlinear Filters In Joint Source Channel Coding Of Images

Ali S. Khayrallah

EE Department

University of Delaware

Newark, DE 19716, USA

Abstract — **We address the problem of data compression and transmission applied to images. We present a differential pulse coded modulation (DPCM) scheme whose prediction filter is the Lℓ filter. We also present an ECC decoder which takes advantage of side information to selectively filter the reconstructed prediction difference sequence, using the weighted median filter. We apply our schemes to some test images, and we compare their performance to the appropriate baseline schemes. Our techniques exhibit a lot of resilience to noise, even for very noisy channels.**

I. Introduction

DPCM is a well known technique for the compression of correlated sources, and has been widely used in the compression of speech, images, and video. It is simple and remarkably effective. DPCM is appropriate in applications where the hardware costs need to be kept low. Examples include many telephone systems and personal wireless communication systems. It is also appropriate in applications such as digital television, where the large data rate requires state of the art high speed electronics, which may prohibit the use of complex compression schemes.

Because of its differential nature, DPCM can suffer from acute sensitivity to bit errors. That is, a single bit error can affect many reconstructed samples. To mitigate this effect, in this paper we propose to use the Lℓ filter of Palmieri and Boncelet. This filter is a good predictor, and it also significantly reduces DPCM's sensitivity to bit errors.

We also present a modified ECC decoder which takes advantage of side information to selectively filter the reconstructed prediction difference sequence. This decoder uses the weighted median filter. This modified decoder further enhances the scheme's resilience to channel errors.

II. The Coding Scheme

We propose the Lℓ filter as a predictor in the DPCM feedback loop. Our filter is given by

$$\hat{w}_{i,j} = \alpha^{n}_{i-2,j}\,\hat{u}_{i-2,j} + \alpha^{n}_{i-1,j-1}\,\hat{u}_{i-1,j-1} + \alpha^{n}_{i-1,j}\,\hat{u}_{i-1,j} + \alpha^{n}_{i,j-2}\,\hat{u}_{i,j-2} + \alpha^{n}_{i,j-1}\,\hat{u}_{i,j-1}$$

where the coefficients α^n depend on the ranking of the elements $\hat{u}$. We consider a special case of Lℓ filter, where the only ranking information used consists of the location of the largest and smallest $\hat{u}$, which we discard by making their coefficients equal to zero. For the remaining three elements, we choose the filter that minimizes the MSE in the absence of a quantizer. One can see that our choice of Lℓ filter behaves like both a linear filter and a median filter. Its linearity makes it a good predictor. Its nonlinearity gives noise immunity to the DPCM decoder, because it provides a mechanism for discarding outliers.

We also propose a modified ECC decoder that takes advantage of two sources of additional information. The first is the residual redundancy in the prediction difference $\hat{v}$. Generally speaking, $\hat{v}$ is not highly correlated, but it retains enough local correlation to help the decoder. The decoder takes advantage of this by filtering the prediction difference estimate $\hat{v}'$. The filter will be applied selectively, only when the decoder has a low confidence in its output. Our choice of filter is the center weighted median filter.

The second source of additional information comes from the ECC decoder, which can produce an estimate for the error pattern $\mathbf{e}$ introduced by the noisy channel. If the weight of $\mathbf{e}$ is zero, the decoder is very confident in its decision. As the weight increases, it becomes less and less confident. We set a threshold $\tau \geq 0$, and let the decoder enable the filter each time the weight exceeds τ.

III. Examples

We apply our schemes to some test images. We compare the compression of DPCM with an Lℓ filter (Lℓ-DPCM) and its behavior in the presence of bit errors to that of two baseline schemes: DPCM with a linear filter (ℓ-DPCM) and DPCM with a median filter (M-DPCM). In terms of quantization, the mean squared error (MSE) performance of Lℓ-DPCM is better than M-DPCM and worse than ℓ-DPCM, but all three are good. Visually, the three schemes are essentially identical for quantizers with 3 bits and above.

In terms of channel response, without ECC, and with ECC and a standard decoder, Lℓ-DPCM sometimes beats ℓ-DPCM in MSE, and sometimes not, while M-DPCM is a distant third. Visually, Lℓ-DPCM does the best job of concealing distortion, since it suffers the least from the very noticeable streaks typical of linear DPCM. Using ECC with the modified decoder helps all three DPCM schemes in MSE, with ℓ-DPCM benefiting the most, and M-DPCM the least, because the median filter is not a particularly good predictor. Again, visually, Lℓ-DPCM wins.

The Polynomial Phase Difference Operator for Parametric Modeling of 2-D Nonhomogeneous Signals

Benjamin Friedlander
Dept. Elec. & Comp. Eng.
University of California
Davis, CA 95616

Joseph M. Francos
Dept. Elec. & Comp. Eng.
Ben-Gurion University
Beer-Sheva 84105, Israel

A fundamental problem in two-dimensional signal processing is the modeling and analysis of nonhomogeneous two-dimensional (2-D) signals. For example, in almost any image taken by a camera, perspective exists, and hence the acquired 2-D signal is nonhomogeneous, even if the original scene was homogeneous. Conventional approaches to the problems of perspective and camera orientation estimation usually involve *local* analysis of the image, by means of edge detection algorithms. Parametric models, when used in image processing, generally assume the observed image to be homogeneous, or piece-wise homogeneous. In this paper we consider a parametric model which is *nonhomogeneous*, and attempts to perform global (or at least, less localized) image analysis. We will study a model consisting of a sine (or cosine) of a polynomial function of the image coordinates.

For practical reasons it is more convenient to work with a complex valued model in which the sinusoidal function is replaced by a complex exponential. In applications where the 2-D signal is real, it can be converted subject to some restrictive conditions, into complex form through the Hilbert Transform. Throughout this paper we will consider 2-D signals which can be represented by a constant amplitude complex exponential whose phase is a polynomial function of the coordinates.

Let $\{v(n,m)\}$ be the 2-D field which is given by

$$\begin{aligned} v(n,m) &= A\exp\{j\phi_{S+1}(n,m)\} \\ \phi_{S+1}(n,m) &= \sum_{(k,\ell)\in I} c(k,\ell)n^k m^\ell \,, \end{aligned} \tag{1}$$

where $I = \{0 \le k,\ell \text{ and } 0 \le k+\ell \le S+1\}$. We shall call $\phi_S(n,m)$ a 2-D polynomial of *total-degree* S. In other words, one might think of the phase polynomial $\phi_S(n,m)$, as if it has S 'layers' since increasing S by one adds a 'layer' of additional $S+2$ parameters to the phase model.

Definition 1: Let τ_m and τ_n be some positive constants. Define

$$\begin{aligned} &\mathrm{PD}_{m^{(q)}}[v(n,m)] = \\ &\mathrm{PD}_{m^{(q-1)}}[v(n,m)]\left(\mathrm{PD}_{m^{(q-1)}}[v(n,m+\tau_m)]\right)^* , \\ &\quad n = 0,1,\ldots,N-1\,, m = 0,1,\ldots,M-1-q\tau_m \end{aligned} \tag{2}$$

$$\begin{aligned} &\mathrm{PD}_{n^{(p)}}[v(n,m)] = \\ &\mathrm{PD}_{n^{(p-1)}}[v(n,m)]\left(\mathrm{PD}_{n^{(p-1)}}[v(n+\tau_n,m)]\right)^* , \\ &\quad n = 0,1,\ldots,N-1-p\tau_n\,, m = 0,1,\ldots,M-1 \end{aligned} \tag{3}$$

where $\mathrm{PD}_{m^{(0)}}[v(n,m)] = \mathrm{PD}_{n^{(0)}}[v(n,m)] = v(n,m)$.

Let $\mathrm{PD}^S[v(n,m)]$ be the 2-D signal obtained by successively applying in some arbitrary sequence, P times the operator $\mathrm{PD}_{n^{(1)}}[\cdot]$, and $S-P$ times the operator $\mathrm{PD}_{m^{(1)}}[\cdot]$, to the signal (1). Then, $\mathrm{PD}^S[v(n,m)]$ is the 2-D exponential

$$\mathrm{PD}^S[v(n,m)] = \exp\left\{ j[\omega_S n + \nu_S m + \gamma_S(\tau_n,\tau_m)]\right\}, \tag{4}$$

whose spatial frequencies are given by $\omega_S = (-1)^S c(P+1,S-P)(P+1)!(S-P)!\tau_n^P\tau_m^{S-P}$, $\nu_S = (-1)^S c(P,S+1-P)P!(S+1-P)!\tau_n^P\tau_m^{S-P}$, and $\gamma_S(\tau_n,\tau_m)$ is neither a function of m nor of n.

We can thus reduce any 2-D nonhomogeneous, polynomial phase signal, $v(n,m)$, whose phase is of total degree $S+1$, to a 2-D single tone whose frequency is (ω_S,ν_S). Hence, estimating (ω_S,ν_S) using any standard frequency estimation technique, results in an estimate of $c(P+1,S-P)$, and $c(P,S+1-P)$. At present we estimate the frequency of the exponential using a search for the maximum of the absolute value of the signal 2-D Discrete Fourier Transform. We have thus obtained an estimate of two of the parameters of the highest order 'layer', $S+1$, of the phase model parameters(i.e., those $c(k,\ell)$'s for which $0 \le k,\ell : k+\ell = S+1$). However, the highest order 'layer', $S+1$, of the phase model parameters has $S+2$ parameters, which need to be estimated. This can be achieved by repeating the procedure which was described above assuming some arbitrary P, for *all* P such that $0 \le P \le S$.

Multiplying $v(n,m)$ by $\exp\{-j\sum_{k=0}^{S+1}\hat{c}(k,S+1-k)m^{S+1-k}n^k\}$ results in a new polynomial phase signal whose total degree is S. By applying to the resulting signal a procedure similar to the one used to estimate the parameters $c(k,\ell)$ for $k+\ell = S+1$, we obtain an estimate of the $S+1$ parameters in the S 'layer'. Let $v^{(s+1)}(n,m)$ denote the 2-D signal, where $s+1$ denotes the *current* total-degree of its phase polynomial. By repeating for all $s = S,\ldots,0$, the two basic steps of estimating the $c(k,\ell)$ parameters of 'layer' $s+1$ through finding the maxima of $\left|\mathrm{DFT}\left(\mathrm{PD}_{m^{(s-P)}}\left[\mathrm{PD}_{n^{(P)}}[v^{(s+1)}(n,m)]\right]\right)\right|$, for all $0 \le P \le s$, followed by multiplying the already reduced order 2-D polynomial phase signal by $\exp\{-j\sum_{k=0}^{s+1}\hat{c}(k,s+1-k)m^{s+1-k}n^k\}$ in the next step, we obtain estimates for all the phase parameters.

In many cases the observed 2-D signal is corrupted by additive white Gaussian noise. In this paper we derive the exact Cramer-Rao Lower Bound (CRLB) on the accuracy of estimating the model parameters in the presence of additive white Gaussian noise. The performance of the algorithm is illustrated by numerical examples, and its performance is compared with the Cramer-Rao bound.

Adaptive image restoration using discrete polynomial transforms

Xavier Neyt, Marc Acheroy
Elect. Eng. Dept., Royal Military Academy, Brussels, Belgium*

Abstract — **This paper presents a restoration algorithm based on a local signal description using discrete polynomials. The algorithm is made adaptive by estimating the local signal-to-noise ratio and by computing the corresponding deblurring filter.**

Furthermore, this method is developed for discrete signals, the input and output images being almost always available as discrete signals.

I. Introduction

Methods to describe, restore and compress signals by mean of polynomials have already been developed by Martens [1, 2] and Philips [4]. The basic idea behind these methods is the computation of filters in order to estimate the polynomial coefficients describing the ideal signal, starting from the degraded signal.

Martens [2], applying these methods to image restoration assumes that each sample of the sampled degraded image corresponds to the zero-order term of the ideal image polynomial expansion. This implies that the blurring kernel is identical to the squared local window function used to describe the signal.

In the proposed method, no other assumption is made about the blurring kernel than a general low-pass behaviour. This allows the choice of arbitrary-shaped blurring functions and of arbitrary positions for the localisation windows.

II. Discrete polynomial transforms

This transform consists in approximating the localised signal using polynomials. These polynomials are orthonormal with respect to a window function $V(i)$, i.e. they are defined by

$$\langle G_n, G_m \rangle = \sum_i V^2(i)\, G_n(i)\, G_m(i) = \delta_{n,m} \qquad (1)$$

and the coefficients of the polynomial expansion are obtained in the usual way.

When the localising function is a binomial, the orthogonal polynomials to be used are the Krawtchouk polynomials.

The extension to two dimensions is trivial when a separable localising function is considered.

III. Non adaptive restoration

The restoration algorithm consists in computing the coefficients of the polynomial expansion of the ideal signal from its degraded version using filters.

Note that because of the noise included in the blurred signal, it is not possible to estimate accurately the high order polynomial coefficients.

The filters to use are obtained by minimising the mean square error between the unknown coeffcients $L_{n,k}$ and their estimate $\hat{L}_{n,k}$.

These filters strongly depends on the local signal to noise ratio in the ideal image, which must be estimated.

Selecting a constant value for the signal to noise ratio yields permanent restoration filters, hence resulting in non adaptive restoration.

IV. Adaptive image restoration

To make the algorithm adaptive, the local signal variance must be estimated in each window and the corresponding filters computed.

Since the SNR of the estimated coefficients of low order is high, even if the SNR of the filters doesn't match that of the image, the coefficients computed using these filters can be used to get an estimate of the local signal variance hence yielding the signal to noise ratio.

Having the SNR of the ideal image in each window, the estimation filters can be computed. These filters applied to the blurred image will give estimates of the coefficients of the ideal image, thus yielding the restored image.

V. Conclusions

A local description is particularly well suited for adaptive restoration methods. Only adaptivity with respect to the SNR has been considered here but a spatially variant blurring filter B could easily be considered since no assumption has been made about the blurring filter. Moreover, due to the property of orientation selectivity of this kind of transform (when extended to 2D), a directional restoration could also be implemented.

Note finally that this local description enables the easy parallelisation of the algorithms.

References

[1] J.B. Martens, "The Hermite Transform — Theory and Applications," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, No. 9, pp 1595 – 1618, September 1990

[2] J.B. Martens, "Deblurring Digital Images by Mean of Polynomial Transforms," *Computer Vision, Graphics and Image Processing*, 50, pp 157 – 176, 1990

[3] W.M. Wells, "Efficient Synthesis of Gaussian Filters by Cascaded Uniform Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp 234 – 239, No 2, March 1986.

[4] W. Philips, "Time-Warped Polynomials for Signal Representation and Coding," *Signal Processing VI: Theories and Applications*, Elsevier 1992, pp 925–928

[5] X. Neyt. *Restauration adaptative d'images basée sur les représentations polynomiales*, thesis presented at the Faculty of Applied sciences at the Université Libre de Bruxelles, May 1994.

*The authors can be contacted via E-mail at the following addresses: Xavier.Neyt@elec.rma.ac.be and Marc.Acheroy@elec.rma.ac.be

Extreme elements and Granulometries in the estimation problem.

Ivan R. Terol Villalobos.
Centro de Inv. y Desarrollo Tec. en Electroquímica. S.C.
Sanfandila Pedro Escobedo. Qurerétaro México.

Abstract - **In this paper, we use the extreme elements to obtain size and shape information of random structures.**

I.- INTRODUCTION.

In textural analysis, Mathematical Morphology M.M. [1]use the probabilistic models and the Granulometry to obtain size and shape information. On the other hand, fractal objets are characterized by applying morphological dilations. Here, we proppose an approach working with the extreme elements of a Granulometry and of the Eroded.

II.-MORPHOLOGICAL TRANSFORMATIONS

In M.M.the basic morphological transformations, the Dilation and Erosion by B (structuring element), are given by,

$\delta_B(X) = X \oplus \check{B} = \bigcup \{X_b : b \in \check{B}\}$

$\varepsilon_B(X) = X \ominus \check{B} = \bigcap \{X_b : b \in \check{B}\}$; $\check{B} = \{-x : x \in B\}$

and the morphological closing and opening are defined by,

$\varphi_B(X) = \varepsilon_{\check{B}}\, \delta_B(X)$ $\gamma_B = \delta_{\check{B}}\, \varepsilon_B(X)$.

Definition 1.. A Granulometry is a family Ψt, for t>0, such that Ψt is antiextensive, increasing for all t and for all s,t>0,

$\Psi_t(\Psi_s(X)) = \Psi_s(\Psi_t(X)) = \Psi_{\sup(s,t)}(X)$

The opening $\gamma_{\lambda B}$, with B a compact convex set satisfies these axioms and two functions are associated; the probability distribution function and its derivate:

$$F(\lambda, X) = \frac{\mu(X) - \mu(\gamma_{\lambda B}(X))}{\mu(X)} \qquad g(\lambda, X) = \frac{d}{d\lambda} F(\lambda, X) \qquad (1)$$

where μ is the Lebesgue measure (area in this case).

For the λ parameter we associate the critical element λ=λn for a given set X. $\lambda n = \sup\{\lambda : \gamma_{\lambda B}(X) \neq \varnothing\}$. In the same way for the erosion case, $\lambda n = \sup\{\lambda : \varepsilon_{\lambda B}(X) \neq \varnothing\}$.

III.- GRANULOMETRY OF CRITICAL ELEMENTS.

Let be $r_\gamma(\lambda, X) = r_\gamma(\lambda) = X - \gamma_{\lambda B}(X)$ the residue operator of X after application of $\gamma_{\lambda B}$ and λn the critical element of X. Invariably we use $\gamma_{\lambda B} = \gamma_\lambda$ and $r_\gamma(\lambda_{n+1}) = X$ for $\lambda n+1 > \lambda n$. In a recursive way, we have $r_\gamma(\lambda_n) = r_\gamma(\lambda_{n+1}) - \gamma_{\lambda B}(r_\gamma(\lambda_{n+1}))$ and for a given k, $r_\gamma(\lambda_k, r_\gamma(\lambda_{k+1})) = r_\gamma(\lambda_{k+1}) - \gamma_{\lambda k}(r_\gamma(\lambda_{k+1}))$, where λk is the critical element of $r_\gamma(\lambda_{k+1})$. In other words,

$\lambda_i = \sup\{\lambda : \gamma_{\lambda B}(r_\gamma(\lambda_{i+1})) \neq \varnothing\}$

We associate two functions; the probability distribution function of critical elements and its derivate:

$$Fc(\lambda_k, X) = \frac{\mu(X) - \sum_{i=k}^{n} \mu(\gamma_{\lambda i}(r_\gamma(\lambda_{i+1})))}{\mu(X)} \qquad (2)$$

We define $Fc(\lambda, X) = Fc(\lambda_k, X), \forall \lambda \in [\lambda_k, \lambda_{k+1})$ and $gc(\lambda, X) = d(Fc(\lambda, X))/d\lambda$. Using a linear structuring element, we have g(λ,X)=gc(λ,X) and F(λ,X)=Fc(λ,X).

To test this approach, we realize a random geometrical characterization by using a deterministic approach. In [2] it is showed, that the deterministic Sierpinski Gasket objet S.G. has a similar behavior, in the percolation studies, than the random S.G. We use this physical assumption. In fact, the Fc and gc functions calculated on the complement of deterministic S.G. are similar than the random case. In this case we have

$Fcri(\lambda_k) = 1 - (\sum_{i=k}^{n} \mu(\gamma_{\lambda i}(r_\gamma(\lambda_{i+1})))/\mu(M)) = ((4-P)/4)^k$ (3)

where M is the mask or the frame and P is the probability filling to create a random S.G. From (3) we obtain a family of straight lines with slope log((4-P)/4),

$\log(Fcri(\lambda_k)) = k \log((4-P)/4)$

By calculating the slope we estimate the filling factor and the fractal dimension. We realized experiments to estimate the fractal dimension of the union of two S.G. objet with the same fractal dimension. We obtain the same fractal dimension.This approach is now used on other fractal objets.

IV.- DEAD LEAVES MODEL

An appropriate model for grains overlaps when the contour of the grains is apparent, is the Dead Leaves Model [3] . A Dead Leaves simulation X, is constructed by implanting independent realizations of primary grains X' at random poins of a Poisson point process (density θ) using a masking law.The probability for a connected set B to be included in a grain is given by,

$P(B, t) = P(B \subset X(t)) = \bar{\mu}(X' \ominus B)/\bar{\mu}(X' \oplus B)[1 - Q(B, t)]$

where $Q(B, t) = \exp(-\theta t \bar{\mu}(X' \oplus B))$

Let X'_i a random disk (radius R_i) with f_i unknow frecuency (dicrete case) for "n classes" and B(r) a ball of radius r. Then,

$H(r) = \pi \sum_{Ri>r} f_i (R_i - r)^2$

where

$$H(r) = \frac{LOG(Q(B(r)))\, P(B(r), t)}{\theta[1 - Q(B(r))]}$$

H(r) can be estimated from the images. Initially, we estimate the value f_n by calculating the size "r" (R_{n-1} -r=0)of the extreme element of the class n-1 (primary grain). Next, a similar procedure is used to estimate f_{n-1} by calculating the extreme element of the classe n-2. We realize the same operation until all the f_i are estimated. The number of classes (limits of application) is four or five.

REFERENCES

[1] Serra J. "Mathematical Morpholgy and Image Analysis". Academic Press 1988 Vol. II.

[2] Clerc et all "The electrical conductibity of binary desordered systems, percolation clusters, fractal and related models".Advances in Physics 1990,Vol.39,No3.

[3] Jeulin D., Terol I. "Application of the Dead Leaves Modelto Powders Morphological Analysis". Acta Stereol 11, Suppl. 1, 105-110.

A Noise Tolerant Algorithm for the Object Recognition of Warning System

Dae-Seong Kang
Satellite Broadcasting System Section
Electronics and Telecommunications Research Institute
P.O. Box 106, Yusong, Taejon, Korea

Abstract — **In this communication, we discuss a noise tolerant traffic sign recognition algorithm which can be utilized in future vehicles to warn drivers that they are approaching traffic signs at an intersection. It is assumed that the vehicle is equipped with a video camera providing chromatic images of the navigational environment. The primary objective of this paper is to develop a noise tolerant color segmentation algorithm and a recognition algorithm which is tolerant to the rotation, position, and scale variations. The treatment of traffic signs by computer vision techniques has been fairly limited in the literature.**

I. Introduction

In outdoor noisy environments, it is not easy to obtain invariant feature vectors from images which have ratation, position, or scale variations. The design of a pattern recognition system for distorted images has long been a challenging goal. In classical pattern recognition methods, the input patterns are required to be standard patterns, since they are very sensitive to rotations, positions, scale variations, or noise. Classical pattern recognitin systems, for example, matched filter, do not operate well under these distortions. Distortions such as rotations, positions, and scale changes of the pattern can be tolerated by using proper geometrical transformations. In a real outdoor environment, the brightness [1] in images constantly varies due to sun angle, weather, clouds, or other conditions. This means the value of brightness is very sensitive to the light source. In such a case, we need a pattern recognition algorithm which is relatively insensitive to brightness variation.

II. System Procedure

The object recognition algorithm consists of two phases: noise tolerant segmentation and object classification invariant to rotation, position, and scale variations. The results of color segmentation depend not only on its segmentation algorithm, but also on choosing the color coordinate system. In this study, the proposed $(u,v,hu\epsilon)$ coordinate system is relatively insensitive to brightness variation, which is useful to measure of color difference between any two arbitrary colors. The proposed segmentation algorithm uses a split and merge concept and an iteration method. Fig. 1 shows the procedure of the noise tolerant segmentation algorithm.

To obtain the above invariances, PLFT(Polar-Log-Fourier transform) is used, which is a powerful method to implement rotation and scale invaiant mapping for 2-dimensional object recognition. Before applying PLFT, we have to do position normalization by moving the object to the center of the image. The network for object classification is a back-propagation network with forty-nine input nodes, one hundred hidden layer nodes, and four output nodes.

III. Experimental Results

Several images were selected to demonstrate the robustness of this approach. Namely: do-not-enter, stop, yield, and other signs were processed under different scales, rotation, and shape variations. The input vector to the classification network was the 7 × 7 array and the output vector had four components corresponding to the traffic signs.

IV. Conclusions

A pattern recognition algorithm which is invariant to noise, brightness variation, rotation, position, and scale change using color classification technique, geometrical transformation, and an artificial neural network has been developed as part of a warning system for approaching traffic signs. The algorithm was tested on a large number of signs with different positions, rotations, scales, and backgrounds. The results of color segmentation phase were tolerant to noise and brightness variations.

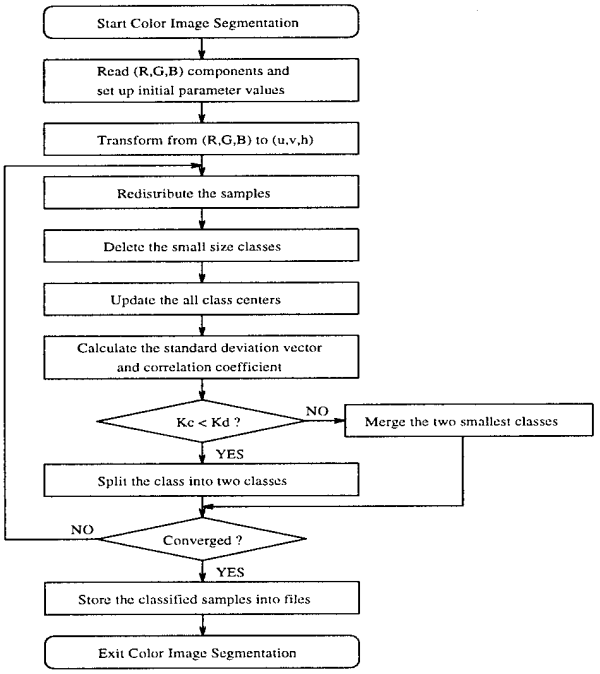


Figure 1: Procedure of the noise tolerant segmentation algorithm.

References

[1] G. Wyszecki and W. S. Stiles, *Color Science : Concepts and Methods, Quantitative Data and Formulas*. John Wiley & Sons, 1967.

Variable Step Search Algorithm for Motion Estimation

Z. Q. Cai and V. N. Tran
Dept. of Communication & Electronic Engineering
Royal Melbourne Institute of Technology, Melbourne 3001, Australia

Summary

Motion estimation(ME) is a key technique in interframe coding. It is the basis of most compression algorithms for video compression, such as the CCITT standard H. 261, MPEG 2 and so on[1]. The performance of ME is decided by two factors: (1) the estimation exactitude; (2) the computational load.

Full search algorithm is the optimal one in the first meaning, but it requires extensive computations. To reduce the computational complexity, many efficient search algorithms have been proposed[3-6]. As described in [6], one step at a time search(OSATS) is the second most efficient algorithm, but it becomes inefficient when the search window is greater than 4 pels/frame. The aim of this paper is to overcome this disadvantage.

The OSATS algorithm[4] looks for a minimum mean absolute error position(MMAEP) in the i-direction first, and from there proceed in the j-direction to find the final MMAEP in the searching window.

On basis of the OSATS algorithm, VSS makes use of variable steps during search, not like in OSATS where one step at a time search. Then the search efficient is greatly advanced. The algorithm is described as follows(Fig. 1).

Step 1: Compare the current block with the block(i, j) in the previous frame, if the value D(i, j) of the distortion function(in simulations, mean absolute error(MAE) is used) is less than a predefined threshold, then the current block is thought to be a nonmoving block and search stops. Otherwise, go to next step.

Step 2: Compute D(i, j), D(i, j-1) and D(i, j+1), a minimum is got. If minimum = D(i, j-1), the block moves left; If minimum = D(i, j+1), the block moves right; Otherwise it goes vertically. Set the search step size "p" equal to half of the search window "w", i.e. $p = w / 2$.

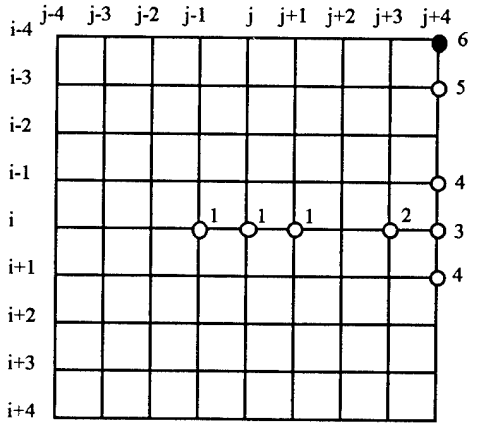
Step 3: Move the coordinate (i, j) to MMAEP(m, n), i.e. i = m and j = n. Find MMAEP(m, n) of the coordinates (i, j), (i, j+sp), where "s" is a sign function which is equal to "1" if the block moves right or "-1" if the block moves left.

Step 4: If p = 1, go to Step 5, otherwise halve the step size "p" and go to Step 3.

Step 5: Keep j-direction fixed after finding MMAEP in j-direction, proceed in i-direction as that of j-direction.

Therefore, for the maximum motion displacement of w pels/frame, the total number of computations is $5 + 2\log_2 w$. A simple example is given in Fig. 1, where w = 4.

From the description of the algorithm in the above section, it can be seen that the two algorithms use the same idea that is first to find MMAEP in i-direction and then keep j-direction fixed, find the position in j-direction with minimum MAE, which is also the final MMAEP in the searching window. Thus, the result of motion estimation in VSS is the same as that of the OSATS. However, the maximum number of search points(MNSP) in VSS is much less than that in OSATS, where MNSP is $3 + 2w$.



● : MMAEP in searching window.

Fig. 1 Variable Step Search

References

[1] R. Aravind, G. L. Cash, D. L. Duttweiler, H. M. Hang, B. G. Haskell, and A. Puri, " Image and Video Coding Standards," AT & T Technical Journal, January/February 1993, pp. 67-89.

[2] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. Commun., vol. COM-29, Dec. 1981, pp.1799-1808.

[3] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion Compensated Interframe Coding for Video Conferencing," in Proc. Nat. Telecommun. Conf., New Orleans, LA, Nov. 29-Dec. 3, 1981, pp. G5.3.1-5.3.5.

[4] R. Srinivasan and K. R. Rao, "Predictive Coding Based on Efficient Motion Estimation," IEEE Int. Conf. Commun., Amsterdam, May 14-17, 1984, pp. 521-526.

[5] A. Puri, H. M. Hang, and D. L. Schilling, "An Efficient Block-matching Algorithm for Motion Compensated Coding," Proc. IEEE ICASSP 1987, pp. 25.4.1-25.4.4.

[6] M. Ghanbari, " The Cross-Search Algorithm for Motion Estimation," IEEE Trans. on Commun., July 1990, Vol. 38, No. 7, pp. 950-953.

A Simple, General, and Mathematically Tractable Sense of Depth

A. Saadat and H. Fahimi
Elect. Eng. Dept., Sharif Univ. of Tech,
Tehran, Iran, e-mail: fahimih@irearn.bitnet

Abstract — **In the most general form the defocusing operator is a linear, circularly symmetric, lowpass, and space variant filter. In this paper without any restriction on the general model, the filter's band-width will be used as a measure of depth. The paper introduces a simple and efficient way to obtain this measure which has a well founded mathematical tractability. Experimental results indicates its high capability to resolve depth.**

I. Introduction

Depth From Defocus methods are based on the relationship between depth and the amount of defocus at each image point [1-3]. For a more effective use of this idea, the most general form of the defocusing operator should be used. The main idea in the proposed method is using the filter's bandwidth at each point as a value related to depth. For well-behaved low pass filters second derivative of the frequency response at origin, is a good measure of their effective band-width. This measure is used here as a sense for depth. In the next section theoretical foundations of the method is explained. An experimental result is given in section III. Section IV concludes this paper.

II. Theoretical Foundations

To extract the second derivative of the frequency response of the defocusing filter at each image point, the defocusing process is analysed in small regions of the image on which, the filter can be assumed space invariant. Consider the following functions in any small region (radius r_m) and in the polar cordinates (r, θ):

$i_o(r,\theta)$	:	focused image
$i_i(r,\theta)$	:	blurred or defocused image
$h(r,\theta) = h(r)$	:	defocusing operator,

Computing $i_o(r)$ and $i_i(r)$ by averaging the first two functions with respect to θ (from 0 to 2π) it can be shown [4] that the parameter d_i, which is defined as

$$d_i = \frac{\int_0^{r_m} i_i(r) r^3 dr}{\int_0^{r_m} i_i(r) r dr} \qquad (1)$$

can be used instead of the second derivative of the Fourier Transform of $i_i(r)$ at the origion. Constructing similar parameters d_o and d_h for i_o and i_h respectively, it can also be shown [4] that the parameters d_i, d_o, and d_h are related by

$$d_i = d_o + d_h \qquad (2)$$

and they can also be interpreted as powers of signals having normalized $ri_i(r)$, $ri_o(r)$, and $rh(r)$ as their density functions in $[0, r_m]$. This inteperetation and the additive form of (2) represents the mathematical tractability of d_i. In other words d_i can be used as a sense of depth in all regions of the image with the same d_o. For instance d_i can be used as a measure of d_h or depth in all regions of the image having the same texture.

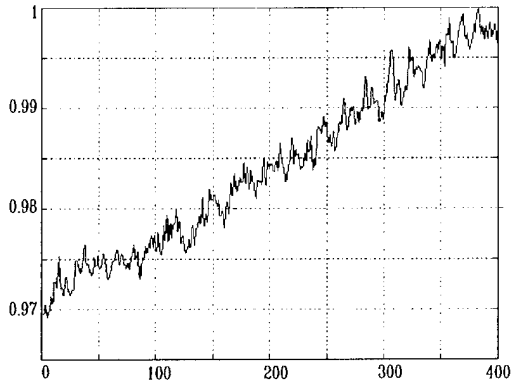


Fig.1. Normalized d_i for $r_m = 10$ pixels.

III. Experimental Results

As an example for evaluating the performance of the proposed sense, the edge texture is selected and used for computing d_i.The image of a black stripe of approximately 400 pixels long and 50 pixels wide on a white flat background, tilted against the camera, is used in a noisy environment. The simple experimental set up is described in [4]. In figure 1 the normalized d_i along one of the located edges in the image is plotted as a function of the length of the stripe. Due to the configuration of the set up, ideal curve should have a monotonically increasing form. Thus the ability of the proposed sense of depth, or d_i, can be seen from this figure.

IV. Conclusions

In this paper, based on the most general form, the parameter d_i is introduced for sensing depth. In the regions of the image with the same texture, this parameter is a good measure of the second derivative of the defocusing filter or depth. Using (1), there is no need for going to Fourier domain and differentiating which is sensetive to measurment noise. Experimental results on the edge texture indicates its ability to resolve depth.

References

[1] A. P. Pentland, "A New Sense for Depth of Field," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.9, no.4, pp.523-531, July 1987.

[2] S. H. Lai, C. W. Fue and, S. Chang, "A Generalixed Depth Estimation Algorithm With a Single Image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.14, no.4, pp.405-411, Apr. 1992.

[3] J. Ens and P. Lawerence, "An Investigation of Methods for Determining Depth From Focus," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.15, no.2, pp.97-108, Feb. 1993.

[4] A. Saadat and H. Fahimi, "A New Criterion for sensing Depth in Images ," *Proc. of 3rd Iranian Conf. on Elect. Engr. Appl. ICEE-74*, May 1995.

On the Cost of Finite Block Length in Quantizing Unbounded Memoryless Sources[1]

Tamás Linder[2] and Kenneth Zeger[3]

[2] Dept. of Telecommunications, Technical University of Budapest, Hungary.
email: linder@vma.bme.hu.

[3] Coordinated Science Lab., Dept. of Elect. and Comp. Engineering, University of Illinois, Urbana-Champaign, IL 61801
email: zeger@uiuc.edu.

Abstract — **The problem of fixed-rate block quantization of an unbounded real memoryless source is studied. It is proved that if the source has a finite sixth moment, then there exists a sequence of quantizers Q_n of increasing dimension n and fixed rate R such that the mean squared distortion $\Delta(Q_n)$ is bounded as $\Delta(Q_n) = D(R) + O(\sqrt{\log n/n})$, where $D(R)$ is the distortion-rate function of the source. Applications of this result include the evaluation of the distortion redundancy of fixed-rate universal quantizers, and the generalization to the non-Gaussian case of a result of Wyner on the transmission of a quantized Gaussian source over a memoryless channel.**

Shannon's source coding theorem with a fidelity criterion [1] showed that by increasing the blocklength n of a lossy source code, it is possible to have the mean squared error approach the distortion-rate lower bound arbitrarily closely. Pilc [3] showed that for finite alphabet sources the convergence of the mean squared error to the distortion-rate function occurs at a rate $O(\log n/n)$. It has recently been shown [2] that for bounded real memoryless sources and squared distortion this convergence occurs at a rate $O(\sqrt{\log n/n})$ This result was used in [2] to analyze the performance of a certain universal quantization scheme. On the other hand, the assumption of bounded support is sometimes a severe restriction in signal quantization, especially since some of the most popular source models have unbounded support, such as the Laplacian. The convergence rate results mentioned above also assume that binary information is transmitted across a lossless channel. In the present paper we eliminate the bounded support requirement and also consider transmission across a noisy channel. In addition we are able to obtain a rate of convergence result for universal lossy source coding.

Theorem 1 *Let $X_1, X_2, \ldots$ be a real i.i.d. source with $\mathbf{E}|X_1|^2 = M_2$ and $\mathbf{E}|X_1|^6 < \infty$. Let $0 < R_1 < R_2$ and assume that $D(R_2) > 0$. Then for any $R \in [R_1, R_2]$ there exists an n-dimensional quantizer Q_n with rate $r(Q_n) \le R$ such that*

$$\Delta(Q_n) \le D(R) + B\sqrt{\frac{\log n}{n}},$$

for all $n \ge 1$, where the constant B depends only on R_1, R_2, and the source distribution. Furthermore, the quantizers satisfy

$$\max_{x \in \mathbb{R}^n} \frac{1}{n}\|Q_n(x)\|^2 \le 2M_2.$$

In [4] Wyner proved that $D_n(R) - D(R) = O(\log n/n)$ for memoryless Gaussian sources, and in [5] showed that $D_n(R) - D(R) = O(\sqrt{\log n/n})$ for any correlated Gaussian source with a sufficiently well-behaved spectral density. Recently Zamir and Feder [6] showed that a $O(\log n/n)$ convergence rate is achievable for correlated Gaussian sources by means of a variable rate coding scheme using subtractive dither.

Corollary 1 *Suppose we are given a real memoryless source $X_1, X_2, \ldots$ with distortion-rate function $D(R)$, satisfying $\mathbf{E}|X|^6 < \infty$, and a discrete memoryless channel of capacity C, accepting one input per source output. Then there exists a source-channel coding scheme with delay n, such that denoting by $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n$ the channel decoder output, we have*

$$\frac{1}{n}\mathbf{E}\left(\sum_{i=1}^{n}|X_i - \hat{X}_i|^2\right) \le D(C) + O\left(\sqrt{\frac{\log n}{n}}\right).$$

Corollary 2 *For any $R > 0$, $k > 8$, and $\epsilon > 2/(k-4)$ there exists a sequence of universal quantizers $\{Q_n\}$ such that*

$$r(Q_n) - R = O\left(\left(\frac{\log n}{n}\right)^{(1/2)-\epsilon}\right),$$

and for any memoryless real source with $\mathbf{E}|X_1|^k < \infty$

$$\Delta(Q_n) - D(R) = O\left(\left(\frac{\log n}{n}\right)^{(1/2)-\epsilon}\right).$$

References

[1] R. G. Gallager. *Information Theory and Reliable Communication.* Wiley, New York, 1968.

[2] T. Linder, G. Lugosi, and K. Zeger. Rates of convergence in the source coding thorem, in empirical quantizer design, and in universal lossy source coding. *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1728-1740, November 1994.

[3] R. Pilc. The transmission distortion of a source as a function of the encoding block length. *Bell System Technical Journal*, 47:827–885, 1968.

[4] A. D. Wyner. Communication of analog data from a Gaussian source over a noisy channel. *Bell System Technical Journal*, pages 801–812, May-June 1968.

[5] A. D. Wyner. On the transmission of correlated Gaussian data over a noisy channel with finite encoding block length. *Information and Control*, 20:193–215, 1972.

[6] R. Zamir and M. Feder. Information rates of pre/post filtered dithered quantizers. submitted to *IEEE Trans. Inform. Theory*, 1993.

[1]The research was supported in part by the National Science Foundation under Grants No. NCR-92-96231 and INT-93-15271 and the Joint Services Electronics Program.

Universal Quantization of Parametric Sources has Redundancy $\frac{k}{2}\frac{\log n}{n}$

P.A. Chou*, M. Effros†, and R.M. Gray‡

*Xerox Palo Alto Research Center, 3333 Coyote Road, Palo Alto, CA 94304

†Dept. of Electrical Engineering, 116-81, Caltech, Pasadena, CA 91125

‡Information Systems Laboratory, Stanford University, Stanford, CA 94305-4055

Abstract — **Let $\{X_i\} \sim P_\theta$, $\theta \in \Lambda \subseteq \mathbb{R}^k$. 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 quantization: for any given Lagrange multiplier $\lambda > 0$, there exist universal fixed-rate and variable-rate quantizers with per-letter Lagrangian redundancy (i.e., distortion redundancy plus λ times the rate redundancy) as low as $\lambda\frac{k}{2}\frac{\log n}{n}$.**

Let $\{X_i\}$ be a stationary ergodic random process over alphabet $\mathcal{X}$ with process measure P_θ, $\theta \in \Lambda \subseteq \mathbb{R}^k$, and let $C^n = \beta^n \circ \alpha^n$ be a length-n quantizer with encoder $\alpha^n : \mathcal{X}^n \to \mathcal{S}$ and decoder $\beta^n : \mathcal{S} \to \mathcal{Y}^n$, where $\mathcal{S} = \{s_1, \ldots, s_M\} \subseteq \{0,1\}^*$ is some binary prefix code and $\mathcal{Y}$ is the reproduction alphabet. Let $d(x^n, y^n) = \sum_i d(x_i, y_i)$ be a single-letter fidelity criterion and let $|s|$ denote the length of the binary string s. The nth order operational distortion-rate function for $\{X_i\}$ is defined

$$\hat{D}_\theta^n(R) = \inf_{C^n} \left\{ \frac{1}{n} E_\theta d(X^n, C^n(X^n)) : \frac{1}{n} E_\theta |\alpha^n(X^n)| \le R \right\},$$

where the infimum is over either fixed-rate or variable-rate quantizers with blocklength n, as appropriate. The support functional of $\hat{D}_\theta^n(R)$ is defined

$$\hat{L}_\theta^n(\lambda) = \inf_{C^n} \left[\frac{1}{n} E_\theta d(X^n, C^n(X^n)) + \lambda \frac{1}{n} E_\theta |\alpha^n(X^n)| \right],$$

where $\lambda > 0$ is a Lagrange multiplier.

We show that there exists a universal sequence of fixed-rate or variable-rate quantizers $\{C^n\}$ such that the per-letter Lagrangian

$$\ell_\theta(\lambda, C^n) = \frac{1}{n} E_\theta d(X^n, C^n(X^n)) + \lambda \frac{1}{n} E_\theta |\alpha^n(X^n)|$$

converges to the support functional $\hat{L}_\theta^n(\lambda)$ as $\lambda\frac{k}{2}\frac{\log n}{n}$ for every $\theta \in \Lambda \subseteq \mathbb{R}^k$. To be precise, assume that for every θ, λ, and n, $\hat{L}_\theta^n(\lambda)$ is achieved by some C^n, say $C_{\theta,\lambda}^n$. Then define

$$\Delta_\lambda^n(\theta||\hat{\theta}) = \ell_\theta(\lambda, C_{\hat{\theta},\lambda}^n) - \hat{L}_\theta^n(\lambda)$$

to be the divergence between the Lagrangian performance of the quantizer matched to $\hat{\theta}$ and the quantizer matched to θ, with respect to θ. We have the following:

Theorem 1 *Let Λ be a subset of $\mathbb{R}^k$ (bounded if we are considering fixed-rate coding but possibly unbounded otherwise). Suppose that for each θ, λ, and n there exists a code $C_{\theta,\lambda}^n$ achieving the support functional $\hat{L}_\theta^n(\lambda)$. Suppose also that the corresponding divergence $\Delta_\lambda^n(\theta||\hat{\theta})$ is locally quadratic such that for each θ and λ there exists a neighborhood $S_{\theta,\lambda}$ of θ and a constant $m_{\theta,\lambda}$ such that $\Delta_\lambda^n(\theta||\hat{\theta}) \le m_{\theta,\lambda}||\theta - \hat{\theta}||^2$ for all $\hat{\theta} \in S_{\theta,\lambda}$ and for all n. Then for each λ there exists a weakly minimax universal sequence of codes $\{C^n\}$ such that for all θ*

$$\ell_\theta(\lambda, C^n) - \hat{L}_\theta^n(\lambda) \le \lambda \frac{k}{2} \frac{\log n + c_{\theta,\lambda}}{n}.$$

If Λ is bounded, and $S_{\theta,\lambda}$ and $m_{\theta,\lambda}$ do not depend on θ, then neither does $c_{\theta,\lambda}$, and the sequence $\{C^n\}$ is strongly minimax universal.

Proof: Fix λ. Construct $C^n = \beta^n \circ \alpha^n$ as follows. For each $n \ge 1$, partition $\mathbb{R}^k$ into a grid of hypercubes $\{A_i^n : i = 1, 2, \ldots\}$ each with side $1/\lceil n^{1/2} \rceil$, such that $\{A_i^n : i = 1, 2, \ldots\}$ refines $\{A_j^1 : j = 1, 2, \ldots\}$. For each hypercube A_i^n that intersects Λ, choose a representative $\hat{\theta}_i^n \in A_i^n \cap \Lambda$ and its matching quantizer $C_i^n = C_{\hat{\theta}_i^n,\lambda}^n$. Then define the encoder α^n to map x^n to the string $s = s_i' s_i'' s_i'''$ where s_i' represents the unit hypercube A_j^1 containing A_i^n, (which can be a fixed-length string if Λ is bounded), s_i'' represents the hypercube A_i^n indexed within A_j^1 (which is a fixed-length string with length $\log \lceil n^{1/2} \rceil^k$), and s_i''' is the string $\alpha_i^n(x^n)$ representing x^n using the quantizer C_i^n. The decoder maps s to the reproduction $y^n = \beta_i^n(s_i''')$. The index i is chosen to minimize the instantaneous Lagrangian $d(x^n, C_i^n(x^n)) + \lambda |s_i' s_i'' s_i'''|$. Thus

$$\begin{aligned} d(x^n, C^n(x^n)) + \lambda |\alpha^n(x^n)| &= \min_i d(x^n, C_i^n(x^n)) + \lambda |s_i' s_i'' \alpha_i^n(x^n)| \\ &\le d(x^n, C_j^n(x^n)) + \lambda |s_j' s_j'' \alpha_j^n(x^n)| \end{aligned}$$

for any particular j. Let j be the index of the cell A_j^n containing θ. Then dividing by n, taking expectations, and subtracting $\hat{L}_\theta^n(\lambda)$, we have

$$\begin{aligned} \ell_\theta(\lambda, C^n) - \hat{L}_\theta^n(\lambda) &\le \ell_\theta(\lambda, C_j^n) - \hat{L}_\theta^n(\lambda) + \frac{\lambda}{n} |s_j' s_j''| \\ &\le \Delta_\lambda^n(\theta||\hat{\theta}_j^n) + \frac{\lambda}{n} \left(b_\theta + \frac{k}{2} \log n \right), \end{aligned}$$

for some constant b_θ. By assumption, $\Delta_\lambda^n(\theta||\hat{\theta}) \le m_{\theta,\lambda}||\theta - \hat{\theta}||^2$ for all $\hat{\theta}$ in a neighborhood $S_{\theta,\lambda}$ of θ. Since $\hat{\theta}_j^n \to \theta$ with $||\theta - \hat{\theta}_j^n||^2 \le k/n$, there exists a constant $a_{\theta,\lambda}$ such that $\Delta_\lambda^n(\theta||\hat{\theta}_j^n) \le a_{\theta,\lambda} k/n$ for all n. Thus the theorem is proved with $c_{\theta,\lambda} = 2a_{\theta,\lambda}/\lambda + 2b_\theta/k$. □

A simple example of a source satisfying the conditions of the theorem is the following. Let $Z_1, Z_2, \ldots$ be an arbitrary real-valued stationary ergodic process with mean 0 and variance 1, and let $X_i = \sigma Z_i + \mu$. Then with $\theta = (\mu, \sigma) \in \Lambda \subseteq \mathbb{R}^2$, under the squared-error distortion measure and fixed-rate quantization of $\{X_i\}$, for all λ, n, θ, and $\hat{\theta}$, $\Delta_\lambda^n(\theta||\hat{\theta}) \le ||\theta - \hat{\theta}||^2$. Hence for any stationary source with unknown mean and variance in a bounded set, there exists a strongly minimax universal sequence of fixed-rate quantizers for which the nth order Lagrangian redundancy is at most $\lambda(k/2)(\log n + c)/n$, where $k = 2$.

On the Encoding Complexity of Scalar Quantizers

Dennis Hui and David L. Neuhoff
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109

Abstract -- It is shown that as rate increases the problem of asymptotically optimal scalar quantization has polynomial-time (or space) encoding complexity if the distribution function corresponding to the one-third power of the source density is polynomial-time (or space) computable in the Turing sense.

I. INTRODUCTION

Shannon's distortion-rate theory describes the optimal tradeoff between rate and distortion of vector quantizers. While it does not address the question of complexity, generally speaking, it is evident that quantizers need to become very complex in order to approach the optimal performance tradeoff, namely, the distortion-rate function. It is well known that the full-search unstructured quantizers with dimension k and rate R has storage and arithmetic complexity increasing exponentially with the dimension rate product kR. However, there are many reduced complexity full-search methods, and the question of how fast complexity must increase as performance approaches the rate-distortion function is open. Moreover, there are many structured vector quantization techniques whose complexities are substantially less than that of full search, but whose performance does not approach the distortion-rate function. It is unclear whether there exist structured quantizers with significantly reduced complexity and distortion close to the optimal.

The approach taken in this paper is to consider how the complexity of (asymptotically) optimal quantization with a given dimension k increases with rate R. Specifically, as an initial effort, we focus on the encoding complexity of scalar quantization.

II. PROBLEM FORMULATION

In stating and deriving the main result we adopt a Turing-like framework for evaluating complexity. Instead of assuming a different encoding machine for each R, whose relative complexities would be difficult, if not impossible to assess, we envision one machine, namely an *oracle Turing machine* M, c.f. [1,2], that is capable of encoding at any integer rate. That is, when rate R is specified, its output in response to a source sample x is an index I, $1 \le I \le 2^R$. We let $d(M,p,R)$ denote the mean-squared error (MSE) that results when this *Turing encoder* is used with an optimum decoder.

In the context of encoding, an oracle Turing machine consists of a finite-state machine, an unlimited tape memory and an oracle that provides a dyadic approximation to the source sample x to the required precision. The time (space) complexity of encoding at rate R with this machine, denoted $c(M,R)$, is the maximum number of steps, (alternatively, the maximum amount of tape memory) required to encode an arbitrary input sample.

We say a source density p is *asymptotically optimally quantizable in polynomial time (or space)*, abbreviated *PTIME-AOQ* (or *PSPACE-AOQ*), if there exists a Turing encoder M and a polynomial g such that

$$c(M,R) \le g(R) \quad \forall R \in Z^+, \quad \text{and} \quad \frac{d(M,p,R)}{D^*(p,R)} \to 1 \ \text{as} \ R \to \infty.$$

where $D^*(p,R)$ is the mean-squared error of the optimum quantizer of rate R.

Intuitively, it is easy to see that some source densities are intrinsically easier to quantize than others. For instance, sources with uniform density can be optimally quantized by simple uniform quantizers. On the other hand, it is also known that the optimal quantization point density for a given source is directly related to the one-third power of the source density. Therefore, it seems reasonable that the possibility of optimal quantization with polynomial complexity should depend on the "complexity" of the desired point density. In order to rigorously analyze this relationship, we adopt the framework of Turing complexity for real-valued functions, c.f. [2]. In this theory, a real-valued function $f:\Re \to \Re$ is said to be *polynmial-time (space) computable* if there is an oracle Turing machine M that is capabable of providing, for any x, a dyadic approximation to $f(x)$ to within an error of 2^{-n} for any pre-specified integer n, and its time (space) complexity is bounded from above by a polynomial function of n.

We are now ready to present the main results of this paper.

III. RESULTS

Proposition 1: Suppose $\lambda(x)$ is a desired quantization point density such that $\int p(x)/\lambda(x)^2 dx < \infty$ and the function $F(x) = \int_{-\infty}^{x} \lambda(y)dy$ is polynomial-time (alternatively, space) computable. Then there exists a Turing encoder that runs in polynomial-time (space) and with the resulting MSE satisfying

$$\limsup_{R\to\infty} \frac{d(M,p,R)}{\overline{D}(\lambda,p,R)} \le 1$$

where $\overline{D}(\lambda,p,R) = \left(2^{-2R}/12\right)\int p(y)/\lambda(y)^2 dy$ is the Bennett integral prediction for the MSE of a quantizer with a given point density.

Corollary 2: If the source density p is such that the function $F(x) = \int_{-\infty}^{x} c\,p(y)^{1/3} dy$, where $c = \left(\|p\|_{1/3}\right)^{-1/3}$, is polynomial-time (space) computable, then p is PTIME-AOQ (PSPACE-AOQ).

By applying Corollary 2, one can easily show that Gaussian, Laplacian and uniform source densities with zero means and unit variances are PTIME-AOQ and PSPACE-AOQ. On the other hand, it is also possible to construct a source density p for which the function F in Corollary 2 is not computable in polynomial time.

REFERENCES

[1] M.R. Garey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, 1979.

[2] Ker-I Ko, *Complexity Theory of Real Functions*, Birkhauser, 1991.

Reduced-Complexity Waveform Coding via Alphabet Partitioning

Amir Said[1]
amir@densis.fee.unicamp.br, Faculty of Electrical Engineering
State University of Campinas (UNICAMP), Campinas, SP 13081, Brazil

William A. Pearlman
pearlman@ecse.rpi.edu, Dept. of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180, U.S.A.

I. Introduction

In this paper we study the waveform coding problem where the data source symbols have a distribution that is simultaneously highly peaked and very long tailed—a situation when the source entropy is small, but the coding process must deal with a very large number of symbols. This type of problem can be found, for example, in the lossless compression of medical images. Those images are digitized with 10–12 bpp, and they are commonly quite smooth. With the adequate reversible transformation (e.g., linear prediction) we have a large fraction of pixel values near zero, but a significant number of pixels have very large magnitudes.

There are many practical difficulties when the data alphabet is large, which get much worse if we try to exploit the statistical dependence left between the source samples by, for example, coding several symbols together or designing conditional codes. Several *ad hoc* methods have been devised to deal with the problems caused by large alphabets. For instance, a popular method uses the "overflow" symbol to indicate which symbols are too large and should be coded separately.

II. The Alphabet Partitioning Method

We study a method to reduce the coding complexity when the source alphabet is large, based on the following coding strategy:

- the source alphabet is partitioned in a relatively small number of sets, with the number of elements in a set equal to a power of two.
- each symbol is coded in two steps: first the number of the set in which the symbol belongs (called *set number*) is coded; afterwards the number of that particular source symbol inside that set (the *set index*) is coded;
- when coding the pair (set number, set index) the set number is entropy-coded with a powerful and complex method, while the set index is left uncoded, i.e., its binary representation is stored or transmitted.

The advantage of this scheme is that it is normally possible to find partitions that simultaneously allow large reductions in the coding complexity and with a very small loss in the compression ratios. This partitioning technique is quite similar to the definition of "buckets" used in [1] for complexity reduction. The set numbers correspond to the bucket number, and can also be used to simplify context-based coding. However, here they have an additional purposes: they allow part of the information to left uncoded, with obvious advantages in speed and complexity. Furthermore, we show the advantages with methods that entropy-code several symbols together (e.g., Huffman, Lempel-Ziv).

To evaluate the loss incurred by leaving the set index uncoded, we assume a source with M symbols, each with probability $p_i, i = 1, \ldots, M$. The source entropy is denoted by $\mathcal{H}$. Partitioning the source symbols in nonempty sets $\mathcal{S}_n$, $n = 1, 2, \ldots, N$, we denote the number of elements in $\mathcal{S}_n$ by $M_n = 2^{K_n}$. We show that the expression for the maximum loss due to leaving the set index uncoded is

$$\Delta\mathcal{H} = \sum_{n=1}^{N} \sum_{i \in \mathcal{S}_n} p_i \log\left(\frac{M_n p_i}{P_n}\right), \tag{1}$$

where

$$P_n = \sum_{i \in \mathcal{S}_n} p_i \tag{2}$$

is the probability that the symbol belongs to the set with number n.

Equation (1) shows that, for each set, the loss should be small under two circumstances:

1. $M_n p_i \approx P_n$, that is, the distribution inside the set is approximately uniform;
2. the contribution of the set n to the entropy is very small.

Using the approach summarized above we consider the alternatives to find the best partitions for a given source, and analyze the trade-off between the coding/decoding complexity and the compression efficiency. Numerical results make clear the advantages of the alphabet partitioning method when used for the lossless compression of medical images. They show that there are simple and efficient methods to define the partitions, and those are quite versatile, i.e., they can be efficiently used for several images of the same type.

[1]This work was supported by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

References

[1] S. Todd, G.G. Langdon, Jr., and J. Rissanen, "Parameter reduction and context selection for compression of gray-scale images," *IBM J. Res. Develop.*, vol. 29, pp. 188–193, March 1985.

Some Results on Quantization of a Narrowband Process

Anurag Bist
Rockwell International, 4311 Jamboree Rd., Newport Beach CA 92658
e-mail: anurag.bist@nb.rockwell.com

Abstract — **We present some results on quantization of a narrowband Gauss-Markov process. The narrowband process is modeled as a lowpass complex envelope with a state space description. We compare the performance of narrowband process quantization schemes with several other previously analyzed schemes.**

I. Introduction

We present some results on quantization of a narrowband Gauss-Markov process. The narrowband process is modeled in a state space framework and several schemes for tracking the inphase and the quadrature components of the narrowband process are considered. These inphase and quadrature components are baseband processes, and thus are much more slowly varying than the original narrowband process. We compare the performance of these schemes with several previously analyzed schemes [1, 2] both with respect to a time averaged smoothed error, and their robustness with respect to the changes in the input spectrum. Finally, we present an analysis of the case when this narrowband process is input to a sigma-delta modulator. By performing an approximate analysis, we arrive at results which are applicable for a large class of inputs and are consistent with other more rigorous analyses [3, 4].

II. Modeling of Narrowband Process

If $x(t)$ is the original narrowband process, the complex envelope is given by : $\tilde{x}(t) = x_c(t) + jx_s(t)$, where $x_c(t)$ and $x_s(t)$ are respectively the inphase and the quadrature components of the narrowband process. Once we obtain the inphase and the quadrature components, we quantize them independently, or together by considering them as a complex state. These quantized values are then used to find an estimate of the original narrowband process. We consider three envelope quantization schemes: i) scalar quantization of the inphase and the quadrature components (complex components), ii) differential quantization of the complex components, and iii) quantization of the complex state $\tilde{\mathbf{S}}(t) = [\, x_c(t) \quad x_s(t) \,]^T$. In [1, 2] we analyzed several source coding schemes for a continuous time Gauss-Markov process. By fixing the overall transmission rate we compared the smoothed error performance of these source coding systems. By performing an identical analysis for the envelope quantization schemes, we can evaluate the optimal tradeoff between sampling interval and the quantization levels for the envelope quantization schemes.

III. Smoothed Error Performance

We compare the performance of the envelope quantization schemes with differential state quantization for a second order narrowband process. The hierarchy of performance within the different schemes is (from best to worst) : i) differential state quantization, ii) differential quantization of the complex state, iii) differential quantization of the inphase and the quadrature components of the complex envelope, iv) scalar quantization of the inphase and the quadrature components, and v) state vector quantization. Thus the envelope quantization schemes perform better than state vector quantization but worse than differential state quantization. Differential state quantization considers the quantization of a state consisting of the proces and its derivatives and was shown to be a superior quantization scheme for Gauss-Markov processes [2].

IV. Comparison of Robustness

We define a measure of robustness for different source coding systems, when the input spectrum changes. We encounter many situations where the input process is changing at regular intervals. In such situations the performance of the system which is designed for one particular input process deteriorates as the input spectrum changes from its original value. We model this change in terms of the state space matrices A and B and quantify the deterioration which accrues due to the changes in the spectrum. We show that for a second order narrowband process, for an N level two dimensional vector quantizer, the normalized change in the smoothed error due to the changes in the input spectrum for both the schemes (differential state quantization [2] and differential quantization of the complex state) is approximately proportional to $\frac{\ln N}{N-1}$. The changes in the smoothed error are marginally less in the differential quantization of the complex state.

V. Results on Sigma-Delta Modulation

Finally, we present a simple analysis of the case when the narrowband process is input to a sigma-delta modulator. We develop an approximate theory to analyze the quantization noise spectra of a sigma-delta modulator when the input is this narrowband Gauss-Markov process of any arbitrary order. The quantization noise spectra can always be approximated by a simple closed form expression in terms of state space matrices.

Acknowledgements

The author wishes to thank his advisor, Prof. N. T. Gaarder, without whose help and guidance this work would not have been possible.

References

[1] A. Bist, "Asymptotic Quantization Analysis of High Order Gauss-Markov Processes," Proceedings of International Symposium on Information Theory, Trondheim, Norway, June 1994 .

[2] A. Bist, "Analysis and Applications of Some Practical Source Coding Systems," Ph. D. dissertation, University of Hawaii, 1994 .

[3] R. M. Gray, "Quantization Noise Spectra," *IEEE Transactions on Information Theory*, Vol. 36, No. 6, pp. 1220-1244, Nov. 1990 .

[4] R. M. Gray, W. Chou, and P. W. Wong, "Quantization noise in single-loop sigma-delta modulation with sinusoidal inputs," *IEEE Transactions on Information Theory*, Vol. COMM-35, pp. 956-968, Sept. 1989 .

A Statistical Analysis of Adaptive Quantization Based on Causal Past

Bin Yu[1]

Statistics Department, University of California, Berkeley, CA 94720-3860, USA.
Email: binyu@stat.berkeley.edu

Abstract — **In this paper, a statistical estimation framework is proposed for adaptive quantization based on causal past. Different estimation methods are given for the marginal density based on the quantized sample. For a stationary and ergodic source process, if its marginal density is in a parametric family with a dimension less than the quantization level, then "adaptation" can be achieved when the sample size is large, i.e., the marginal density can be estimated consistently.**

SUMMARY

Adaptive lossless encoding/decoding based on the causal past is equivalent to the predictive version Rissanen's MDL (1989) or the prequential approach to statistical inference of Dawid (1984). In other words, at time $N+1$, a lossless code can be designed based on the causal past data $x^N = (x_1, x_2, ..., x_N)$ to encode the next data point x_{N+1}. Since the causal past data is available to both the encoding and decoding ends and as long as both ends agree to the same lossless coding rule depending on the causal past data, the encoding and decoding can be done "on the fly". In statistical terms, a lossless code based on the causal past amounts to a predictive density for x_{N+1} based on x^N. Such a predictive density can be obtained either parametrically or non-parametrically. The parametric predictive density can simply be the plug-in density estimator $f(\cdot|\hat{\theta}_N)$ where $f(\cdot|\theta)$ is a pre-determined parametric family such as the Gaussian family with unknown mean and variance, and $\hat{\theta}_N$ is a good estimator (e.g. the maximum likelihood estimator) of θ based on x^N. On the other hand, the nonparametric predictive density can be any good nonparametric density estimator based on x^N, for example the kernel or log-spline density estimators. In this paper, we show that a parallel estimation theory can be established based on quantized or lossy data.

Recently, Ortega and Vetterli (1994) proposed an adaptive quantization algorithm based on the causal past and they also presented convincing experimental results. Their approach differs from other adaptive quantization in that there is no separate training data set needed for the quantization – the quantizer is re-designed sequentially based on causal past quantized sample. Following Ortega and Vetterli (1994), we divide the problem into an estimation part and a quantization part. For the former, the underlying density is estimated based on the causal quantized data, and for the latter a new (optimal) quantization algorithm (e.g. Lloyd-Max) is designed based on the estimated density. Here we concentrate mainly on the estimation part.

We now describe a statistical estimation model which lends itself to a theoretical analysis. This model is a good approximation to situations where the quantization levels are stablized, and these levels don't have to be the optimal levels based on the unknown density.

Let an L-level quantization of [a,b] (which can be an infinite interval) correspond to an (interval) partition $\{A_i\}_{i=1}^L$, and assume we only observe the quantized causal past data x^N, i.e., we observe only the indicators $I_{\{x_j \in A_i\}}$. Denote by $n_i(N)$ (i=1,..., L) the counts of x's falling into intervals A_i. Assume the source process is stationary and ergodic with a k-dimensional parametric marginal density $f(x|\theta)$ ($\theta \in R^k$). Let $P_i = P_i(\theta) := \int_{A_i} f(x|\theta)dx$. Under regularity conditions on the parametric family, when $k = L$, the above equations uniquely determine θ in terms of P_i's: $\theta = g(P_1, ..., P_L)$. By the Ergodic Theorem, for N large, $n_i/N \approx P_i$; hence $\hat{\theta} := g(n_1/N, ..., n_L/N)$ tends to the true θ as N gets large. That is, quantized sample leads to consistent estimation of the unknown density when the source is stationary and ergodic, and as long as the marginal density is parametric with dimension less than the level of quantization – "one needs at least the number of equations as the number of unknowns." (In the case that $k < L$, we solve for θ using the k equations corresponding to the k largest n_i/N; or we minimize $\sum_i (P_i(\theta) - n_i/N)^2$.) When the CLT holds for the stationary process, the asymptotic normality of $\hat{\theta}$ is expected.

If we further assume that the source process is memoryless, then maximum likelihood method can be used to estimate θ based on n_i: $\hat{\theta}_{mle} =$ arg. max. $\prod_i P_i(\theta)^{n_i}$. In particular, the Monte-Carlo EM (Expectation-Maximization) or data argumentation algorithm (cf. Wei-Tanner, 1990) can be used to find $\hat{\theta}_{mle}$ if we view the unobserved x's as the complete data and the n_i's as the observed incomplete data. This algorithm is especially useful when the MLE based on the complete data has a closed form such as in the case of the Gauss family. Moreover, the linear-interpolation estimation method in Ortega and Vetterli (1994) can be viewed in our framework as follows: let $\theta = (f(a_1), ..., f(a_k))$, where a's are pre-chosen points in $[a, b]$, for example, centers of A_i's. Then $f(\cdot|\theta)$ is the density determined by linearly interpolating the $f(a)$'s.

Currently under investigation are non-parametric estimation methods and using MDL to select window sizes on which the causal past is based. Simulation studies are also planned to test the estimation methods when used together with a quantization algorithm such as the Lloyd-Max algorithm.

REFERENCES

[1] Dawid, P. "Present position and potential developments: some personal views, statistical theory, the prequential approach," *J. Royal Statist. Soc. A* **147** 278-292, 1984.

[2] Ortega, A. and Vetterli, M. "Adaptive quantization without side information," preprint.

[3] Rissanen, J. *Stochastic complexity in statistical inquiry.* World Scientific, Singapore.

[4] Wei, C. and Tanner, M. "A Monte Carlo implementation of the EM algorithm and the Poor Man's data augmentation algorithm," *J. Amer. Statist. Assoc.* **85** 699-704, 1990.

[1]This work was partially supported by ARO Grant DAAH04-94-G-0232 and NSF Grant DMS-9322817.

Probability Quantization for Multiplication-Free Binary Arithmetic Coding[1]

Kar-Ming Cheung
Jet Propulsion Laboratory, 4800 Oak Grove Dr.,
Pasadena, CA 91109

Abstract — **We describe an efficient probability quantization scheme for binary arithmetic code implementation. We show that this scheme is simple to implement and has better compression efficiency than some existing schemes.**

I. Introduction

The binary arithmetic code is a crucial element of many practical state-of-the-art lossless and lossy compression schemes. The key to an efficient implementation of the binary arithmetic coding procedure is to avoid performing the time-consuming multipication and division operations in the probability update for each binary symbol sent. IBM's QM-coder [1] keeps track of two fixed-length registers A and C, where A represents the size of the current interval, and C indicates the base of the current interval. By means of a normalization process A and C are kept within a specific range. By a simple approximation that requires A to be in the range of $0.75 \leq A < 1.5$, the QM-coder replaces multiplications with simple additions and subtractions. Using a binary entropy argument, the worst-case efficiency can be shown to be about 97.0%. Langdon et. al. proposed a more intuitive approach to perform binary arithmetic coding [2]. Langdon's binary arithmetic coding procedure keeps track of two values *high* and *low*, where *high* and *low* correspond to the top and the bottom of the current interval. Langdon suggested to constrain the probability of the less probable symbol to the nearest integral power of $\frac{1}{2}$, so that multiplications can be replaced by simple shifts. The worst-case efficiency of Langdon's binary arithmetic code can be shown to be about 95.0%.

II. Probability Quantization Scheme

In this article we improve upon Langdon's results by approximating the probability of the less probable symbol with a fraction of the form 2^{-l} or $2^{-l-1} + 2^{-l-2}$ for $l = 1, 2, \ldots$. It is easy to show that multiplying a number by $2^{-l-1} + 2^{-l-2}$ is equivalent to right-shifting it by $l + 0.415$ bits. Computationally this correponds to replacing a multiplication operation with 2 shifts and an add. As we will show later, this scheme improves the worst-case coding efficiency to 98.5%.

The following is a sketch on how to optimally quantize the probability of the less probable symbol to achieve the aforementioned computational efficiency. We use a similar approach as Langdon's. Let p, $0 < p \leq 0.5$, be the true probability of the less probable symbol. The question is to choose a step-wise probability quantization function $Q(p)$ of p such that the average code length per symbol, namely $p\log_2(Q(p)) - (1-p)\log_2(1-Q(p))$, is minimized. The design of $Q(p)$ is complexity-driven, not performance-driven. However as we will show later, that we do not sacrifice much by quantizing p into the form 2^{-l} or $2^{-l-1} + 2^{-l-2}$ for $l = 1, 2, \ldots$. We examine two different cases.

Case 1: $2^{-l-1} + 2^{-l-2} < p \leq 2^{-l}$
This corresponds to finding the breakpoint p' such that

$$p'(l + 0.41504) - (1-p')\log_2(1 - 2^{-l-1} - 2^{-l-2}) = p'l - (1-p')\log_2(1 - 2^{-l})$$

Case 2: $2^{-l-1} < p \leq 2^{-l-1} + 2^{-l-2}$
This corresponds to finding the breakpoint p' such that

$$p'(l+1) - (1-p')\log_2(1 - 2^{-l-1}) = p'(l + 0.41504) - (1-p')\log_2(1 - 2^{-l-1} - 2^{-l-2})$$

We use the same performance efficiency definition as Langdon's, which is given by the entropy as a fraction of the average code length,

$$\text{efficiency} = \frac{-p\log_2 p - (1-p\log_2(1-p)}{pQ(p) - (1-p)\log_2(1-2^{-Q(p)})}.$$

We tabulate the optimal probability range and the worst-case efficiency for each quantized probability value (Figure 1).

References

[1] W. Pennebaker, and J. Mitchell, *JPEG: Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993.

[2] G. Langdon and J. Rissanen, "A Simple General Binary Source Code, *IEEE Trans. Inform. Theory*, Vol. IT-28, Sept. 1982.

Prob. Range	Right-Sft	Probability	Efficiency
0.437 - 0.500	1	0.5	0.988
0.310 - 0.437	1.415	0.375	0.985
0.218 - 0.310	2	0.25	0.994
0.155 - 0.218	2.415	0.1875	0.991
0.109 - 0.155	3	0.125	0.996
0.077 - 0.109	3.415	0.09375	0.994
0.054 - 0.077	4	0.0625	0.997
0.039 - 0.054	4.415	0.046875	0.995
0.027 - 0.039	5	0.03125	0.998
0.019 - 0.027	5.415	0.0234375	0.996
0.014 - 0.019	6	0.015625	0.998

Figure 1

[1]This work was carried out by Jet Propulsion Laboratory, California Institute of Technology, under a contract with National Aeronautics and Space administration

Affine Index Assignments for Binary Lattice Quantization with Channel Noise[1]

András Méhes and Kenneth Zeger

Coordinated Science Lab., Dept. of Elect. and Comp. Engineering, University of Illinois, Urbana-Champaign, IL 61801
email: zeger@uiuc.edu.

Abstract — **A general formula is given for the MSE performance of affine index assignments for a binary symmetric channel with an arbitrary source and a binary lattice quantizer. The result is then used to compare some well-known redundancy free codes. The binary asymmetric channel is considered for a uniform input distribution and a class of affine codes.**

Two major issues in noisy channel vector quantization are complexity and sensitivity to channel errors. Structured vector quantizers and index assingments provide a low complexity solution for enhancing channel robustness.

A d-dimensional, n-bit *noisy channel VQ* with *index set* $\mathcal{I} = \{0, 1, \ldots, 2^n - 1\}$, and *code book* $\mathcal{C} = \{\mathbf{y}_i \in \mathbb{R}^d : i \in \mathcal{I}\}$ is a functional composition $Q = \mathcal{D} \circ \pi^{-1} \circ \eta \circ \pi \circ \mathcal{E}$, where $\mathcal{E}: \mathbb{R}^d \to \mathcal{I}$ is the *quantizer encoder*, $\mathcal{D}: \mathcal{I} \to \mathcal{C}$ is the *quantizer decoder*, $\pi: \mathcal{I} \to \mathcal{I}$ is the *index permutation*, and $\eta: \mathcal{I} \to \mathcal{I}$ is a random permutation representing the channel.

A *binary lattice quantizer* is a vector quantizer, whose codevectors are of the form $\mathbf{y}_i = \mathbf{y}_0 + \sum_{l=0}^{n-1} \mathbf{v}_l i_l$ for $i \in \mathcal{I}$, where the ordered set of vectors $\mathcal{V} = \{\mathbf{v}_l\}_{l=0}^{n-1}$ is called the *generating set*, and $i_l \in \{0, 1\}$ is the l^{th} bit in the binary expansion of the index i (here i_0 is the LSB). A binary lattice quantizer is equivalent to a direct sum quantizer (or multistage or residual quantizer) with two code vectors per stage. Examples include truncated lattice vector quantizers (e.g. uniform quantizers). A binary lattice VQ is similar to the non-redundant version of the LMBC-VQ (VQ by Linear Mappings of Block Codes) presented in [3].

An *affine index assignment* is an assignment of the form

$$\pi(i) = \vec{i}\mathrm{G} \oplus \vec{d}, \qquad \pi^{-1}(i) = (\vec{i} \oplus \vec{d})\mathrm{F}, \quad (\mathrm{F} = \mathrm{G}^{-1})$$

where G is the *generator matrix*, $\vec{d}$ is the *translation vector*, and the operations are performed over $GF(2)$. Many popular redundancy free codes are affine, including the Natural Binary Code (NBC), the Folded Binary Code (FBC), and the Gray Code (GC).

For a given source $\mathbf{X}$, the *Hadamard transform* of its distribution is defined as $\hat{P}_l = \sum_{i \in \mathcal{I}} \mathrm{P}[\mathcal{E}(\mathbf{X}) = i](-1)^{\langle \vec{l}, \vec{i} \rangle}$.

The MSE of a quantizer that satisfies the centroid condition, can be decomposed as $D = D_S + D_C$, where

$$D_S = \sum_{i \in \mathcal{I}} \mathrm{E}\left[\|\mathbf{X} - \mathbf{y}_i\|^2 \mid \mathcal{E}(\mathbf{X}) = i\right] \mathrm{P}[\mathcal{E}(\mathbf{X}) = i]$$

$$D_C = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \mathrm{P}[\mathcal{E}(\mathbf{X}) = i] \mathrm{P}[\pi(j) | \pi(i)].$$

Theorem 1 *The channel distortion of a 2^n point binary lattice vector quantizer with generating set $\{\mathbf{v}_l\}_{l=0}^{n-1}$, which uses an affine index assignment with generator matrix G to transmit across a binary symmetric channel with crossover probability q, is given by*

$$D_C = \frac{1}{4} \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \langle \mathbf{v}_k, \mathbf{v}_l \rangle \hat{P}_{2^k + 2^l} \times \left(1 - 2(1-2q)^{w(\vec{f}^T_{n-k})} + (1-2q)^{w(\vec{f}^T_{n-k} \oplus \vec{f}^T_{n-l})}\right),$$

where $w(\cdot)$ denotes Hamming weight, $\vec{f}^T = [f_{1,k}, \ldots, f_{n,k}]$ is the k^{th} column of $\mathrm{F} = \mathrm{G}^{-1}$, $\hat{P}_l$ is the l^{th} component of the Hadamard transform of the induced discrete distribution on the encoder cells, and $\oplus$ indicates modulo 2 addition.

Let FBC* denote the "best" Folded Binary Code obtained by reordering the generating set $\mathcal{V}$ to minimize D_C, and let $\mathbf{U}$ be a uniform discrete random variable on the code points.

Corollary 1 *Given the conditions of Theorem 1 (and $q < 1/2$), $D_C^{(FBC^*)} > D_C^{(NBC)}$ if and only if*

$$\mathrm{Var}[Q(\mathbf{X})] + \mathrm{E}^2[Q(\mathbf{X}) - \mathbf{U}] > \frac{\max_{\mathbf{v} \in \mathcal{V}} \|\mathbf{v}\|^2}{\sum_{\mathbf{v} \in \mathcal{V}} \|\mathbf{v}\|^2} \mathrm{Var}[\mathbf{U}]$$

For a uniform (discrete) distribution on the code vectors, and a binary symmetric channel the NBC is the optimal index assignment [1], [2]. An *affine translate of the NBC* is an index assignment of the form $\pi(i) = \vec{i} \oplus \vec{d} = \pi^{-1}(i)$.

Theorem 2 *If a 2^n point binary lattice vector quantizer induces equiprobable encoder cells for a given source, and transmits an affine translation of the Natural Binary Code across a binary asymmetric channel with crossover probabilities $\mathrm{P}[1|0] = p$ and $\mathrm{P}[0|1] = q$, then the channel distortion is minimized if and only if the translation vector $\vec{d}$ satisfies*

$$\vec{d} = \operatorname*{argmin}_{i \in \mathcal{I}} \|\mathbf{y}_i - \mathrm{E}[\mathbf{U}]\|$$

[1]The research was supported in part by the National Science Foundation under Grants No. NCR-92-96231 and INT-93-15271.

References

[1] T. R. Crimmins, H. M. Horwitz, C. J. Palermo and R. V. Palermo "Minimization of Mean-Square Error for Data Transmitted Via Group Codes," *IEEE Trans. Info. Theory*, IT-15, no. 1 pp. 72-78, January 1969.

[2] S. W. McLaughlin, D. L. Neuhoff and J. J. Ashley, "Optimal Binary Index Assignments for a Class of Equiprobable Scalar and Vector Quantizers," preprint.

[3] R. Hagen and P. Hedelin, "Design Methods for VQ by Linear Mappings of Block Codes," *Proc. IEEE Int. Symp. Information Theory*, Trondheim, Norway, p. 241, 1994.

[4] A. Méhes and K. Zeger, "Redundancy Free Codes for Binary Discrete Memoryless Channels," Proceedings of the 1994 CISS, Princeton, NJ, pp. 1057-1062.

Optimal Quantization for Finite State Channels[1]

Tolga M. Duman Masoud Salehi
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115

Abstract — **A quantizer design algorithm for transmission over finite state channels is presented. Optimal design algorithms for a variety of conditions regarding the knowledge of the state information at the transmitter and the receiver are derived. Both cases of noiseless and noisy observations are considered.**

I. Summary

We want to transmit the output of an information source to a receiver over a finite state channel with two states. In general, the entropy rate of the source is too high, and therefore we need to quantize the source output to make it suitable for transmission. Our objective is to design the quantizer to minimize the mean squared error (MSE) when the channel is in state S_1, subject to a constraint on the MSE when the channel is in state S_2. In other words, the problem is to minimize

$$D_1 = E[(X-\hat{X})^2|\text{channel state is } S_1]$$

subject to $D_2 = E[(X-\hat{X})^2|\text{channel state is } S_2] \leq D$

where X is the source output, $\hat{X}$ is the reconstructed output at the receiver, and D is the maximum allowable distortion when the channel state is S_2.

Let $P_m(k|i)$ be the probability of receiving k as the channel output when the channel input is i and the channel state is S_m, where $m = 1,2$, $i \in \{1,2,...,N_1\}$ and $k \in \{1,2,...,N_2\}$. We also assume that noisy state information is available both at the transmitter and the receiver. Let t_{jl} and r_{jl} be the probability that state S_l is perceived as state S_j at the transmitter and the receiver, respectively. Denote the i^{th} quantization region by A_{mi} when the channel state is perceived as S_m at the transmitter, and the k^{th} reconstruction level by $g_n(k)$ when the channel state is perceived as S_n at the receiver.

To design the quantizer, our approach is to convert the constrained optimization problem to an equivalent unconstrained minimization problem by the method of Lagrange multipliers, i.e., to minimize $L = D_1 + \lambda(D_2 - D)$, $\lambda \geq 0$ is a constant. By optimizing the encoder structure for a fixed decoder and the decoder structure for a fixed encoder, we obtain the necessary conditions for the optimality of the quantizers. This results in the following algorithm for the quantizer design as derived in [1].

Algorithm: Optimal quantizer design to minimize L for a fixed λ.

1) Start with an initial encoder structure.

2) Find the optimal decoder structure for the current encoder structure by using

$$g_n(k) = E[X|\text{channel output is } k \text{ and channel state is } S_n] \quad (1)$$

for all $k \in \{1,2,...,N_2\}$, $n = 1,2$.

3) Find the optimal encoder structure for the current decoder structure by using

$$A_{mi} = \{x : 2\alpha_{mi}x - \beta_{mi} \geq 2\alpha_{mi'}x - \beta_{mi'}, \ \forall i' \neq i,\ i', i \in \{1,2,...,N_1\}\} \quad (2)$$

where

$$\alpha_{mi} = \sum_{j=1}^{2}\lambda_j \sum_{n=1}^{2} t_{mj}r_{nj} \sum_{k=1}^{N_2} P_j(k|i)g_n(k),$$

$$\beta_{mi} = \sum_{j=1}^{2}\lambda_j \sum_{n=1}^{2} t_{mj}r_{nj} \sum_{k=1}^{N_2} P_j(k|i)g_n^2(k)$$

$i \in \{1,2,...,N_1\}$, $m = 1,2$, $\lambda_1 = 1$, $\lambda_2 = \lambda$.

4) If the change in L is below a prespecified threshold, stop. Otherwise, go to step 2.

The Lagrangian is non-increasing at each step and is bounded from below, therefore the algorithm is always convergent. A numerically efficient algorithm to find A_{mi} in (2) is presented in [2].

In order to complete the solution of the problem one has to vary the Lagrange multiplier ($\lambda \geq 0$), apply the design algorithm, obtain a set of achievable distortion pairs, and convexify these points. The last step is justified by the use of time-sharing. In [1] it is illustrated that, in general, time-sharing is necessary to obtain the optimal performance of the system. A set of numerical examples where the above algorithm is employed is also presented.

We also consider the quantizer design problem when the observation is noisy. It is shown that the problem of optimal quantizer design for noisy observations can be separated into two parts — first estimating X from the observation in the MSE sense, and then using the quantizer design algorithm for no observation noise.

References

[1] T. M. Duman, "Multiterminal quantization for distributed detection and estimation," M.S. Thesis, Northeastern University, Boston, MA, May 1995.

[2] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory.*, vol. IT-39, pp. 821-834, May 1993.

[1] This work was partially supported by the NSF Grant NCR-9101560

Performance of the Adaptive Quantizer

Nina I. Pilipchouk
(Moscow)

Abstract — **The probabistic analysis of the adaptive quantizer *DH* is presented. For the case when the number of quantizer levels is equal to 4, Mean square error - Average Entropy functions are calculated.**

I. Introduction

The Adaptive *DPCM* (*ADPCM*) has been recommended by CCITT [1] to implement in communication.

The exact mathematical analysis of the *APCM* systems is rather sophisticated [2] and therefore there is not the complete analysis of any of them. In this paper, it is presented the complete probabilistic analysis of the simple variant of the adaptive quantizer *DH* [3] as well as the comparison with Max's quantizer [4].

II. Main Performance

Consider the adaptive uniform quantizer *DH* with $N = 4$ quantizer levels, the variable size of the quantizer step h, and the variable size of the quantizer range d. The step and the range at the sampling instant t_{k+1} depends on their values and a value of an input signal at the preceding sampling instant t_k (see, for details, [3]). The adaptive quantizer is equivalent to the two virtual quantizers with steps $h_1 = h$ and $h_2 = 2h$, respectively. The first quantizer is used for the small values of the input signal, and the second one is used for the large values.

The adaptive quantizer is designed to reduce the value of the entropy of quantized signal for a given error in comparison with Max's nonadaptive quantizer [4].

Main performance is the Mean square error - Average Entropy function. In [3], it is shown that the joint probability distribution of the input signal and parameters $w_i(y)$ $(i = 1, 2)$ is given by the equations

$$w_1(y) = \int_{|x|\le h_1} (w_1(x,y)+w_2(x,y))\,dx; w_2(y) = w(y)-w_1(y),$$

where $w(y)$ is one-dimensional probability density of the input signal, $w_i(x,y)$ is the joint probability of the samples x, y and parameters $i = 1, 2$. Put $h_1 = h$, $h_2 = 2h$.

For this case the solution of the equations is as follows:

$$w_1(y) = \int_{|x|\le h} w(x,y)dx;\ w_2(y) = \int_{|x|>h} w(x,y)dx;$$

where $w(x,y)$ is two-dimensional probability density of the input signal.

We consider the Gaussian input signal with zero mean value, variance 1 and correlation coefficient between two adjacent samples ρ. For this case , the average mean square error and the average entropy can be rewritten as follows:

$$\varepsilon^2 = 2\int_0^h (y-h/2)^2 w(y) f_1(y)dy + 2\int_h^\infty (y-3h/2)^2 w(y) f_1(y)dy +2\int_0^{2h} (y-h)^2 w(y) f_2(y)dy + 2\int_{2h}^\infty (y-3h)^2 w(y) f_2(y)dy,$$

$$H = -2\int_0^h w(y)f_1(y)dy \log(1/Q_1 \int_0^h w(y)f_1(y)dy) -2\int_h^\infty w(y)f_1(y)dy \log(1/Q_1 \int_h^\infty w(y)f_1(y)dy) -2\int_0^{2h} w(y)f_2(y)dy \log(1/Q_2 \int_0^{2h} w(y)f_2(y)dy) -2\int_{2h}^\infty w(y)f_2(y)dy \log(1/Q_2 \int_{2h}^\infty w(y)f_2(y)dy),$$

where

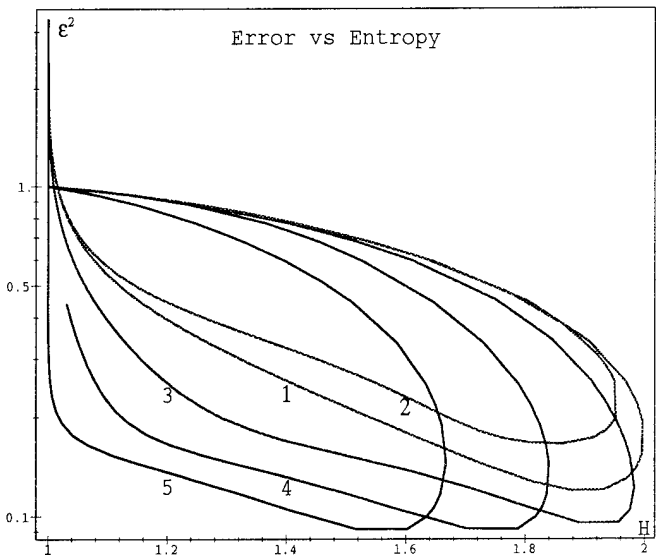
$w(y) = \exp(-y^2/2)\sqrt{2\pi}$;

$f_1(y) = 1/\sqrt{2\pi}\int_{-(h+\rho y)/\sqrt{1-\rho^2}}^{(h-\rho y)/\sqrt{1-\rho^2}} \exp(-z^2/2)dz$; $f_2(y) = 1 - f_1(y)$,

$Q_1 = \int_{|x|\le h} w(x)dx$; $Q_2 = 1 - Q_1$.

III. Numerical Results

Main performance, i.e., the Mean square error - Average entropy function, is presented in the Fig. for the nonadaptive quantizer (1), and for the adaptive quantizer and $\rho =$ 0, 0.9, 0.99, 1 (2, 3, 4, 5). One can see that in the region where the error is minimal the value of the error changes slowly with H. Hence, we can get an extra gain in data compression if we use suboptimal values of h. For example, if it is allowed to have the error $\varepsilon^2 = 0.119$, which is minimal for the nonadaptive quantizer, then we can get the gain about 0.62 bit per sample for large ρ.



References

[1] M.H. Sherif, D.O. Bowker, G. Bertocci, B.A. Orford, G.A. Mariano, "Overview and Performance of CCITT/ANSI Embedded ADPCM Algorithms," *IEEE Trans. Commun.*,Vol. COM-41, pp. 391-399, Feb. 1993.

[2] N.I. Pilipchouk, G.R. Nadezhdina, "APCM Performance Improvement by Means of Entropy Coding," In *Proceedings of the IEEE ISIT*, page 196 , Trondheim, Norway, July 1994.

[3] N.I. Pilipchouk, V.P. Jakovlev, "Adaptive Pulse Code Modulation,"-*Moscow: Radio i Svjaz*, 1986 (in Russian).

[4] J.Max, "Quantizing for Minimum Distortion," *IRE Trans.Inform. Theory*, pp.7-12, March 1960.

Combined Multipath and Spatial Resolution for Multiuser Detection: Potentials and Problems

Howard C. Huang, Stuart C. Schwartz, and Sergio Verdú
Department of Electrical Engineering; Princeton University; Princeton NJ, 08544

I. Introduction

Mobile telephony in CDMA channels encounters a variety of communication challenges including fading due to multipath (MP) and multiaccess interference (MAI) due to simultaneous transmissions from interfering users. Detectors which employ multiuser detection and temporal (RAKE-type) combining have been shown [1] to provide near-far resistant solutions which effectively combat both of these impediments. In the first part of this paper, we address the potential gains of using spatial combining in conjunction with multiuser detection and temporal combining. In the second part of the paper, we examine an adaptive multiuser detector which is well suited for MAI-limited MP channels.

II. Multiuser Array Detection for Multipath Channels

Recently, efforts have been made to combine the use of temporal combining and spatial combining. Most of these efforts are based on conventional detection schemes which have been shown to be near-far limited. In the first part of this paper, we combine results from [1] and [2] to derive a class of near-far resistant detectors which uses a linear multiuser detector in conjunction with spatial and temporal combiners. It is shown that the optimum (in terms of near-far resistance) linear multiuser detector with an array of P sensors consists of a bank of match filters at each sensor matched to the users' delayed spreading codes, followed by a spatial combiner (which acts as a beamformer pointing in the direction of each users' MP signals), a temporal combiner (which coherently combines a user's MP components), and a linear transformation which decorrelates the users. Since this decorrelation process relies on the estimates of the signals' spatial and MP parameters, this detector (known as the spatial-temporal decorrelator (stD)) is near-far limited when the estimates are not exact. By interchanging the order of the three processors, we can obtain two suboptimum detectors, the sDt and Dst, which, respectively, retain their near-far resistant characteristics when there is MP parameter mismatch and when there is both MP and spatial parameter mismatch. If all of the system parameters are known exactly, we have the following relationship among their respective bit error rates as a function of the noise level: $P_{stD}(\sigma) \leq P_{sDt}(\sigma) \leq P_{Dst}(\sigma)$. This result is illustrated in Figure 1 for a 2-user synchronous, coherent system where each user contributes $L = 2$ MP components and where there are $P = 2$ sensors.

III. Blind Adaptive Detection for Multidimensional Signals

Motivated by the need for a noncoherent multiuser detector for MP channels which has no *a priori* knowledge of the interfering users, in the second part of the paper, we derive an extension of the blind adaptive detector [3] for differentially encoded, multidimensional signals. Such a detector is ideally suited for MP channels since, if we assume negligible ISI, the spanning set for the multidimensional subspace is given by the truncated, delayed translates of the desired user's spreading code. Given the L-dimensional subspace in which the desired user's signal lies, we can obtain an arbitrary orthogonal basis $\mathbf{z}_1 \ldots \mathbf{z}_L$. The resulting detector consists of a bank of L linear filters followed by an inner-product operation between the current filter bank output and that from the previous bit interval; the bit estimate is the hard-limit of this inner-product. The l^{th} filter consists of a real part $(\mathbf{z}_l + \mathbf{x}_l^R)/||\mathbf{z}_l + \mathbf{x}_l^R||$, $l = 1 \ldots L$ which operates on the real part of the received signal and a corresponding imaginary part. The $\mathbf{x}_l^R$ and $\mathbf{x}_l^I$ are each constrained to be orthogonal to all of the basis vectors $\mathbf{z}_1 \ldots \mathbf{z}_L$ and are obtained adaptively using the output energy of the respective real and imaginary part of the l^{th} filter. Each of the $\mathbf{x}_l^R$ and $\mathbf{x}_l^I$ can be adapted independently, exhibits global convergence, and requires knowledge of only $\mathbf{z}_l$ and the timing (bit-epoch) of the desired user. Hence this detector requires even less knowledge than the conventional RAKE receiver; yet as seen in Figure 1, for a 2-user system with $L = 2$, it essentially achieves the same performance as the optimum linear MP multiuser detector (equivalent to the differentially coherent stD with P=1).

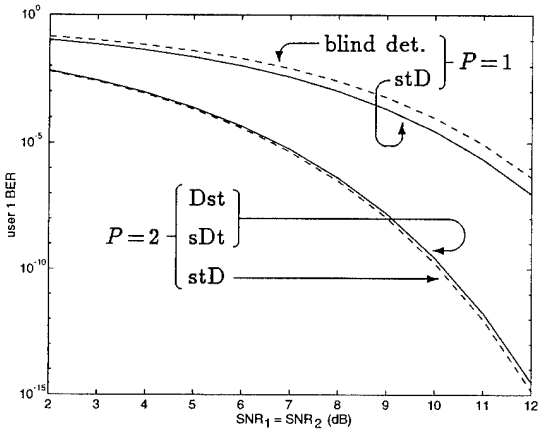


Figure 1: Performance of Multiuser Detectors ($L = 2$)

References

[1] H. Huang and S. Schwartz, "A comparative analysis of linear multiuser detectors for fading multipath channels," in *1994 Globecomm Proceedings*, pp. 11 – 15.

[2] H. Huang and S. Schwartz, "Robust multiuser detection using array sensors," submitted to *IEEE Transactions on Communications*.

[3] M. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Transactions on Information Theory*, July 1995.

MULTI-USER COMMUNICATION WITH MULTIPLE SYMBOL RATES

Michael L. Honig
Dept. of EECS
Northwestern University
Evanston, IL 60208

Sumit Roy
Division of Engineering
University of Texas at San Antonio
San Antonio, TX 78249

Summary

Multi-user communication scenarios, to date, have almost exclusively focussed on the situation in which all users share a common symbol rate $1/T$. However, future multi-media services will require that users with different (and possibly time-varying) data rates share a common transmission channel. We consider the problem of optimizing a multi-user receiver when the users transmit with different symbol rates. The problem of optimizing the transmitter pulse shaping filters for each user assuming different transmitted symbol rates is also considered. We assume the Minimum Mean Squared Error (MMSE) performance criteria.

The kth user generates a sequence of pulses

$$s_k(t) = \sum_i b_k[i]\delta(t - iT_k)$$

where $\{b_k[i]\}$ is the sequence of symbols corresponding to user k, and $1/T_k$ is user k's symbol rate. This signal is the input to a pulse shaping filter with transfer function $P_k(f)$. The channel corresponding to user k is $H_k(f)$, and the additive noise $n(t)$ is assumed to be white. The received signal is therefore

$$y(t) = \sum_{k=1}^{K} \sum_i b_k[i]\{p_k * h_k(t - iT_k)\} + n(t)$$

where K is the number of users, and $p_k * h_k$ is the convolution of the transmitted pulse shape with the channel impulse response. We will assume that there exist nonnegative integers $m_1, \ldots, m_K$ such that $T_1 : T_2 : \cdots : T_K = m_1 : m_2 : \cdots : m_K$, where $m_1 \le m_2 \le \cdots \le m_K$, implying $T_1 \ge T_2 \ge \cdots \ge T_K$.

For systems with multiple rates, the optimum receiver is *periodically time-varying*, due to the underlying cyclostationarity of the sampled received signal. The approach we take is to embed the optimum receiver design problem into an equivalent higher-dimensional problem that is wide-sense stationary. To do so, we 'decompose' the input data stream from user k into $LCM(\mathbf{m})/m_k$ low-rate streams each with a *common* symbol period

$$T_s = \frac{LCM(\mathbf{m})}{m_k} T_k$$

where $\mathbf{m} = (m_1, \ldots, m_K)$, and $LCM(\cdot)$ denotes the least common multiple of the elements of the vector argument. It is convenient to think of the additional streams created by this process as 'fictitious' new users in the system. This procedure effectively yields an equivalent higher dimensional, *single-symbol-rate* multi-input, multi-output communication system with input dimension (corresponding to the total number of 'users') equal to $\sum_{k=1}^{K} \frac{LCM(\mathbf{m})}{m_k}$. Note that for the case $m_k = 1$ for all k, this reduces to multi-user communication with identical symbol periods. Accordingly, the vector of channel transfer functions which corresponds to the embedded system with equal symbol rates is given by

$$\mathbf{H}(f) = \Big[H_1(f), H_1(f)e^{j2\pi f T_1}, \ldots, H_1(f)e^{j2\pi f \tilde{m} T_1}; \cdots; H_K(f), H_K(f)e^{j2\pi f T_K}, \ldots, H_K e^{j2\pi f \tilde{m} T_K} \Big]$$

where $\tilde{m} = LCM(\mathbf{m})/m_K - 1$.

We also consider the optimization of the transmitter pulses $p_1(t), \ldots, p_K(t)$ subject to the power constraints $\int_{-\infty}^{\infty} |P_k(f)|^2 df \le \Pi_k$, $k = 1, \ldots, K$, assuming a linear MMSE receiver. Using the preceding decomposition technique, the problem can again be embedded in a higher-dimensional problem in which the users transmit with the same symbol rate. Necessary conditions for optimality can be derived, and show that FDMA achieves a local optimum (which may be globally optimal). The FDMA solution differs from that given in [1], in that for a particular frequency $0 < f < 1/(2m_1T_1)$, user 1 (user 2) can place power at up to m_1 (m_2) different Nyquist translates, (that is, $f + i/(m_1T_1)$ for different nonnegative integers i).

Numerical results will be presented that illustrate the tradeoff between changing symbol rates and changing the number of constellation points to achieve a given mix of data rates.

Reference

[1] M. Honig and U. Madhow, "Optimization of Transmitter Pulses for 2-User Data Communications", *Proc. 1992 Int. Symposium on Information Theory*, San Antonio, TX, Jan. 1992.

Multisensor Multiuser Receivers for Time-Dispersive Multipath Fading Channels

M. Stojanovic and Z. Zvonar*

Elect. & Comp. Eng. Dept., Northeastern Univ., Boston, MA 02115

*Communications Div., Analog Devices, Wilmington, MA 01887

Abstract — **Receiver structures based on joint MMSE diversity combining, equalization and multiple access interference suppression are discussed. It is shown that receiver complexity can substantially be reduced by exploiting the structure of multipath. Experimental results, obtained in an underwater acoustic channel, demonstrate superior capabilities of the receivers proposed.**

I. Introduction

Due to their superior performance, multiuser receivers are being considered for applications ranging from wideband CDMA systems to bandwidth-efficient multiple-access underwater acoustic (UWA) communication channels [1], [2]. In severely dispersive time-varying channels, multipath propagation presents a major limitation to the system performance. In such a case, multisensor signal processing offers potentials of robustness to fading, reduction of residual intersymbol interference (ISI) [3] and suppression of multiple-access interference (MAI).

II. Receiver Structure

We address the general case of a multipoint-to-point communication system where multiuser signals are subject to ISI and may overlap in both time and frequency. Assuming the presence of L users in a system with K receiving elements, the optimal receiver consists of a combiner followed by a sequence detector, as shown in Fig.1. The l^{th} combiner is optimally

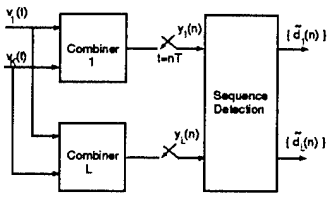


Figure 1: Optimal receiver.

represented as a bank of K matched filters whose outputs are summed and sampled at the symbol rate. The L discrete-time combiner outputs are processed by an $L \times L$ detector, chosen as a MIMO DFE [4]. When the channel is not known, the combiners are realized as banks of fractionally spaced adaptive filters.

III. Reduced-Complexity Adaptive Processing

Although the use of an equalizer eliminates the exponential complexity of the optimal (MLSE) detector, the resulting combiner/equalizer structure may still have complexity prohibitively high for many practical cases. Besides the increase in computational time, a critical disadvantage of large adaptive filters lies in their high noise enhancement, which ultimately limits the gain obtained by increasing the number of input channels. These issues motivate the search for a different combining strategy in which the size of the combiner will be reduced, but multichannel processing gain preserved.

By modeling the channel as consisting of a finite number of propagation paths, it is revealed that the optimal combiner can equivalently be realized using fewer matched filters. The resulting adaptive combiner is shown in Fig.2. The pre-

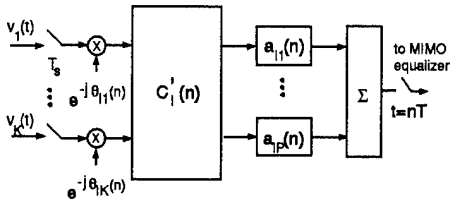


Figure 2: Reduced-complexity adaptive combiner.

combiners $\mathbf{C}_l$ perform spatial processing only, reducing the number of channels from K to $P < K$ for subsequent multichannel equalization. Shown also is the multichannel phase-locked loop which is an essential part of a practical receiver.

When the multipath structure is not known, the approach most beneficial is to conduct unconstrained optimization of the combiners and the equalizers. To preserve performance of the full-complexity receiver, the pre-combiners and the multichannel equalizers need to be optimized jointly. An adaptive algorithm suitable for application in rapidly time-varying channels is a combination of the second-order gradient updates for the carrier phases, and a multiple RLS updates for the coefficients of the combiners and the equalizers.

The methods described above were applied to the real data obtained from experiments conducted in the shallow water acoustic channel, characterized by rapidly time-varying ISI which extends over several tens of symbol intervals. Due to the bandwidth limitation of the channel, only very low spreading ratios can be used (e.g., 3), resulting in increased MAI. The proposed techniques demonstrated superior performance in such conditions.

References

[1] S.Verdu, "Adaptive multiuser detection," in Proc. *IEEE Third International Symposium on Spread Spectrum Techniques and Applications*, Oulu, Finland, July 1994, pp 43-49.

[2] M.Stojanovic and Z.Zvonar, "Adaptive spatial/temporal multiuser receivers for time-varying channels with severe ISI," in Proc. *28th Annual Conference on Information Sciences and Systems*, Princeton, NJ, March 1994, pp 127-132.

[3] P.Monsen, "MMSE equalization of interference on fading diversity channels," *IEEE Trans. Comm.*, vol. COM-32, pp. 5-12, Jan. 1984.

[4] A.Duel-Hallen, "Equalizers for multiple input/multiple output channels and PAM systems with cyclostationary input sequences," *IEEE J. Sel. Areas Comm.*, vol. JSAC-10, pp. 630-639, Apr. 1992.

An Iterative Multiuser Receiver: The Consensus Detector

Alex J. Grant[†][1] and Christian Schlegel[‡]

† Mobile Communications Research Centre
University of South Australia
The Levels, 5095, Australia

‡ Dept. Electrical Engineering
University of Texas at San Antonio
San Antonio, TX 78249, USA

Abstract — **A flexible iterative receiver is proposed for the multiple access channel. The receiver splits the detection problem into a "single user" decoding step followed by a combining step. The structure of the receiver is suitable for many different types of multiple access channels.**

I. Introduction

Motivated by the need for reduced complexity, the success of iterative methods for decoding of concatenated codes [1], and recent theoretical results concerning successive cancellation receivers [2], we propose an iterative detector for co-operative detection of multiuser systems, which for reasons that will become apparent, we have named the consensus decoder.

The detector divides the detection operation into two parts - a "single user" estimation step (in which soft decisions are produced), and a "multiuser" combining step.

We shall consider a general m-user multiple access system, in which user i transmits X_i, drawn independent of other users from a finite alphabet, $\mathcal{X}_i$, $i = 1, 2, \ldots, m$; according to the distribution $p_i(x_i)$. The channel produces output symbols, Y, members of the alphabet $\mathcal{Y}$, according to transition probabilities, $p(y \mid x_1, x_2, \ldots, x_m)$.

II. The Consensus Detector

Consider an m-user system. The operation is as follows. User i adds redundancy to its source data, U_i, via an encoder, producing X_i. We shall restrict each X_i to be drawn from an identical alphabet, $\mathcal{X}$. Without loss of generality, denote the members of $\mathcal{X} = \{0, 1, \ldots, J-1\}$. The channel outputs $Y \in \mathcal{Y}$, according to some transition probability, $p(Y \mid X_1, X_2, \ldots, X_m)$. We shall denote the output alphabet $\mathcal{Y} = \{0, 1, \ldots, K-1\}$.

The detector operates as follows.

1. User i attempts to estimate X_i given Y, treating other users as noise. At each symbol interval, each single user detector outputs soft information, $\underline{p}_i$ which is a vector of probability estimates for each channel input symbol. $\underline{p}_i = [P(X_i = 0 \mid Y), \ldots, P(X_i = J-1 \mid Y)]$
2. The symbol estimator for user i forms a list of possible channel outputs, $\underline{y}_i$ due to the other $m-1$ users. This can be interpreted as an estimate of the channel, treating the other users as part of the channel. Each element of $\underline{y}_i$ has associated with it a probability, which is determined from $\underline{p}_i$.
3. The detector for user i now estimates X_i given Y and the list of possible channel outputs, once again outputting soft information. $\underline{p}_i = [P(X_i = 0 \mid Y, \underline{y}_i), \ldots, P(X_i = J-1 \mid Y, \underline{y}_i)]$
4. Steps 2 and 3 are now repeated as many times as desired.

This procedure separates the detection into a single user step (Step 3) and a combining step (Step 2).

In practice, it is impossible for the symbol estimator to form the full list of possible channel inputs, since there will in general be K^m combinations. For example a 10 user system with 8 channel input symbols, there are already about 1 billion possibilities. Therefore, we shall only keep the L most likely symbols, which can be found with a simple M-algorithm. This is where the system complexity is reduced.

The final output of the detector is in a sense the set of sequences to which the m detectors have "agreed" to, hence the name consensus detector. It is also simple to include "confidence levels" in particular users as follows. If we define a parameter, C_i, to be the confidence we have in user i, 0 denoting no confidence and 1 denoting complete confidence, we adjust the probability estimates from a particular user $\underline{p}_i^* = C_i \underline{p}_i + (1 - C_i)\underline{u}$ where $\underline{u}$ is the uniform distribution, $P(X_i = j) = \frac{1}{J}$, for all $0 \leq j \leq J-1$. This has the effect that as we have less faith in a particular user, we flatten out their distribution, placing more uncertainty (entropy) in their decision.

III. Discussion

The advantages of the proposed detector are as follows.

The complexity of the system may be easily varied to provide different levels of performance. Simulation results have shown that in practice only $L \approx 10$ likely symbols need to be retained. The number of iterations can also be varied. In general at most $4-5$ iterations are required, usually less. The reduced complexity nature of the detector gives a complexity that increases only linearly with the number of users.

The structure of the consensus detector is suitable for use with many multiple access channels. All that is required is a suitable single user detector to perform step 3 of the algorithm.

The system may be biased according to previous knowledge about the users, for example different power levels, through the use of confidence levels.

References

[1] C Berrou, A Glavieux, and P Thitmajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes", in *IEEE Int. Conf. Comms.*, 1993.

[2] A Grant, R Urbanke, B Rimoldi and P Whiting, "Single user coding for the discrete memoryless multiple access channel", in *IEEE Symp. Info. Theory*, 1995.

[1]Supported in part by Telecom Australia under Contract No.7368 and by the Commonwealth of Australia under International S & T Grant No.56.

Blind Multiuser Deconvolution in Fading and Dispersive Channels[1]

Javier R. Fonollosa†, José A. R. Fonollosa†, Zoran Zvonar‡and Josep Vidal†

†Universitat Politècnica de Catalunya
Department of Signal Theory and Communications
08034 Barcelona, SPAIN
fono@tsc.upc.es

‡Analog Devices
Communications Division
Wilmington MA 01887-1024, USA
Zoran.Zvonar@analog.com

Abstract — **An adaptive near-far resistant technique for the blind joint multiuser identification and detection in asynchronous CDMA systems is analyzed in fading and dispersive GSM channels.**

I. Introduction

Multiuser detection in CDMA systems usually requires either knowledge of the transmitted signature sequences and channel state information or use of a known training sequence for adaptation. Consequently blind adaptive multiuser receivers have gained considerable attention [1]. We recently proposed a joint multiuser deconvolution scheme [2] characterized by:

- No knowledge of timing, channel state information or signatures nor use of training sequences is required for any user.
- The estimate of the signature sequence of each user convolved with its physical channel impulse response is provided after initial convergence.
- The blind multiuser detector is near-far resistant.

The purpose of this paper is to further investigate the behavior of this scheme in fading and dispersive channels.

II. System Model

We consider the asynchronous CDMA channel

$$r(t) = \sum_n \sum_{k=1}^{K} b_k[n] h_k(t - nT, t) + \sigma w(t) \qquad (1)$$

where $h_k(t - nT, t)$ is the overall complex channel impulse response, given by the convolution of the signature sequence, physical radio channel and the receiving filter impulse responses. It incorporates the amplitude and the delay for user k, and its duration is assumed to be smaller or equal to L symbols, i.e. $h_k(\tau, t) = 0, \tau < 0, \tau > LT, \forall t$. The total number of active users is K and their transmitted sequences are binary independent symbols $b_k[n] \in \{1, -1\}$. The symbol rate is $1/T$ and $w(t)$ is normalized white Gaussian noise. The CDMA channel is sampled at a rate $M/T = 1/T_s$ to derive the vector sequence $\mathbf{r}[n]$

$$\mathbf{r}[n] = [r(nT), r(nT+T_s), \ldots, r(nT+(M-1)T_s)]^T \ . \qquad (2)$$

The observation $\mathbf{r}[n]$ is modeled as a probabilistic M length vector sequence of a state vector $\mathbf{s}[n]$

$$\mathbf{r}[n] = \mathcal{H}[n]\mathbf{s}[n] + \mathbf{w}[n] \ , \qquad (3)$$

where $(M \times KL)$ matrix $\mathcal{H}[n]$ depends of the overall discrete impulse response for all users and $\mathbf{w}[n]$ is the normalized noise vector. There are $N = 2^{LK}$ possible state vectors corresponding to L binary symbols of K users.

III. Blind Identification and Detection Algorithm

If the overall impulse response for each user was known, that is if the signature sequence, physical channel impulse response, amplitude and delay corresponding to each user were available, then using this information, the Viterbi algorithm could be employed to determine the multiuser maximum-likelihood transmitted sequence. In the method we presented however, the Viterbi algorithm is applied with current estimates of the overall impulse responses which are updated recursively after arbitrary initialization. The number of users (K) is assumed known together with a bound for the impulse response duration (L). A similar approach was proposed for the blind equalization of single user channels using the Viterbi algorithm [3] and the Baum-Welch identification algorithm [4]. Specific to the multiuser approach is the procedure which overcomes the convergence to a local minimum [2].

IV. Behavior in Fading and Dispersive Channels

The blind multiuser algorithm has been tested using the mobile radio channel model for typical urban areas (Type 1) TUX60, as defined in [5]. Simulations indicate that, for moderate Doppler frequency (50 Hz) and multipath spread (1.35 symbols), convergence can still be attained within few hundred symbols. Afterwards, the algorithm is still able to track slow channel variations. Possible modification of the receiver, after the initial convergence, may include a simpler decision-directed MMSE scheme.

[1] Work supported by CIRIT of Catalonia (GRQ93-3021).

References

[1] S. Verdú, "Adaptive Multiuser Detection," *Proc. Third International Symposium on Spread Spectrum Techniques and Applications*, Oulu, Finland, pp. 43-50, July 1994.

[2] J. R. Fonollosa, J. A. R. Fonollosa, Z. Zvonar, and J. Vidal, "Blind Multiuser Identification and Detection in CDMA Systems," *Proc. IEEE ICASSP-95*, pp. 1876-1879, May 1995.

[3] N. Seshadri, "Joint Data and Channel Estimation using Blind Trellis Search Techniques," *IEEE Trans. on Communications*, vol 42, pp. 1000-1011, March 1994.

[4] J. A. R. Fonollosa and J. Vidal, "Application of Hidden Markov Models to Blind Channel Characterization and Data Detection," *Proc. IEEE ICASSP-94*, pp. IV 185-188, April 1994.

[5] GSM recommendation 05.05 (version 3.11.0).

On the Least Possible Decoding Error Probability for Truly Asynchronous Single Sequence Hopping

Sándor Csibi
Dept. of Telecom., Tech. Univ. of Budapest,
Stoczek u. 2, H-1111 Budapest, Hungary

Abstract — **Unslotted asynchronous multiple access without feedback is considered. Poisson population, least length single sequence hopping and interleaved outer coding with guard spaces are assumed. Bounds from both sides on the least decoding error probability are proved to vanish with rate $\frac{1}{\kappa'}$ (from a given finite source block length κ' on) and under further conditions defined precisely in a companion preprint.**

I. Introduction

For slotted (frame) asynchronous least length single sequence hopping and a single inner R-S code, bounds from both sides on the least possible decoding error probability have been already obtained by the same author, that disappear with the source block length κ at rate $\frac{1}{\kappa}$, far not exponentially ([1]). This is the price (in error probabity) of being constrained to this simple kind of multiple access. It is, obviously, a question of interest under what additional conditions can, for the same decoding error probability, the very same decay rate $\frac{1}{\kappa}$ be proved also for truly (unslotted) asynchronous access, without assuming any common clock for signal trasmission. It will be shown next that (i) proper kind of interleaving, and (ii) keeping silence (inserting a dummy guard space, just at one end of each message carrying interval as in [2]) are the additions to the model, sufficient for so doing. The question will be answered by Theorem 1.

II. More on the underlying model

Infinite source population is assumed, with demands due to a Poisson process of given parameter λ, called total demand rate. One of the sources is activated next to each demand, never active before. Just time hopping is considered, for simplicity and also because of the actual tasks kept in mind by the author. A message of νkm symbols, sent next to each demand, is taking values in $GF(q)$. $n = q - 1$. Each of the consecutive m subblocks of each source block of length k are encoded by means of m distinct (n, k) Reed-Solomon component codes over $GF(q)$ of the same kind. Along each frame superslots are defined consecutively, each of $m+\mu$ slots. (Superslot duration is defined as time unity. The last μ slots of each superslot are kept dummy.) The same binary hop sequence s_0 of length N, of weight n, of complete cyclic order, and of cyclic correlation $c = 1$ is assigned to each potential source. Multiple access erasure channel is assumed with neither noise nor delay.

Definition 1 *Consider a register step t at which match is declared. There is frame front coincidence at t provided frame fronts from at least two distinct sources occur at t within the correlator window, within supeslot distance (mod N) (from the rear end of the window (mod N)).*

III. Results

Choose

$$A = A' := n - k + 1$$

as activity threshold. Consider any correlator step t at which match is declared. Denote by k' the value of the subblock length k (associated with each component code) at which $R_{sum} := (1 + \frac{\mu}{m})^{-1} \frac{A'\kappa}{N}$ takes it largest possible value, given n, m, N, and A'. (Obviously $\kappa = km$.) Denote by N' the shortest possible hop sequence (frame) length at which decoding is error free, at any t, with match but with neither frame front coincidence nor overflow (with respect to activity threshold A').

Denote by A_0 the largest possible zero error activity threshold, given n, $k = k'$. Let $N = N'$.

Lemma 1

$$A_0 = k',$$

given any N, n, and $k = k'$.

Call peak factor the ratio $(1 + \delta)$ of A_0 to $\lambda 2N'$. Confine the study to $0 < \delta$. Denote, at any instant t, by $\mathcal{C}_t$ the configuration of all frame fronts that are just window active at t; and by

$$\mathbf{P}(dec\ err),$$

the decoding error probability at any t with match, but with neither frame front coincidence nor overflow with respect to $A = A_0 = k'$. Refer to

$$\mathbf{P}(dec\ err)' := \mathbf{P}(dec\ err)$$

at any register step t at which $\mathcal{C}_t$ equals one of the worst possible configurations (in the sense that the number of erasures along the considered codeword is the possible largest).

Theorem 1 *Assume the considered model for truly (unslotted) asynchronous single sequence hopping (with $0 < \delta$, block length $k' \geq 2$, the number of frames $\nu \geq 3$ next to each demand, and $k' \geq 10$. Then*

$$(1 - g_1)\frac{1}{4(1+\delta)(1+\frac{1}{\nu})}\frac{1}{k'} \leq \mathbf{P}(dec\ err)'$$

$$\leq (1 + g_2)(1 + g_3)\frac{1}{e(1+\delta)(1+\frac{1}{\nu})}\frac{1}{k'}.$$

(Expressions of $g_l (l = 1, 2, 3)$ are precisely given in the companion preprint. g_l $(l = 1, 2, 3)$ exceed 1, tend to 1 as $\kappa \to \infty$, and are close to 1 for usual values of k'. Recollect that the source blocklength of κ' q-ary symbols equals mk'.)

References

[1] S. Csibi, "Two sided bounds on the decoding error probability for structured hopping, a single common sequence and Poisson population," *Proc. 1994 IEEE Internat. Symp. Inform Theory*, Trondheim, Norway, June 27 – July 1, p. 290, 1994.

[2] J.L Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans on Inform. Theory*, vol. IT-31, pp. 192-194, 1985.

On the Performance of Partial-Response DS/SS Systems in a Specular Multipath Environment

Yi-Pin Wang, Wayne E. Stark[1]
The University of Michigan
Ann Arbor, Michigan 48105

Abstract — **Direct sequence spread spectrum systems using partial-response signals in a specular multipath environment are investigated. Instead of using the conventional precoder-decoder combination for non-spread partial-response signals, a RAKE receiver is employed to take advantage of the resolvability provided by wide-band DS/SS signals and the inherent diversity of partial-response signals. The performance measure of interest is signal-to-interference ratio (SIR). Our results suggest partial-response signals perform well in an outdoor mobile DS/SS system with high chip rate. The technique developed in this paper can be extended to any type of partial-response signals.**

I. Introduction and System Model

Partial-response signals have been widely used in many non-spread communication and magnetic recording systems because they allow transmission at the Nyquist rate by introducing known interference. Since the interference is known, it can be removed by certain processing. In addition, partial-response signals confine all the signal power to the main lobe. This feature makes filter design very straightforward when out-of-band power emission has to be strictly limited. Among many variations of partial-response signals, class I and class IV signals are most widely used because of their spectral shapes and simpler decoding operations at the receiver.

Two types of decoding algorithms are normally used for partial response systems: symbol-by-symbol decoding and maximum likelihood sequence detection (PRML) PRML performs better than symbol-by-symbol decoding. However, the performance of PRML depends on the size of the decoder memory and the decoder complexity is proportional to the size of the memory. Several technical difficulties are usually associated with conventional partial response systems. First the receiver has to estimate the power level of the received signal even when binary signaling is used. The inaccuracy of the power level estimate of the received signal degrades the noise immunity of the decoder. The degradation could be significant when a large signal set is used. Moreover, channel distortion or other types of interference requires the receiver to employ an equalizer.

In this paper, a direct-sequence spread-spectrum (DS/SS) system using class-I partial-response (PR-I) and class-IV partial-response (PR-IV) signals is considered. The self-interference introduced by partial-response signaling is treated as a form of multipath interference with known delays and amplitudes, and a RAKE receiver is used to take advantage of the known multipath interference. The main advantage of using a RAKE receiver instead of a conventional precoder-decoder combination for a system using partial-response signaling is the reduction of the complexity of the decoder. It was mentioned previously that conventional precoder-decoder structure needs to estimate the power level of the received signal, to equalize the channel distortion and multipath interference, and to use a sequential decoder to maximize the performance. However for a binary partial-response DS/SS system with a RAKE receiver, the decoder does not need the information about the power level of the received signal and can sustain the channel distortion and multipath interference to a certain degree without having to equalize the channel. Moreover, symbol-by-symbol detection should perform fairly well for such a receiver. The transmitter, channel and receiver model were all detailed in [1].

II. Numerical Results and Conclusions

We calculated the SIR's for a DS/SS system using PR-I, PR-IV, and filtered rectangular chip waveforms. Here, for fair comparisons, the filtered rectangular chip is a unit amplitude pulse low-pass filtered by an ideal brickwall filter with cut-off frequency equal to one half of the chip rate to produce a DS/SS signal of the same bandwidth as its partial-response counterparts. Our results show that when random spreading sequences are used filtered rectangular chips outperform PR-I and PR-IV by about 0.5–1 dB. On the other hand, with m-sequences and differential delays longer than 3 chips duration but no longer than $N-3$ chips duration, where N is the number of chips per bit, partial-response signals perform better than filtered rectangular in certain cases. This suggests that partial-response signals may be attractive in an outdoor mobile radio environment with high chip rate. Another feature of partial-response signals is they can be designed to match to the frequency response of the channel or the frequency band allocations. It may be possible to optimize the system to minimize the self-interference.

The RAKE receiver structure presented in this paper also makes the decoding process for other types of partial-response signaling straightforward, whereas for the conventional precoder-decoder structure the decoding process becomes cumbersome when the number of controlled interference terms of partial-response signaling is greater than two.

A potential disadvantage of partial-response signals is that they do not have uniform amplitude, and hence may suffer from non-linear amplification. Therefore, the transmitter power amplifier must operate in the linear range.

References

[1] Yi-Pin Wang and Wayne E. Stark, "Performance of a DS/SS System Using PR-I and PR-IV Chip Waveforms in a Specular Multipath Environment," *Submitted to IEEE Transactions on Vehicular Technology*, 1995.

[1] This paper was partially supported by the National Science Foundation under Grant NCR-9115969 and Intra Tech Systems Inc.

Distributed Access Control in Wireline and Wireless Systems

Christopher J. Hansen and Gregory J. Pottie

Department of Electrical Engineering
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024

Abstract - Practical frequency hopped spread spectrum (FHSS) wireless networks and multitone modulated wireline systems can be modeled as sets of interference channels. In these systems, it is desirable to optimize a cost function over the network that includes the transmission rates, blocking probability, and dropping probability for users. This optimization can be approximated using distributed algorithms that do not require explicit communication between pairs of users. We present one such algorithm that is designed to quickly identify a suboptimal, but reasonable solution. The performance of this algorithm is evaluated with simulations of prototype wireline and wireless systems.

I. Introduction

Both frequency hopped wireless networks [3] and multitone modulated wireline networks [1,2] can be modeled by a gain matrix plus additive white gaussian noise. In the wireline case, user pairs sharing a twisted pair cable interfere with each other through near end cross talk (NEXT) and far end cross talk (FEXT). Likewise, in a cell based wireless system, communication pairs formed between base stations and users interfere because of the shared radio channel. We assume here that the base station receiver decodes the received signals from different users independently, as is done in practice. For both systems, the resulting channel model is a set of interference channels.

Current digital wireless systems (IS-54 TDMA) and wireline systems (discrete multitone ADSL) use fixed reuse patterns to guarantee a minimal signal to interference ratio (SIR) and a minimal level of service. These reuse patterns take the form of cellular frequency planning in wireless systems, and fixed transmitter power levels in wireline systems. Since reuse patterns are designed for the worst case (cell boundaries for wireless and 1% worst case interferers for wireline) they are inherently inefficient. Capacity improvements can be made in both systems by adapting to the actual interference and avoiding the worst case situation when two users with high interference levels share the same channel.

Recent work has shown that high capacity can be achieved in wireless systems with frequency hopping over orthogonal hopping patterns [4]. With orthogonal hopping patterns, a user sees a different set of independent interferers in each hop. Power and bits can be allocated to those hops where interference is relatively low and a high SIR can be maintained for all users on the channel. A similar situation exists for the multitone wireline system, except the K channels are accessed in parallel. Certain pairs of users will interfere strongly. We choose to allocate power among users and channels to avoid this situation.

II. Cost Function Optimization

Assuming the interfering signals are independent and Gaussian, the aggregate bit rate over K channels is the average of the achievable bits rates over the channel set. For the wireline case, the transmission rate is increased by a factor of K because the channels are accessed in parallel. The relevant measure of the system performance is a cost function with call blocking, dropping, and system capacity as arguments. The goal is to optimize the cost over all users in the network so that the maximum revenue for network operation can be maintained, subject to specific service constraints.

Since the SIRs of the users in the network are all interconnected by the transmit powers, the optimal choice of the feasible set is difficult. Any algorithm for optimizing the cost function must operate jointly over all the users. Our approach is to choose a suboptimal solution based on admission control. New users rapidly probe all K channels to determine which can be used without excessive interference to previously active users. The algorithm is modeled after [5] but extended to handle multiple constellation sizes and periodic adaptation of active users. Active users make an effort to accommodate new users, but only if doing so will allow them to maintain the transmission rate they achieved when entering the network. The balance between blocking and dropping probabilities is controlled by the aggressiveness of new users and the ability of active users to block new users when necessary. After admission, an active user will use a distributed power control algorithm [6] to maintain the SIRs on all allocated channels.

III. Conclusions

At a fundamental level, multitone digital subscriber lines and frequency hopped wireless networks have formally similar channels and network cost functions. Optimizing the capacity and utility of these systems can be achieved by algorithms that operate in a distributed fashion, using only the knowledge of one's own channel characteristics and the interference from other users. The differences between wireline and wireless systems lie in the magnitude of the interference between user pairs and the associated costs of blocking and dropping. When these factors are incorporated into the adaptation algorithms, good performance can be achieved with both systems.

Acknowledgments

This work was supported by grants from ARPA/ETSO, University of California MICRO, PairGain Technologies, and Rockwell.

References

[1] Kalet, I. "The Multitone Channel", IEEE Transactions on Communications, Vol. 37, No. 2, February 1989, pp119-124.

[2] Chow, J.S., et al, "A Discrete Multitone Transceiver System for HDSL Applications", IEEE Journal on Selected Areas in Communications, Vol. 9, No. 6, August 1991, pp 895-908.

[3] G.J. Pottie and A.R. Calderbank, "Channel coding strategies for cellular radio," submitted to IEEE Trans. Vehic. Tech.

[4] C.C. Wang and G.J. Pottie, "Dynamic channel resource allocation in frequency hopped wireless systems," 1994 International Symposium on Information Theory.

[5] C.J. Hansen, C.C. Wang, and G.J. Pottie, "Distributed Dynamic Channel Resource Allocation in Wireless Communication Systems," Proceedings of the 1994 Asilomar Conference on Signals, Systems, and Computers.

[6] S. Chen, N. Bambos, and G. Pottie, "Admission Control Schemes for Wireless Communication Networks with Adjustable Transmit Powers", Proc. IEEE Infocom '94, pp 21-8, vol 1.

A Bayes coding algorithm for FSM sources

Toshiyasu MATSUSHIMA[1] and Shigeichi HIRASAWA
School of Science and Engineering, Waseda University 3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169 JAPAN

I. Introduction

The optimal universal code for FSMX sources[1] with respect to Bayes redundancy criterion[2] is deduced under the condition that the model, the probabilistic parameters and the initial state are unknown. The algorithm is not only Bayes optimal for FSMX sources but also asymptotically optimal for a stationary ergodic sources. Moreover the algorithm is regarded as a generalization of the Ziv-Lempel algorithm. In the basic CTW algorithm, the algorithm needs the initial context $x_{1-d}x_{2-d}\cdots x_0$, where a finite constant d is the depth of the context tree, for calculating the coding probability of x_1. For the problems of the initial situation and the infinite depth tree, the extensions to the CTW algorithm have been proposed in [3]. The optimal algorithm proposed in this paper gives a solution against these problems from another new point of view.

II. The probability of a sequence from a FSMX source

If arithmetic coding is used for universal coding, the main problem is deciding coding probability $P_C(x^n)$ or $P_C(x_t|x^{t-1})$ which is the probability assumed to code a source sequence $x^n : x_1x_2\cdots x_n$ where $x_i \in A$. Let m be an FSMX source model. The state set of m is represented by a l-ary complete tree $T(m)$ called a context tree. Let $S(m)$ be the set of all states in m. $S(m)$ corresponds to the set of all leaf nodes in $T(m)$. The state of a model m at t is determined by the postfix of a source sequence x^t. This mapping from x^t to a state $s \in S(m)$ is denoted by $f_m(x^t)$. The node corresponding to a postfix x_{t-j}^t is denoted by $s(x_{t-j}^t)$. All interior nodes of a tree $T(m)$ is denoted by $S^I(m)$.

For efficiency of the calculation of Bayes coding, we introduce a parametric representation for the probability of FSMX sources. Let $\theta(m)$ be a transition probability $\{P(x|s)|x \in A, s \in S(m)\}$. Moreover, the initial transition probability $\theta^I(m) = \{P^I(x|s)|x \in A, s \in S^I(m)\}$ is introduced. The probability of a sequence x^t is represented by

$$P(x^t|\theta(m),\theta^I(m),m) = P^I(x_1|\lambda)\cdots P^I(x_J|s(x_1^{J-1}))\prod_{i=J}^{t-1}P(x_{i+1}|f_m(x_i)), \quad (1)$$

where $J = \arg\min_j\{s(x_0^j)|s(x_0^j) \in S(m)\}$.

III. A recursive calculation of the coding probability

The Bayes optimal redundancy code for hierarchy source models such as FSMX models given an initial state was presented in our previous paper. In the case that the initial condition is unknown, the Bayes code of the FSMX models represented by Formula (1) is given in this section. The recursion formulas of the adaptive coding probability of the code are induced by using special classes of the prior : $q(s)$, $P(\theta(s)|s)$[4] and $P(\theta(s)^I|s)$.

$P^I(x_t|x^{t-1},s)$ and $P^S(x_t|x^{t-1},s)$ are defined as follows:

$$P^I(x_t|x^{t-1},s) = \int\cdots\int P(x_t|x^{t-1},\theta^I(s),s)P(\theta^I(s)|x^{t-1},s)d\theta^I(s), \quad (2)$$

$$P^S(x_t|x^{t-1},s) = \int\cdots\int P(x_t|x^{t-1},\theta(s),s)P(\theta(s)|x^{t-1},s)d\theta(s), \quad (3)$$

where $P(\theta^I(s)|x^{t-1},s)$ is the posterior probability of $\theta^I(s)$ given (x^{t-1},s).

Theorem 1 *Let $q(s|x^{t-1})$ be the posterior probability of $q(s)$ given x^{t-1}. The adaptive coding probability of Bayes code with respect to Formula (1) is given by the following recursion formula:*

$$P_C(x_t|x^{t-1}) = q(x_t|x^{t-1},s_\lambda), \quad (4)$$

$$q(x_t|x^{t-1},s) = \begin{cases}(*1) & \text{if } s = s(x_1^{t-1})\\ (*2) & \text{otherwise,}\end{cases} \quad (5)$$

$(*1) = q(s|x^{t-1})P^S(x_t|x^{t-1},s) + (1-q(s|x^{t-1}))P^I(x_t|x^{t-1},s)$,

$(*2) = q(s|x^{t-1})P^S(x_t|x^{t-1},s) + (1-q(s|x^{t-1}))q(x_t|x^{t-1},s')$,

where s' is a child node of s, and $s', s \in \{s(x_{t-j}^{t-1})|j = 1,\ldots,t-1\}$.

IV. The proposed algorithm

Using Theorem 1, we propose a practical Bayes coding algorithms for FSMX sources. The context tree used in the algorithm, which is not always an l-ary complete tree, grows according as the length of the source sequence increases. The set of the paths from the root to the leaves in the context tree with respect to the sequence x^t contains all parsing blocks of x^t by the Z-L algorithm. This means that the FSMX sources implicitly assumed in the Z-L algorithm are included in the context trees in the proposed algorithm. Although the Z-L algorithm assumes a single FSMX source for parsing, our algorithm uses a mixture model with respect to the set of FSMX sources which includes the single FSMX source. The proposed algorithm is regarded as a generalization of the Z-L algorithm.

References

[1] J. Rissanen. Universal modeling and coding. *IEEE Trans. Inf. Theory*, 27(1):12–23, Jan 1981.

[2] L. D. Davison. Universal noiseless coding. *IEEE Trans. Inf. Theory*, 19(6):783–795, Nov 1973.

[3] F. M. J. Willems. Extensions to the context tree weighting method. In *Proc. Int. Symp. of Infromation Theory*, page 387, 1994.

[4] T. Matsushima, , and S. Hirasawa. A bayes coding algorithm using context tree. In *Proc. Int. Symp. of Infromation Theory*, page 386, 1994.

[1] E-mail: toshi@matsu.mgmt.waseda.ac.jp

A CTW Scheme for Some FSM Models

Joe Suzuki
Dept of Mathematics, Osaka University,
Toyonaka, Osaka 560, Japan

Abstract — **The presented paper addresses a modified version of the CTW (Context Tree Weighting) which deals with some FSM (Finite State Machine) models as well as the FSMX (FSM X) models at little expense of computing in encoding/decoding.**

The FSMX model is an FSM model $g \in G_D$ ($D \geq 0$: integer) in which each state s the data x_{t+1}, $t = 0, 1, \cdots, n-1$ ($n \geq 1$: integer), to be encoded depends on is expressed as the shortest sequence $x_{t-d+1}x_{t-d+2}\cdots x_t$ ($d \leq D$) such that no state $s \in S(g)$ is a postfix of any other state, where G_D is the set of the models whose depth d is at most D, and $S(g)$ is the set of the states for $g \in G_D$. In general, the length $l(x_1^n|x_{-\infty}^0)$ given $x_{-\infty}^0 \in X^\infty$ is expressed by $l(x_1^n|x_{-\infty}^0) = -\log\{\sum_{g\in G_D} W(g)\prod_{s\in S(g)} Q_s(x_1^n|x_{-\infty}^0)\}$, $x_1^n \in X^n$, where $W(g)$, $g \in G_D$, satisfys $\sum_{g\in G_D} W(g) \leq 1$ (model weighting technique). Then, for each model $g \in G_D$, the probability $Q_s(x_1^n|x_{-\infty}^0) = \prod \frac{n_t[x_{t+1}, s] + 1/2}{n_t[s] + \alpha/2}$ is assigned to each state $s \in S(g)$, where the product is taken over $t = 0, 1, \cdots, n-1$ such that the state at time instance $t+1$ is $s \in S(g)$, and $n_t[x_{t+1}, s]$ and $n_t[s]$ are respectively the occurrence of $x_{t+1} \in X$ given $s \in S(g)$ and that of $s \in S(g)$ in $t = 0, 1, \cdots n-1$.

The CTW gives length $l(x_1^n|x_{-\infty}^0) = -\log P^\lambda(x_1^n|x_{-D+1}^0)$, $x_1^n \in X^n$, by setting $x_{-D+1}^0 \in X^D$ and constants $0 \leq \beta_s \leq 1$ for $s \in \cup_{d=0}^{D-1} X^d$ ($\beta_s = 0$ for $s \in X^D$), and applying the following equation recursively:

$$P^s(x_1^n|x_{-D+1}^0) = \begin{cases} (1-\beta_s)Q_s(x_1^n|x_{-D+1}^0) \\ \quad +\beta_s \prod_{x\in X} P^{xs}(x_1^n|x_{-D+1}^0) & (0 \leq |s| < D) \\ Q_s(x_1^n|x_{-D+1}^0) & (|s| = D) \end{cases} \tag{1}$$

where $xs \in \cup_{1\leq d\leq D} X^d$ is the concatenation of $x \in X$ and $s \in \cup_{0\leq d\leq D-1} X^d$, and $D \geq 0$ is some constant. Then, $W(g)$, $g \in G_D$, are expressed as $W(g) = \prod_{s\in S(g)}(1-\beta_s)\prod_{t\in T(g)-S(g)}\beta_t$, where $T(g)$ is the set of any postfixes of $s \in S(g)$ including s itself. For example, $W(g)$, $g \in G_D$, are obtained for five models ($D = 2$) and $\beta_s = 1/2$, $s \in \{\lambda\} \cup X$, as depicted in Figure 1. Notice that just $O(Dn)$ computation and $O(\alpha^D)$ storage are needed for the encoding/decoding although the depth is bounded by the finite constant D [1].

In this paper, we remove the constraint in the CTW that the source should be an FSMX model [1]. Although the proposed scheme does not yet cover the general FSM models with bounded depth D, the upperbound of the individual redundancy coincides with that of the original scheme except the length of model $g \in G_D$ (Theorem 1). In addition, the computation complexity is shown to be $O(2^D n)$ (Theorem 2). Although the $O(2^D n)$ computation may seem to be enormous compared with that of the original scheme, $O(|G_D|n)$ computation is required to realize model weighting technique for general FSM models. The number of possible models which we deal with in this paper is proved to be $|G_D| = O(2^{\alpha^D})$ (Theorem 3).

The model class we deal with is such that each state which x_{t+1}, $t = 0, 1, \cdots, N-1$, depends on is expressed as an element in $(X \cup \{*\})^D$ rather than that in $\cup_{0\leq d\leq D} X^d$, where "*" refers to a don't care symbol meaning that the state does not depend on the value of the position. Then, Eq. (1) is replaced by the following recursive equation

$$P^s(x_1^n|x_{-D+1}^0) = \begin{cases} (1-\beta_s)P^{*s}(x_1^n|x_{-D+1}^0) \\ \quad +\beta_s \prod_{x\in X} P^{xs}(x_1^n|x_{-D+1}^0) & (0 \leq |s| < D) \\ Q_s(x_1^n|x_{-D+1}^0) & (|s| = D) \end{cases} \tag{2}$$

Then, $W(g)$, $g \in G_D$, are expressed as $W(g) = \prod_{r\in R(g)}(1-\beta_r)\prod_{t\notin R(g)}\beta_t$, where $R(g)$ is the set of $r \in \cup_{0\leq d\leq D-1}(X \cup \{*\})^d$ such that concatination $*r$ is a postfix of any state in model $g \in G_D$. For example, $W(g)$, $g \in G_D$, are obtained for six models ($D = 2$) and $\beta_s = 1/2$, $s \in \{\lambda\} \cup X$, depicted as in Figure 2. Note that any FSMX model can be expressed as a specific case where, once "*" is emitted by some node, the "*" does not stop until the leaf, thus such a model as Figure 2 (e) is excluded in the original scheme.

The procedure of the update at time instance $t+1$, $t = 0, 1, \cdots, n-1$, is summaized as follows: Replace some place of the α-nary sequence $x_{t-D+1}^t \in X^D$ with "*"'s to obtain 2^D $(\alpha+1)$-nary sequence of length D; update $Q_s(x_1^n|x_{-D+1}^0)$, $n_t[x_{t+1}, s]$, $x_{t+1} \in X$, and $n_t[s]$ for the 2^D states $s \in (X \cup \{*\})^D$ generated in step 1; and generate $P^s(x_1^n|x_{-D+1}^0)$ by recursively applying Eq. (2) to the updated $P^{xs}(x_1^n|x_{-D+1}^0)$, $x \in X \cup \{*\}$, until $P^\lambda(x_1^n|x_{-D+1}^0)$ is obtained.

References

[1] F.M.J. Willems, Y.M. Shtarkov and T.J. Tjalkens, "Context Tree Weighting : A Sequential Universal Source Coding Procedure for FSMX Sources," IEEE Int. Symp. on Inform. Theory, San Antonio, Texas, Jan. 17-22, 1993, p. 59.

[2] J. Suzuki, "On a Generalized Context Tree Weighting Scheme", the Fourth Benelux-Japan Workshop of Information Theory, page 11, Eindhoven, Neitherland, June 22-23, 1994, p. 11.

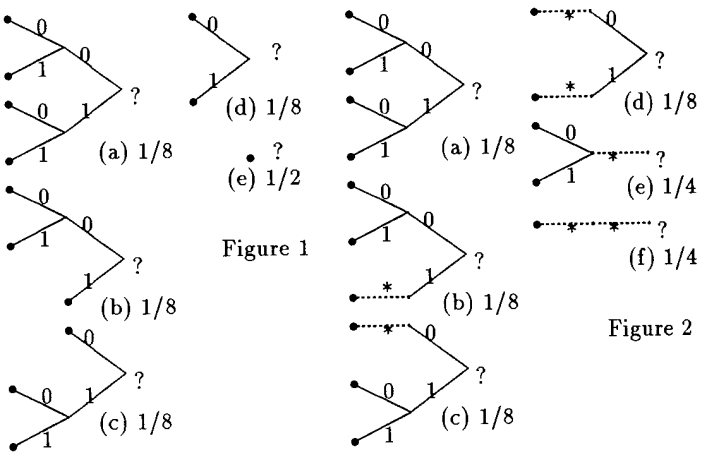


Figure 1

Figure 2

On Tree Sources, Finite State Machines, and Time Reversal

Gadiel Seroussi and Marcelo Weinberger
Hewlett-Packard Laboratories, Palo Alto, California, U.S.A.

Abstract — **We investigate the effect of time reversal on tree models of finite-memory processes. This is motivated in part by the following simple question that arises in some data compression applications: when trying to compress a data string using a universal source modeler, can it make a difference whether we read the string from left to right or from right to left? We characterize the class of finite-memory *two-sided tree processes*, whose time-reversed versions also admit tree models. Given a tree model, we present a construction of the tree model corresponding to the reverse process, and we show that the number of states in the reverse tree might be, in the extreme case, quadratic in the number of states of the original tree. This answers the above motivating question in the affirmative.**

I. Summary

Tree models[2] provide a reduced parametrization of finite-memory (Markov) sources, which can be efficiently and optimally modeled using Algorithm Context[1, 2], thus allowing a model size that is not necessarily exponential in the Markov order. In this work, we investigate the effects of *time reversal* on the structure of the minimal tree model of a finite-memory source. Time reversal of stationary Markov processes is well understood in the literature. In particular, it is known that time reversal preserves both the order and the entropy of a stationary Markov process (see, e.g., [3, Ch. 4]). This still leaves the question of the effect on the minimal tree parametrization open, and our interest in it stems from its implications (through *model cost*) on the rate of convergence to the entropy of a universal modeler.

Let A be an alphabet of α symbols, and let λ denote the empty string. For a string $u{=}u_1u_2\ldots u_k{\in}A^*$, let $\overline{u}{=}u_ku_{k-1}\ldots u_1$ denote the reverse of u. A *process* (or *information source*) over A is defined as a probability assignment $P : A^* \rightarrow [0,1]$ satisfying $P(\lambda) = 1$ and $P(u) = \sum_{a\in A} P(ua) \quad \forall u \in A^*$. Consider an arbitrary sequence $x^n = x_1x_2\cdots x_n$ over A. A process P has the *finite-memory property* (see, e.g. [2]) if the function $p(a|x^n) \triangleq P(x^na)/P(x^n)$ (a conditional probability by the properties of P) satisfies

$$p(\cdot|x^n) = p(\cdot|u\bar{s}(x^n)) \quad \forall u \in A^*, \tag{1}$$

where $s(x^n) = x_nx_{n-1}\cdots x_{n-\ell+1}$ for some ℓ, $0 \le \ell \le m$, not necessarily the same for all x^n (the case $\ell = 0$ is interpreted as defining the empty string). Such a string $s(x^n)$ is called a *state*. In a minimal representation of the model, $\bar{s}(x^n)$ is the shortest suffix of x^n (or *context*) satisfying (1). The set S of states defines a complete α-ary tree T, with the branches labeled by symbols of the alphabet, and S as the set of leaves. The pair $\mathcal{T} = <T, p(\cdot|\cdot)>$ is called a *tree model* for the process P, which is called *minimal* if for every node w in T such that all its successors wb are leaves, there exists $a, b, c \in A$ such that $p(a|wb) \neq p(a|wc)$. Conversely, we prove that given a tree model $\mathcal{T} = <T, p(\cdot|\cdot)>$, there exists one and only one *two-sided tree process* P modeled by $\mathcal{T}$, such that the *reverse assignment* $\overline{P}(u) \triangleq P(\overline{u})$ is also a finite memory process (called also the *reverse process* of P).

The reverse process $\overline{P}$ admits a minimal tree model $<\widehat{T}_p, \hat{p}(\cdot|\cdot)>$. The underlying tree $\widehat{T}_p$ in this model depends on both T and $p(\cdot|\cdot)$. In contrast, we define the *reverse tree* of T as $\overline{T} = \bigcup_{\text{all } p(\cdot|\cdot)} \widehat{T}_p$, which depends solely on T. The tree $\overline{T}$ is the minimal representation for the reverses of *all* the processes whose minimal tree models have T as underlying graph. We can also see $\overline{T}$ as the minimal tree of a reverse process $\overline{P}_z$, where P_z is a "symbolic process" with a minimal tree model $<T, p_z(\cdot|\cdot)>$ in which we have substituted $(\alpha{-}1)$ symbolic indeterminates $z_{a,s}$ for the free parameters $p(a|\bar{s})$ at each state s. Notice that, while there is a symmetry between T and $\widehat{T}_p$, so that $\widehat{(\widehat{T}_p)}_{\hat{p}} = T$, no such symmetry exists between T and $\overline{T}$, and we might have $\overline{\overline{T}} \neq T$. We present a combinatorial construction of $\overline{T}$, and use it as a tool to bound the size difference between T and $\widehat{T}_p$, noting that the latter is a subtree of $\overline{T}$. The construction and proofs rely on the characterization of tree models that have the *finite-state machine poperty*, i.e., whose leaves uniquely define a *next-state function*. Let $|T|_L$ denote the number of leaves of a complete α-ary tree T.

Theorem 1. (a) *Let T be such that $|T|_L{=}N$. Then,*

$$|\overline{T}|_L \;\le\; \frac{1}{2(\alpha-1)}N^2 + O(N).$$

(b) *For every $N{>}0$ such that $N \equiv 1 \bmod(\alpha{-}1)$, there exists a complete α-ary tree T with $|T|_L{=}N$, such that $|\overline{T}|_L$ attains the upper bound of part (a) up to an additive term $O(N)$.*

Corollary 1. *Let P be a process with minimal tree model $<T, p(\cdot|\cdot)>$, and let $N{=}|T|_L$. Then, the minimal tree model $<\widehat{T}_p, \hat{p}(\cdot|\cdot)>$ of $\overline{P}$ satisfies*

$$\sqrt{2(\alpha-1)N} - O(1) \;\le\; |\widehat{T}_p|_L \;\le\; \frac{1}{2(\alpha-1)}N^2 + O(N)\,.$$

It follows from Corollary 1 that, when using tree sources to model data, there might be significant differences between the size of the tree estimated when reading the data from left to right and the one estimated from right to left. These differences, in turn, affect the model cost incurred by the modeling algorithm. This behavior is a consequence of the choice of class of models targeted by the algorithm, since the number of free parameters determining the reverse process is identical to the number of parameters in the original process. On the other hand, it is this choice of model class that allows for an efficient estimation algorithm.

References

[1] J. Rissanen, "A universal data compression system," *IEEE Trans. Inform. Theory*, IT-29, pp. 656–664, Sep 1983.

[2] M. J. Weinberger, J. Rissanen, and M. Feder, "A universal finite memory source," *IEEE Trans. Inform. Theory*, IT-41, pp. 643–652, May 1995.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

Approximation of Bayes code for Markov sources

Jun-ichi Takeuchi† and Tsutomu Kawabata††

tak@SBL.CL.nec.co.jp kawabata@cas.uec.ac.jp

†C&C Res. Labs., NEC Corp., 4-1-1 Miyazaki, Miyamae-ku, Kawasaki, Kanagawa 216, Japan.

††Dept. of Communications & Systems, Univ. of Electro-Communications, 1-5-1 Choufugaoka, Chofu, Tokyo 182, Japan.

In this abstract, we give an approximation formula for the predictive Bayes code for the FSMX' models (subspaces of Markov models). Moreover, we empirically show that the code using our approximation formula with the Jeffreys prior employed gives shorter code length than the one using the Laplace estimator for the first order Markov models.

Let $\mathcal{A}$ be a set of m symbols. Suppose that $\mathcal{A}$ contains symbol ω and let $\mathcal{A}'$ denote $\mathcal{A}-\{\omega\}$. Let T be a subset of $\mathcal{A}^*$. When, for all $s \in T$, any postfix of s belongs to T (e.g., the postfixes of a_1a_2 are a_1a_2, a_2 and λ (the null sequence)), T is called a context tree. Define $\partial T \equiv \{as | a \in \mathcal{A}, s \in T\} \cup \{\lambda\} - T$. Each element of ∂T is called a leaf of T or a context. For any $a^i = a_1a_2,...,a_i$ ($i \geq d(T) \equiv \max_{s\in\partial T}|s|$), let $s(a^i)$ denote a postfix of a^i which belongs to ∂T. $s(a^i)$ is called the context of a^i defined by T. Now, we define the FSMX' source $p(\cdot|\eta, T)$ as $p(a^N|\eta,T) = p(a^{d(T)}|\eta,T)\prod_{i=d(T)+1}^{N-1}\eta^{a_{i+1}}_{s(a^i)}$, where η^a_s denotes in general the probability that a is produced at the context s (i.e. $\eta^a_s > 0$ and $\sum_{a\in\mathcal{A}}\eta^a_s = 1$ hold.) and $p(a^{d(T)}|\eta,T)$ denotes the initial probability determined by the stationary probabilities. Let η be the $(|\mathcal{A}|-1)\cdot|\partial T|$-dimensional vector whose components are η^a_s ($s \in \partial T$, $a \in \mathcal{A}'$) and $H(T)$ denote the range of η. We can write $p(a^N|\eta,T) = p(a^d|\eta,T)\prod_{s\in\partial T}\prod_{a\in\mathcal{A}}(\eta^a_s)^{n^a_s}$, where n^a_s denotes the number of times a is generated at the context s in the sequence $a_{d+1}...a_N$, and we let $n_a = \sum_{a\in\mathcal{A}} n^a_s$. An FSMX' model $M(T)$ is defined as $M(T) = \{p(\cdot|\eta,T)|\eta \in H(T)\}$. (When $\forall s \in \partial T$ $\forall a \in \mathcal{A}$ $sa \notin T$ holds, $M(T)$ is called an FSMX model.) By introducing the another parameter θ, $p(a^N|\eta)$ can be rewritten as follows.

$$p(a^N|\eta,T) = p(a^d|\eta,T)\prod_{s\in\partial T}\exp(n_s(\sum_{a\in\mathcal{A}'}(\theta^a_s)\hat{\eta}^a_s - \psi(\theta_s))), \quad (1)$$

where we let $\theta^a_s = \ln(\eta^a_s/\eta^\omega_s)$, $\hat{\eta}^a_s = n^a_s/n_s$, and $\psi(\theta_s) = -\ln\eta^\omega_s = \ln(1+\sum_{a\in\mathcal{A}'}\exp\theta^a_s)$. ('ln' denotes the natural logarithm.) We let $\Theta(T)$ denote the range of θ as η varies over $H(T)$. Note that a class of probability distributions written as $\exp(n_s(\sum_{a\in\mathcal{A}'}(\theta^a_s)\hat{\eta}^a_s - \psi(\theta_s)))$ is called an exponential family. We use $p(\cdot|\theta,T)$ as a short hand notation for $p(\cdot|\eta(\theta),T)$.

Next, we define the Bayes code for $M(T)$. We fix a context tree T and let $p(\cdot|\theta)$ denote $p(\cdot|\theta,T)$. We assume a prior $w(\theta)d\theta$ over $\Theta(T)$. Then, the predictive Bayes code with prior w is given by $p_w(a_{N+1}|a^N) \equiv \int p(a_{N+1}|a^N,\theta)w(\eta|a^N)d\theta = \int \eta^{a_{N+1}}_{s(a^N)}w(\eta|a^N)d\theta$, where $w(\theta|a^N)$ denotes the posterior density of θ. Now, we can state our main result.

Theorem 1 *Let $w(\theta)$ be the prior defined on the measure $d\theta$. Under a certain weak condition, for every $a \in \mathcal{A}'$,*

$$p_w(a|a^N) = \hat{\eta}^a_{sc} + \frac{1}{n_{sc}}\frac{\partial\ln(p(a^d|\hat{\theta})w(\hat{\theta}))}{\partial\theta^a_{sc}} + O(\frac{\sqrt{\ln N}}{n_{sc}\sqrt{N}}), \quad (2)$$

holds, where sc and $\hat{\theta}$ denote $s(a^N)$ and $\theta(\hat{\eta})$, respectively.

Remark: The key of the proof is expression (1). This extends the approximation formula for the Bayes code for any (i.i.d.) exponential family given in [3].

We let $\tilde{\eta}$ denote the first term plus the second term of (2). Then we can use $\tilde{\eta}^{a_{N+1}}_{s(a^N)}$ as an approximation formula for $p(a_{N+1}|a^N)$. Let w_J denote Jeffreys prior, which is defined as $w_J(\theta)d\theta = (\det J(\theta))^{1/2}/c$ (c is a normalization constant and $J(\theta)$ is the Fisher information matrix with respect to θ). We refer to p_{w_J} as the Jeffreys code. It is known that p_{w_J} for the i.i.d. case (i.e. $T = \emptyset$) is asymptotically minimax in terms of redundancy ([1]) and almost equals the Laplace estimator, which is used in CONTEXT[2] and CTW method[4].

Now, we compare our approximation formula for w_J with the Laplace estimator. Let $\mathcal{A} = \{0,1\}$ ($\omega = 0$) and $T = \{\lambda\}$. ($\partial T = \{0,1\}$) Suppose that $s(a_N) = 0$. By Theorem 1, the approximation of the Jeffreys code for this case is given by

$$\tilde{\eta}^1_0 = \hat{\eta}^1_0 + \frac{1}{n_0}\big(1 + a_1 - 1.5\hat{\eta}^1_0 - \frac{2\hat{\eta}^1_0(1-\hat{\eta}^1_0)}{\hat{\eta}^1_0 + \hat{\eta}^0_1}\big). \quad (3)$$

Note that the difference between $\tilde{\eta}^1_0$ and the Laplace estimator $(n^1_0 + 0.5)/(n_0 + 1)$ equals $\Omega(1/n_0)$.

We have compared the redundancy of the code using (3) with that of the one using the Laplace estimator (let p_L denote it) by a computer simulation. In general, the redundancy of code q is defined as $R_N(\theta,q) = E_\theta(-\log q(a^N) - (-\log p(a^N|\theta)))$, where $q(a^N)$ is the block probability given to a^N by q and E_θ denotes the expectation with respect to $p(\cdot|\theta)$. ('log' dnotes the logarithm to the base 2.) We have estimated the expectation with respect to $p(\cdot|\theta)$ by performing a large number of trials using pseudo random numbers. We show the result with $N = 50$ in Table 1. The number of trials is 1000. In each cell, the right hand sides and left hand sides denote $R_N(\theta,p_{w_J})$ and $R_N(\theta,p_L)$, respectively. The vertical and horizontal axis correspond to the values of η^1_0 and η^0_1 of the actual source respectively. We can see that $R_N(\theta,p_J) < R_N(\theta,p_L)$ holds for all cases. This seems to support our conjecture ([5]) that the Jeffreys code is minimax for FSMX' models as well.

	0.1	0.5	0.9
0.1 p_J / p_L	3.59 / 3.72	3.52 /3.72	3.71 / 4.01
0.5		3.44 /3.83	3.36 / 3.78
0.9			3.41 / 3.86

Table 1: Redundancy

Our approximation formula requires not only the n^a_{sc}'s but also the n^a_s's for all $s \in \partial T$. On the other hand, the Laplace estimator can be calculated based on the n^a_{sc}'s alone. Both CONTEXT and CTW methods make use of such property of the Laplace estimator. Hence, there is a difficulty in introducing our formula to CONTEXT or CTW.

References

[1] B. Clarke & A. Barron, "Jeffreys prior is asymptotically least favorable under entropy risk," *the JSPI*, 1994.

[2] J. Rissanen, "A universal data compression system," *IEEE Trans. IT*, Vol. 29, No. 5, pp.656-664, 1983.

[3] J. Takeuchi, "Characterization of the Bayes estimator and the MDL estimator for exponential families," *IEEE ISIT*, 1995.

[4] M. Willems et. al., "The context tree weighting method: basic properties." *IEEE trans. IT*, Vol. 41, No. 3, pp. 653-664, 1995.

[5] T. Kawabata, "Bayes codes and context tree weighting method," (in Japanese) *TR of IEICE*, IT93-121, pp. 7-12, 1994.

Markov Random Field Models for Natural Language

Kevin E. Mark, Michael I. Miller, Ulf Grenander
Department of Electrical Engineering
Washington University
St. Louis, Missouri 63130

I. Introduction

Markov chain (N-gram) source models for natural language were explored by Shannon and have found wide application in speech recognition systems. However, the underlying linear graph structure is inadequate to express the hierarchical structure of language necessary for encoding syntactic information. Context-free language models which generate tree graphs are a natural way of encoding this information, but lack the modeling of interword dependencies.

In this paper, we consider a hybrid tree/chain graph structure which has the advantage of incorporating lexical dependencies in syntactic representations. Two Markov random field probability measures are derived on these tree/chain graphs from the maximum entropy principle.

II. Stochastic Context-Free Grammars

A stochastic context-free grammar G is specified by the quintuple $< V_N, V_T, R, S, P >$ where V_N is a finite set of non-terminal symbols, V_T is a finite set of terminal symbols, R is a set of rewrite rules, S is a start symbol in V_N, and P is a parameter vector. If $r \in R$, then P_r is the probability of using the rewrite rule r.

An important measure is the probability of a derivation tree T. Using ideas from the random branching process literature [1, 4], we specify a derivation tree T by its depth L and the counting statistics $z_l(i,k), l = 1, \ldots, L, i = 1, \ldots, |V_N|$, and $k = 1, \ldots, |R|$. The counting statistic $z_l(i,k)$ is the number of non-terminals $\sigma_i \in V_N$ rewritten at level l with rule $r_k \in R$. With these statistics the probability of a tree T is given by

$$\pi(T) = \prod_{l=1}^{L} \prod_{i=1}^{|V_N|} \prod_{k=1}^{|R|} P_{r_k}{}^{z_{l-1}(i,k)}. \tag{1}$$

In this model, the probability of a word string $W_{1,N} = w_1 w_2 \ldots w_N$, $\beta(W_{1,N})$, is given by

$$\beta(W_{1,N}) = \sum_{T \in Parses(W_{1,N})} \pi(T) \tag{2}$$

where $Parses(W_{1,N})$ is the set of parse trees for the given word string. For an unambiguous grammar, $Parses(W_{1,N})$ consists of a single parse.

III. Markov Random Field Models

We now consider adding bigram relative frequencies as constraints on our stochastic context-free trees inducing linear constraints on the leaves of the CF tree. For a given word string $W_{1,N} = w_1 w_2 \cdots w_N$, the relative frequency of the word pair $v_i v_j$ is

$$\frac{C_{v_i v_j}(W_{1,N})}{N-1} = \frac{1}{N-1} \sum_{k=1}^{N-1} 1_{v_i v_j}(w_k, w_{k+1}) \tag{3}$$

where $v_i, v_j \in V_T$.

Theorem 1 [3]

The probability distribution on trees, $p(T)$, minimizing the relative entropy with respect to the distribution $\pi(T)$ defined by a stochastic context-free grammar,

$$\sum p(T) \log \frac{p(T)}{\pi(T)} \tag{4}$$

subject to the bigram constraints $\{E\left[\frac{C_{v_i v_j}(W_{1,N})}{N-1}\right] = H_{v_i v_j}\}_{v_i, v_j \in V_T}$ is

$$p(T) = \frac{1}{Z} \exp\left(\frac{1}{N-1} \sum_{v_1 \in V_T} \sum_{v_2 \in V_T} \alpha_{v_1 v_2} C_{v_1 v_2}(W_{1,N}) \right) \pi(T)$$

where Z is the normalizing constant and the $\alpha_{v_1 v_2}$ are the Lagrange multipliers chosen to satisfy the constraints.

This distribution is a Markov random field with the following neighborhood structure on the leaves:

$$p(w_i | T \backslash w_i) = p(w_i | w_{i-1}, w_{i+1}, \gamma_i) \tag{5}$$

where γ_i is the part-of-speech of w_i. Note that because of the added lexical neighbors, the distribution is no longer context-free.

A second, more computationally efficient model which retains the neighborhood structure of the MRF above is given by the distribution

$$p(T) = \frac{1}{K} \pi(T^*) p(w_1 | \gamma_1) \prod_{i=2}^{N} p(w_i | w_{i-1}, \gamma_i) \tag{6}$$

where T^* is a tree down to the preterminal, or part-of-speech, level. This model is interpreted as a SCF model generating a sequence of parts-of-speech with word attachment according to a non-stationary Markov chain.

References

[1] Harris, T. E., *The Theory of Branching Processes*, Springer-Verlag, Berlin, 1963.

[2] Mark, K. E., Miller, M. I., and Grenander, U., "Constrained Stochastic Language Models," to appear in *Image Models (and their Speech Model Cousins)*, ed. S. E. Levinson and L. Shepp.

[3] Mark, K. E., Miller, M. I., Grenander, U., and Abney, S., "Parameter Estimation for Constrained Context-Free Language Models," in *Proc. DARPA Speech and Natural Language Workshop*, Morgan Kauffman, New York, 1992.

[4] Miller, M. I., and O'Sullivan, J. A., "Entropies and Combinatorics of Random Branching Processes and Context-Free Languages," *IEEE Trans. on Information Theory*, March, 1992.

A Multialphabet Arithmetic Coding with Weighted History Model

Meng-Han Hsieh and Che-Ho Wei[1]

Dept. of Electronics Engineering and Center for Telecommunications Research
National Chiao Tung University, Hsinchu, Taiwan 30050, ROC

Abstract — A multialphabet arithmetic coding with weighted history model is presented for variable length coding of the video symbols in video compression applications.

I. Introduction

A limited past history model introduced by Ghanbari[1] uses a limited number of past symbols to estimate the probability distribution. This model takes relatively large buffer to achieve its optimal compression performance. Here we present a weighted history model that uses less buffer and obtain better performance.

II. Weighted History Model

Suppose there are p possible occurrences, and the alphabet used in arithmetic coding is defined as $S_1, \cdots, S_p$. The buffer size used in the limited past history model is M, and the occurrence of S_i in the buffer is represented by O_i for all index i lies between 1 and p. Adding all occurrence in the buffer thus obtains the buffer size, i.e., $O_1 + O_2 + O_3 + \cdots + O_p = M$, and the relative frequency of symbol S_i can be obtained by $freq(S_i) = \frac{O_i+1}{p+M}$. Therefore the corresponding cumulative frequency of symbol S_i is $cum_freq(S_i) = \sum_{k=1}^{i} freq(S_k)$.

The major disadvantages of the limited past history model is caused by the requirement that the occurrence of each symbol is at least one for arithmetic coding. The limited past history model overestimates the probability of each symbol by $\frac{1}{p+M}$, and the total overhead probability is equal to $\frac{p}{p+M}$. When the buffer size M is small, the overhead probability is almost one. That is, the probability distribution obtained by the limited past history buffer is nearly invariant to occurrence in the history buffer, and the statistical property of the source data is not reflected by this model.

To enforce the relations between the probability distribution and the occurrence in the buffer, we can simply induce a weight to the buffer. Therefore, the frequency of the ith symbol is $freq(S_i) = \frac{O_i \cdot W+1}{p+M \cdot W}$. The total overhead probability of the weighted history model is $\frac{p}{p+M \cdot W}$, which is much smaller than that of the limited past history model, especially when the buffer size is small.

The weighted history model uses less buffer than the limited past history model does. Consequently, the weighted history model has a faster adaptation and the local redundancy can be exploited more. The performance of the arithmetic coding with weighted history model for various buffer sizes and various weights was investigated. Fig. 1 uses a coded data of the pyramid VQ[2] as the source data. Five different weights of the weighted history model with various buffer sizes are shown. From this figure, it can be seen that the weighted history model really outperforms the limited past history model, especially when the buffer size is small. The large weight will reduce the probability of the symbols that are not in current buffer. If the next symbol is not in the history buffer, a long codeword will be assigned to represent this symbol because of low probability. From our experiments, an appropriate weight for the weighted history model is in the range from 16 to 128.

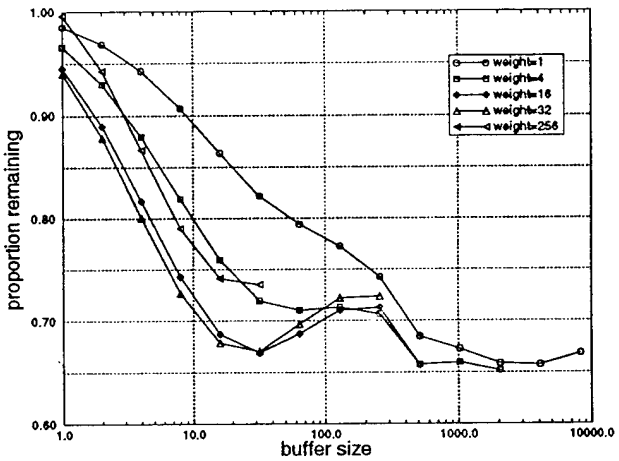


Fig. 1: Performance of arithmetic coding with weighted history model. Data source: coded data form pyramid VQ.

III. Hardware Implementation

The weighted history model uses a smaller history buffer to model the cumulative density function of the arithmetic coder, and uses smaller counters to record the cumulative frequencies than the limited past history model. This is because each occurrence in the history buffer is multiplied by a weight, thus all bits below the weight are not changed if the weight is an integral power of 2. Because of smaller buffer size and counters, the weighted history model is well suited for hardware implementation in conjunction with the multiplication-free multialphabet arithmetic coder proposed in [3].

IV. Conclusion

A weighted history model can solve the disadvantages of the limited past history model. The performance of the weighted history model is better than the limited past history model, and the history buffer used in the weighted history model is smaller. From the experiments, it can be seen that the arithmetic coding with weighted history model is good for image coding.

References

[1] M. Ghanbari, "Arithmetic coding with limited past history," *Elec. Letters*, vol. 27, no. 13, pp. 1157-1159, Jun. 1991.

[2] A. Gersho, R. M. Gray, *Vector quantization and signal compression*, Massachusetts: Kluwer Academic Publishers, 1992.

[3] J. Rissanen and K. M. Mohiuddin, "A multiplication-free multialphabet arithmetic code," *IEEE Trans. on Commun.*, vol 37, no. 2, pp. 93–98, Feb. 1989.

[1]This work was supported by National Science Council, ROC under the contract NSC82-0404-E009-338

Bit-Wise Arithmetic Coding for Data Compression

Aaron B. Kiely[1]
Jet Propulsion Laboratory, MS 238-420, 4800 Oak Grove Drive, Pasadena, California, 91109, USA

Abstract — **Consider the problem of compressing a uniformly quantized IID source. A traditional approach is to assign variable length codewords to the quantizer output symbols or groups of symbols (e.g., Huffman coding). Here we propose an alternative solution: assign a *fixed length* binary codeword to each output symbol in such a way that a zero is more likely than a one in every codeword bit position. This redundancy is then exploited using a block-adaptive binary arithmetic encoder to compress the data. This technique is simple, has low overhead, and can be used as a progressive transmission system.**

I. Encoding Procedure

A continuous source with probability density $f(x)$ is quantized by a uniform quantizer whose output symbols are mapped to b bit codewords. The first codeword bit indicates the sign of the quantizer reconstruction point. Each successive bit gives a further level of resolution and is assigned so that zeros are more concentrated near the origin. Figure 1 illustrates this mapping for $b = 4$.

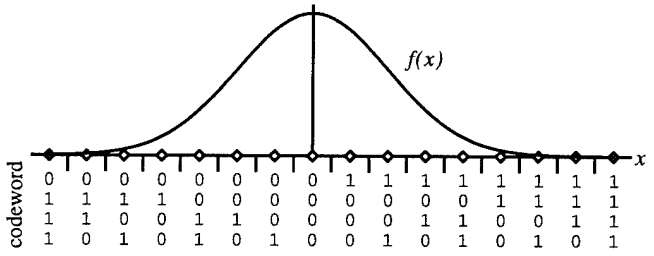


Fig. 1: Example of a pdf and codeword assignment for a four bit uniform quantizer.

We assume that $f(x)$ is symmetric about $x = 0$ and nonincreasing with $|x|$ so that the probability is more concentrated near the origin. Such sources are not uncommon in practice. Because of this assumption, the codeword assignment ensures that a zero will be more likely than a one in every bit position.

Codewords corresponding to N adjacent source samples are grouped together. The N sign bits of the codeword sequence are encoded using a block-adaptive binary arithmetic encoder. Then the N next most significant bits are encoded, and so on. Each bit sequence is encoded independently– at the ith stage the arithmetic coder estimates the *unconditional* probability that the ith codeword bit is a zero. This can be viewed as a simple progressive transmission system– each subsequent codeword bit gives a further level of detail about the source.

The obvious loss is that we lose the benefit of inter-bit dependency. E.g., the probability that the second bit is a zero is not in general independent of the value of the first bit, though the encoding procedure acts as if it were. However, for many sources (e.g., Gaussian and Laplacian), this loss is small, and this technique often has lower redundancy than Huffman coding, because the arithmetic coder is not required to produce an output symbol for every input symbol.

The independent treatment of the codeword bits provides some benefits. The overhead required increases linearly in b. By contrast, because the number of codewords is 2^b, the overhead of block-adaptive Huffman coding increases exponentially in b unless we are able to cleverly exploit additional information about the source [2].

II. Arithmetic Encoder Operation

A binary arithmetic encoder has a single parameter P, the anticipated probability of a zero. We encode an N-length sequence of bits block-adaptively, i.e., the encoder output sequence is preceded by overhead bits that identify to the decoder the value of P being used. By using $\log_2 N$ bits of overhead, we could specify the exact frequency of zeros in the sequence, but by using fewer bits we can exchange accuracy for lower overhead. If m overhead bits are used, we can select 2^m probabilities $\{\rho_1, \rho_2, \ldots \rho_{2^m}\}$ that can be used as values for P. This amounts to using line segments to approximate the binary entropy function [1].

Omitting the remaining details, we find that for large N, to minimize the maximum redundancy (including overhead), the probability values are

$$\rho_i \approx \frac{1}{2}\left[1 - \sin\left(\frac{\pi}{2^{m+1}}[1 + 2^m - 2i]\right)\right]$$

and the optimal number of overhead bits m is approximately

$$m \approx \frac{1}{2}\log_2 N + \log_2 \pi - 1.$$

The encoder counts the number of zeros in the input sequence to determines the probability index i. We transmit m bits to identify i, followed by the arithmetic encoder output sequence. The encoder and decoder both use parameter $P = \rho_i$.

III. Performance

The rate R of the bit-wise arithmetic coder is approximately

$$R \approx H(Q) + \mathcal{R} + \frac{b}{N}\left[\frac{1}{2\ln 2} + \log_2 \pi - \frac{1}{2} + \frac{1}{2}\log_2 N\right]$$

here $H(Q)$ is the entropy of the quantized source and $\mathcal{R}$ is the redundancy due to independent treatment of the codeword bits.

References

[1] A. B. Kiely, "Bit-Wise Arithmetic Coding for Data Compression," *TDA Progress Reports* 42-117, pp. 145-160, January-March 1994.

[2] R. J. McEliece and T. H. Palmatier, "Estimating the Size of Huffman Code Preambles," *TDA Progress Reports* 42-114, April-June 1993, pp. 90-95, August 15, 1993.

[1]The research described in this paper was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Fast Enumerative Source Coding

Boris Ryabko
Novosibirsk Telecommunication Institute, Kirov st. 86, 630102, Russia

Abstract — **The problem of enumerative coding was considered in [1] for the first time. By coding words of a length n the method from [1] has an encoding and decoding speed which equals to $0(n)$ when $n \to \infty$. We propose a code which has the high speed: $0(\log^2 n \log\log n), n \to \infty$. This code is close to author's method from [2].**

I. Introduction and the Main Idea

The problem of enumerative coding is well known in Information Theory and widely applied to retrieval problems and combinatorial analysis [1]. The suggested fast code uses the method from [2]. The simplest but important example of enumerative coding is the problem of translation numbers from one number system to another. We use this example for the description of the main idea of the proposed method . Let we have to translate the number $x_1...x_n$ from the m-system ($m > 2$) to the binary system. A "common" method is based on well - known Horner scheme :

$$(1)\quad code(x_1x_2...x_n) = (...(x_1m + x_2)m + x_3)...m + x_n$$

When we calculate in the binary - system, we obtain the value $x_1...x_n$ in the binary - system . We shall assess the calculation time by the number of operations on single - bit words. We use the Schonhager- Strassen method of multiplication and division of numbers.For this method the time of multiplication of two numbers with L digits each, is equal to $O(L \log L \log\log L), L \to \infty$. [3]. It is easy to see that the time for calculation by (1) is not less than $cn^2, c > 0, n \to \infty$. Hence, the speed is not less than cn. We suggest computing by the scheme

$$(2)\quad code(x_1...x_n) = ((...((x_1m + x_2)(mm) + (x_3m + x_4)))$$
$$((mm)(mm)) + (x_5m + x_6)(mm) + (x_7m + x_8)...)$$

In this case the main part of multiplications will be implemented on comparatively small numbers and when (2) is used, the time for computing is equal to $O(n \log^2 n \log\log n), n \to \infty$ and the encoding speed is equal to $O(log^2 n \log\log n)$... We can see that "proper" arrangement of brackets allows to decrease the calculation time essentially. It is worthy of noting that the described method is known as "divide and conquer" principle [3].

II. Main Result

We use definitions from [1]. Let $A = \{0, 1, ..., m - 1\}$ be an alphabet of m letters,$m \geq 2$, A^n be the set of all words of length n over the alphabet A. Let an arbitrary $S \subset A^n$ be a source. Let's give the lexicographic order to words S, and for the integer $1 \leq k \leq n$ and for the word $x_1...x_k \in A_k$, denote by $N_s(x_1...x_k)$ the quantity of words produced by S and having the prefix $x_1...x_k$ In [1] the code by formula

$$(3)\quad code(x_1...x_n) = \sum_{i=1}^{n} \sum_{a<x_i} N_s(x_1...x_{i-1}a)$$

was proposed. Let's define for $x_1...x_n \in S$.

$$(4)\quad P(x_1) = N_s(x_1)/|S|, \qquad P(x_k/x_1...x_{k-1}) =$$
$$N_s(x_1...x_k)/N_s(x_1...x_{k-1}), k = 2, ..., n$$

$$(5)\quad q(x_k/x_1...x_{k-1}) = \sum_{a<x_k} P(a/x_1...x_{k-1}), k = 1, ..., n$$

From (3), (4), (5) it is easy to obtain

$$(6)\quad code(x_1...x_n) = |S|(q(x_1) + q(x_2/x_1)P(x_1)+$$
$$q(x_3/x_1x_2)P(x_1)P(x_2/x_1) + ...)$$

The scheme of the proposed method is following : Each $P(x_k/x_1...x_{k-1})$, $q(x_k/x_1...x_{k-1}), x_1...x_k \in A^k$ can be written in the form of a word with $2\log n + 0(1)$ digits. Then (6) resulted in the form

$$code(x_1...x_n) = |S|((q(x_1) + q(x_2/x_1)P(x_1))+$$
$$(P(x_1)P(x_2/x_1))(q(x_3/x_1x_2) + q(x_4/...)P(x_3/...))+$$
$$((P(x_1)P(x_2/x_1)(P(x_3/...)P(x_4/...))(q(x_5/...)+$$
$$q(x_6/...)P(x_5/...)...)$$

(Here we used "proper" arrangement of brackets , as if we go over to (2) from (1)). Decoding is constructed similarly, by using division. It is easy to calculate that the encoding and decoding speed is equal to $O(\log^3 n \log\log n)$ when $n \to \infty$.

References

[1] T.M.Cover, "Enumerative source encoding", *IEEE Trans.Inf.Theory,* vol. 19, n 1, pp. 73-77, 1973.

[2] B.Ya.Ryabko, "Fast and effective coding of information sources", *IEEE Trans. Inf. Theory,* vol. 30, n 1, pp.96-99, 1994.

[3] A.V.Aho, J.E.Hopcroft, J.D.Ullman, "The design and analysis of computer algorithms", *Addison-Wesley,* 1976.

A New Spectral Shaping Scheme Without Subcarriers

Yongwen Yang & Lloyd R. Welch
Panasonic Technologies, Inc./CSTL & EE Dept, Univ. of Southern California

Abstract — **This study is concerned with the selection of waveform structure and message redundancy for reliable reception in a noisy channel and for not interfering with a secondary use of the signal such as establishing frequency and phase synchrony. We will propose and analyze the Bit Reversal(BR) encoding scheme of inserting redundancy to minimize the spectral energy near zero frequency. The primary mathematical tool for this analysis will be Markov chains with finite number of states and constant transition probabilities.**

I. Introduction

The analysis of power spectrum density (PSD) of synchronous baseband digital signals plays a fundamental important role in the design of communication and signal processing systems. From [1,2,3], we see that the PSD is characterized by modulator design (this factor determines the structure of the model and hence the transition matrix) and signal design (chooses a set of waveforms used). In this paper, we will propose and analyze a spectral shaping scheme, the Bit Reversal(BR) encoding scheme that increases the data bandwidth by a very small fraction, yet reduces the spectral energy near D.C. to nearly zero. The primary mathematical tool for this analysis, like most digital signal format, will be Markov chains with finite number of states and constant transition probabilities.

II. The Proposed BR Encoding Scheme

The idea of the BR encoding model is to attempt to balance the number of +1's and -1's in the transmitted stream by inserting a redundant bit every $L-$th bit. This bit indicates whether the L-block is transmitted directly or sign reversed before transmission. The decision is based on the excess of +1's or -1's in the message of the $t-$th L-block versus the excess in the transmitted stream from time zero.

More precisely, assume L be an even integer; let m_n be a sequence of ± 1's, representing the message stream and let x_n be the transmitted stream. For $1 \leq k \leq \infty$ define

$$C_0 = 1 \quad \text{and} \quad C_k = C_0 + \sum_{n=1}^{k*L} x_n \qquad (1)$$

to actively maintain the digital sum variation (DSV) of the transmitted stream. The BR encoding scheme is summarized in Fig 1.

III. PSD of a BR Encoding Scheme of L=4

Let's consider a special case of $L = 4$ for BR encoding scheme, i.e., there three message bits in each frame of four bits (one redundant bit). Its Markov model consists of four states of $\{E_3, E_1, E_{-1}, \text{and} E_{-3}\}$ as shown by Fig 2. This is the model where waveforms being probabilistic functions of state transitions. Applying theorem in [3], we have the PSD:

DSV state		output bit
C>0	$\sum_{i=2}^{L} m_i > 0$	-1 $-m_2$ $-m_3$ ..... $-m_L$
	$\sum_{i=2}^{L} m_i < 0$	1 m_2 m_3 ..... m_L
C<0	$\sum_{i=2}^{L} m_i > 0$	1 m_2 m_3 ..... m_L
	$\sum_{i=2}^{L} m_i < 0$	-1 $-m_2$ $-m_3$ ..... $-m_L$

Fig. 1: The Bit Reversal Encoding Scheme

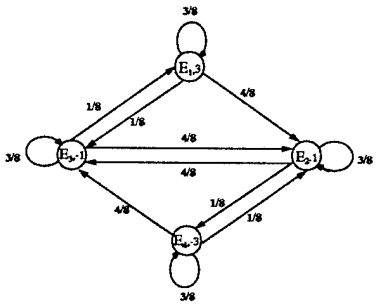


Fig. 2: The BR Encoding Markov Model of L=4

$$G(f) = \frac{4sinc^2(\pi fT)}{T} - \frac{2sinc^2(\pi fT)}{3(17 - 8\cos(2\pi fT))T}(5 + 8\cos(2\pi fT) + 19\cos(4\pi fT) + 13\cos(6\pi fT) + 13\cos(8\pi fT) - 4\cos(10\pi fT))$$

which equal to zero at D.C. The data bandwidth is increased by a very small fraction.

IV. Conclusion

We proposed and analyzed a new spectral shaping scheme without subcarriers, the BR encoding scheme. Unlike the traditional method modulating the data onto a subcarrier which will result in a much larger bandwidth than necessary to transmit the data, it increases the data bandwidth by a very small fraction, yet reduces the spectral energy near D.C. to nearly zero. We apply a general formula given in [3] to compute the PSD of the BR encoding scheme.

References

[1] R. C. Titsworth and L.R. Welch, "Power spectra of signals modulated by random and pseudorandom sequences," *JPL Tech. Report*, no. 32-140, Oct. 10, 1961.

[2] P. Galko and S. Pasupathy, "The mean power spectral density of Markov chain driven signals," *IEEE Trans. IT*, vol. IT-27, no. 6., pp. 746-754, Nov., 1981.

[3] Yongwen Yang, "Power spectrum density of Markov chain driving processes," Ph.D. Dissertation, CSI-91-12-03, Univ. of So. Cal., Los Angeles, CA 90089, 1991.

On Optimizing Multicarrier Transmission

T.J. Willink[1] and P.H. Wittke[2]

[1] Communications Research Centre, Ottawa and

[2] Department of Electrical and Computer Engineering, Queen's University at Kingston, Ontario

Abstract — **Optimum conditions for maximizing the throughput of an orthogonal frequency division multiplexed (OFDM) system are derived, and an algorithm for achieving them is presented. Theoretical bounds on performance are derived and used to compare OFDM with conventional equalized single carrier QAM for both the conventional error probability criterion and a criterion based on the mean-squared error. OFDM is shown to achieve greater throughput than equalized single carrier QAM, especially at low to intermediate signal-to-noise ratios and on channels with poor spectral properties.**

I. Summary

Orthogonal frequency division multiplexing (OFDM), a form of multicarrier transmission, has attracted attention as an alternative to equalized single carrier transmission over channels with spectral nulls, multipath or fading.

The principle of OFDM is to modulate many parallel subcarriers by dividing the high rate transmission data into lower rate sub-streams. For a correctly chosen subcarrier spacing, the modulated sub-streams are orthogonal, and hence interchannel interference (ICI) is avoided. For sufficiently narrow subchannel bandwidths, the system can be considered to be a set of parallel Nyquist I channels. Subchannels which have severe attenuation can then be avoided, and subchannels with good gain-to-noise characteristics can be exploited by allocating them more power and data. As each subcarrier is modulated using low rate data, the symbol period is far greater than for a single carrier modulated at the same total data rate. This mitigates the effects of impulsive noise and fading.

Early commercial multicarrier modems used guard intervals in the time and frequency domains to reduce the effects of intersymbol interference (ISI) and interchannel interference (ICI). Each subcarrier was modulated using the same power and data rate. Towards the end of the sixties, a number of authors, notably Chang [1], used overlapping orthogonal spectra to increase the efficiency of multicarrier systems. More recently, Kalet [2] introduced the concept of adjusting the power and data assigned to each subcarrier to increase the throughput further.

Kalet stated that maximum throughput would be achieved when the data and power assignments were such that each subcarrier achieved the same symbol error probability. Based on these assumptions, Zervos and Kalet [3] concluded that OFDM would not yield significantly greater throughput than decision feedback equalized single carrier transmission.

In this paper, we do not constrain the error probability to be the same over all the subcarriers. An optimization procedure is used to determine the conditions which must be met to achieve maximum throughput, and an iterative algorithm is presented which will rapidly achieve these conditions.

It is shown that, using the conventional error probability criterion, OFDM will in fact always outperform decision feedback equalized single carrier QAM. The increase in throughput is most significant at low and intermediate signal-to-noise ratios, where error propagation renders the DFE impractical.

As an example, the NEXT-dominated high-speed digital subscriber loop is considered. Using the results presented in [4] the exact error probability of single carrier QAM using a DFE will be compared to optimized OFDM, and it is seen that OFDM gives significantly better performance. At a data rate of 1.28 Mbps, the equalized single carrier can be used over wire lengths up to 11.5 kft at a bit error probability of 10^{-5}. For the same data rate and bit error probability, OFDM can be used for lengths up to 15.5 kft.

It is well-known that the mean-square error (mse) is a tractable criterion in the design of linear and decision feedback equalizers. We present a criterion for optimizing OFDM transmission which is based on the mse. It enables a direct comparison to be made between OFDM and equalized single carrier transmission in which the equalizer is designed using the minimum mse criterion. Examples show that OFDM again outperforms equalized single carrier QAM, especially at low and intermediate SNRs and on channels with poor spectral properties.

References

[1] R.W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission", *Bell Syst. Tech. J.*, vol. 45, pp.1775–1796, Dec. 1966.

[2] I. Kalet, "The multitone channel", *Proc. ICC*, pp. 1704–1710, 1987.

[3] N.A. Zervos and I. Kalet, "Optimized decision feedback equalization versus optimized orthogonal frequency division multiplexing for high-speed data transmission over the local cable network", *Proc. ICC*, pp. 1080–1085, 1989.

[4] T.J. Willink, P.H. Wittke and L.L. Campbell, "Error probability calculations for Decision Feedback Equalizers", *Proc. ISIT*, p. 356, Trondheim, 1994.

[5] T.J. Willink and P.H. Wittke, "Optimization and performance evaluation of multicarrier Transmission", *to be submitted to IEEE Trans. Inform. Theory.*

Multitone Modulation and Demodulation for Channels with SNR Uncertainty

Carl W. Baum and Keith F. Conner
Dept. of Elect. and Comp. Eng., Clemson University, Riggs Hall, Clemson, SC 29634-0915

Abstract — **A multitone transmission scheme that uses a nonlinear binary code to specify multitone signal constellations is proposed and motivated. A technique for designing the nonlinear code is presented, and methods for making symbol decisions and erasures without knowledge of signal and noise strength parameters are given. The performance of a system using this scheme with Reed-Solomon error control coding is discussed.**

I. Introduction

Multicarrier modulation schemes have received increasing attention for wireless communication systems including messaging systems [1] and microwave radio [2]. In this paper, we consider multicarrier schemes that employ on-off keying on orthogonally spaced subcarriers. Such systems are amenable to low-complexity noncoherent demodulation, and because multiple bits are transmitted per symbol, these systems also benefit from the advantages of long symbol durations including simulcasting capability and reduced requirements for channel equalization.

A natural approach is to use parallel channels independently; i.e., the data is partitioned into separate bit streams that are each modulated on separate subcarriers. In this work, we explore the use of a binary code *across* the separate bit streams to introduce and exploit dependencies between the modulated subcarriers.

II. System Model

The multitone modulation scheme we consider can be viewed as a generalization of M-ary frequency shift keying. A multitone channel encoder uses a binary (n, k) code to specify the mapping from data to multitone signal constellations $\{\mathbf{c}_i\}$, $i = 1, \ldots, 2^k$. The 1s in a codeword dictate which tones are transmitted simultaneously. An (N, K) singly extended Reed-Solomon (RS) code with $N = 2^k$ is employed to provide error control. Multitone symbols are interleaved to mitigate the effects of channel fading.

The demodulator consists of a bank of n energy detectors whose outputs are denoted by the vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. The output $\mathbf{y}$ is used by a decision device that makes symbol decisions or declares symbol erasures. The decision device is followed by a RS decoder that employs errors-and-erasures bounded-distance decoding.

Because the multitone constellations are not orthogonal (in general), a maximum likelihood detector requires knowledge of signal and noise parameters. In many practical applications, these parameters will not be known, and they may vary significantly with time. We therefore consider the use of decision devices based on simple linear combinations of the y_is. We have investigated decision rules of the following form:

$$\text{Choose} \begin{cases} \mathbf{c}_i \text{ if } i = \arg\max_m d^{(m)} \text{ and } d^{(i)} > b; \\ \text{erasure otherwise} \end{cases}$$

where b is a fixed threshold and $d^{(m)}$ defined as one of the following:

$$d^{(m)} = \frac{\mathbf{y} \cdot \mathbf{c}_m}{\|\mathbf{c}_m\|} \quad \text{or} \quad \frac{\mathbf{y} \cdot (\mathbf{c}_m - \bar{\mathbf{c}}_m)}{\|\mathbf{c}_m\|} \quad \text{or} \quad \frac{\mathbf{y} \cdot \mathbf{c}_m}{\|\mathbf{c}_m\|} - \frac{\mathbf{y} \cdot \bar{\mathbf{c}}_m}{\|\bar{\mathbf{c}}_m\|}$$

($\bar{\mathbf{c}}_m$ is the ones complement of $\mathbf{c}_m$, $\|\cdot\|$ denotes Hamming weight, and $\mathbf{y} \cdot \mathbf{0}/\|\mathbf{0}\| \equiv 0$).

III. Multitone Code Design

Using linear codes for specifying the multitone constellations results in very poor performance when the decision rules described above are used. We have therefore focused on the use of nonlinear codes for this purpose. Our approach is to choose a linear code with good Hamming distance properties and generate various nonlinear codes from this code by applying different combinations of bit inversions (inverting the ith bit of every codeword in the code for various values of i). The resulting codes can be have the same distance properties as the original linear code, but with different weight distributions.

Numerical results show that codes containing codewords with very small or very large Hamming weight perform worse than codes with less variation in weight. This fact may lead one to conclude that constant weight codes should be used. However, the Hamming distance properties of constant weight codes are usually inferior to those of codes based on the best linear codes. We have shown that the guaranteed error correcting capability t_m of constant weight codes must satisfy

$$\left[\sum_{\substack{j=0 \\ j \text{ even}}}^{t_m} \binom{\ell}{\frac{j}{2}} \binom{n-\ell}{\frac{j}{2}} \right] 2^k \le \binom{n}{\ell},$$

where the value(s) of ℓ that maximize t_m must include $\ell = \lfloor \frac{n}{2} \rfloor$ or $\lceil \frac{n}{2} \rceil$ (t_m must also satisfy $t_m \le \min(2\ell, 2(n-\ell))$). The error correcting capability of many known linear codes exceeds this bound.

IV. System Performance

We have shown that, for a system using errors-only decoding of the RS code, the combination of a good nonlinear multitone code and the decision rule described above gives performance in AWGN close to that of the corresponding linear multitone code with maximum likelihood decoding. Additionally, we have shown that incorporation of errors-and-erasures decoding provides significant performance improvements in channels subject to AWGN and Rayleigh fading.

References

[1] R. Petrovic, W. Roehr, and D. Cameron, "Multicarrier modulation for narrowband PCS," *IEEE Trans. Veh. Technol.*, pp. 856-862, Nov. 1994.

[2] S. Aikawa, Y. Nakamura, and H. Takanashi, "Performance of trellis coded 256 QAM super-multicarrier modem using VLSIs for SDH interface outage-free digital microwave radio", *IEEE Trans. Commun.*, pp. 1415-1421, Feb./Mar./Apr. 1994.

Achievable Rates for Tomlinson–Harashima Precoding

Richard D. Wesel[1] and John M. Cioffi

Information Systems Laboratory, Stanford University, Stanford, California 94305

Abstract — **The maximum achievable information rate of the zero–forcing Tomlinson–Harashima precoder (ZF–THP) is given exactly. Bounds are provided for the minimum mean square error (MMSE) THP. Performance of THP is characterized on an example channel, and discussed for arbitrary channels.**

Consider the power–constrained additive white gaussian noise (AWGN) channel with intersymbol interference (ISI) where a real input sequence[2] $X(D)$ with $E[x_k{}^2] \leq P$ is filtered by $H(D)$ and distorted by real AWGN n_k.

For this channel, we compute the reduction in achievable rate from capacity incurred by Tomlinson–Harashima precoding (THP) combined with codes designed for AWGN without ISI. Loss due to finite complexity codes will be neglected.

Figure 1 shows a general THP system. Linear time invariant filters $F(D)$ and $B(D)$ are chosen to minimize optimality criteria discussed below. $B(D)$ must be causal and monic.

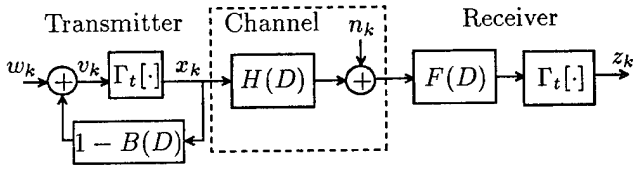


Figure 1: Communication system using THP

Γ_t is a mapping from $\mathcal{R}$ to $(-t/2, t/2]$ where $t \in \mathcal{R}^+$. Specifically, $\Gamma_t[v_k] = v_k + a_k$ where a_k is the integer multiple of t for which $\Gamma_t[v_k] \in (-t/2, t/2]$. Figure 2 is equivalent to Figure 1 with a_k as defined above. The noise $\hat{n}$ is n filtered by $F(D)$.

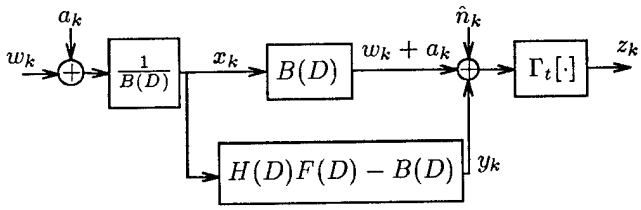


Figure 2: Communication system equivalent to Figure 1.

ZF–THP is the scheme originally proposed in [1, 2] with $F(D)$ and $B(D)$ chosen so that $y_k = 0$ and $\hat{n}_k$ is white. The ZF–THP system is a memoryless channel with input w and output $\Gamma_t[w + \hat{n}]$. The channel inputs are constrained by $w \in (-t/2, t/2]$ and $E[x^2] \leq P$. THP transmitter output power is roughly $t^2/12$ for large alphabet PAM [3]. Thus we restrict our attention to the choice of t which obeys the power constraint with equality (i.e. $t = \sqrt{12P}$). This system's achievable rate is

$$I_{ZF-THP} = \log_2(t) - h(\Gamma_t[\hat{n}]) \qquad (1)$$

where $h(\cdot)$ denotes differential entropy.

The MMSE–THP is obtained by choosing $F(D)$ and $B(D)$ to minimize $\text{VAR}(\hat{n} + y)$. Ideal interleaving is assumed which produces a memoryless channel with input w and output $\Gamma_t[w + y + \hat{n}]$, where w is constrained as above. Our bounds for this system are

$$I_{MMSE-THP} \geq \log_2(t) - h(\mathcal{T}(\sigma^2, t)) \qquad (2)$$

$$I_{MMSE-THP} \leq \log_2(t) - h(\Gamma_t[\hat{n}]) \qquad (3)$$

where $\mathcal{T}(\sigma^2, t)$ is a zero mean Gaussian truncated to $(-t/2, t/2]$ with variance $\sigma^2 = \text{VAR}(y + \Gamma_t[\hat{n}])$ after truncation. Note that $\text{VAR}(\hat{n}_k)$ depends on $F(D)$ and thus the right hand sides of Equations (1) and (3) are not equivalent.

Figure 3 plots Equations (1) (2) and (3) as well as capacity for a 50 tap bandpass ISI channel with AWGN.

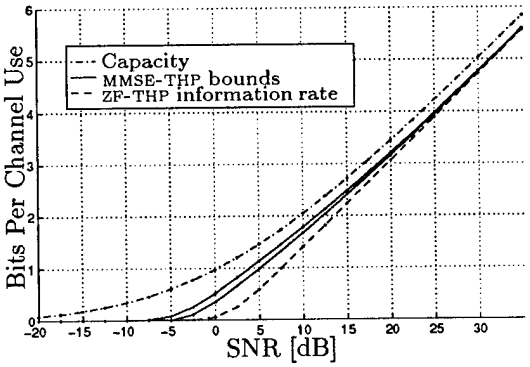


Figure 3: Information rates for example channel

For any $H(D)$ both MMSE bounds converge to $\log_2(t) - \frac{1}{2}\log_2(2\pi e E[\hat{n}^2])$ as SNR $\to \infty$. At high SNR, ZF–THP and MMSE–THP identically suffer only the 1.53 dB or .255 bit "shaping loss" from capacity regardless of $H(D)$.

At low SNR, THP achievable rates can still be considerably below capacity. Here, the loss is due entirely to the receiver Γ_t (and interleaving for the MMSE–THP). This behavior is the reverse of that observed at high SNR where the loss was entirely due to the transmitter Γ_t.

In Figure 3 the MMSE–THP outperforms the ZF–THP. This may be true for all H(D). We have shown that the ZF–THP rate will never be more than .08 bits per channel use above the MMSE–THP rate.

ACKNOWLEDGEMENT

Shlomo Shamai provided several helpful suggestions including using the maximum entropy property of the truncated Gaussian. We enjoyed interesting discussions with Amos Lapidoth, Erik Ordentlich, and Yoichi Matsumoto.

REFERENCES

[1] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Letters*, 7:138–139, March 1971.

[2] H. Harashima and H. Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE T. Comm.*, 20(4):774–780, August 1972.

[3] J. Mazo and J. Salz. On the transmitted power in generalized partial response. *IEEE T. Comm.*, 24(3):348–352, March 1976.

[1] Email: *wesel@isl.stanford.edu*. This work was supported by an AT&T Foundation Fellowship and NSF grant NCR-9203131.

[2] Sequences will be denoted by their formal D-transforms $X(D) = \sum_k x_k D^{-k}$.

Phase-Shifted Linear Partial-Response Modulation

Amir Said[1]
amir@densis.fee.unicamp.br, Faculty of Electrical Engineering
State University of Campinas (UNICAMP), Campinas, SP 13081, Brazil

John B. Anderson
anderson@ecse.rpi.edu, Dept. of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180, U.S.A.

I. Introduction

Partial-response modulations are widely used for spectrum shaping and bandwidth reduction. Linear partial-response signaling (PRS) is a well-known form of linear coded modulation which traces back to Lender [2] in 1963. Due to its "age," it is widely believed that the properties of linear PRS have been thoroughly investigated. However, this is not the case. Due to the lack of efficient algorithms for maximum-likelihood detection the research on the subject concentrated on the simplest forms of PRS, and mainly with suboptimal detection [1].

In this paper we analyze the properties of PRS signals generated from complex-valued functions. Those signals have not only the usual intentional intersymbol interference (ISI), but also an intentional interference between the quadrature components of the RF modulated signal. Our objective is to show how these two forms of interference should be used together.

The theory presented here is particularly important in the design of PRS generators for bandwidth efficient coded modulation [3], which are schemes with severe ISI. Furthermore, since the results are valid for intentional or non-intentional ISI, they can also be used to improve performance of communications in non-ideal channels.

The lowpass representation of a linear PRS is defined by

$$s_1(t) = \sum_{n=-\infty}^{\infty} u[n]h(t-nT), \tag{1}$$

where $u[n]$ is the data sequence, and $h(t)$ is a spectrum shaping (generator) pulse. Note that here both $u[n]$ and $h(t)$ may be complex-valued, with the real and imaginary parts of $s_1(t)$ corresponding to the *in-phase* and *quadrature* components of the RF modulated signal. In most of the literature about linear PRS coding the shaping pulse $h(t)$ is considered to be real-valued, even when the data sequence $u[n]$ is complex-valued (e.g., partial-response QAM modulation). Here we assume that the imaginary part of $h(t)$ adds interference between the quadrature components. We stress that the signals $s_1(t)$ considered in this paper do not exist at baseband as a real signal; rather, they take form only as RF signals. One can consider the problem here as synthesis of coded signals directly at RF.

The spectrum used by (1) is defined by the spectral power density

$$\Phi_{s_1}(f) = \alpha|H(f)|^2, \tag{2}$$

where α is a constant, and $H(f)$ is the Fourier transform of $h(t)$. Note that the dependence between the in-phase and quadrature components makes the spectral power density asymmetric, i.e., $H(f) \neq H(-f)$.

[1]This work was supported by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

II. Phase-Shifted-Data (Generalized) PRS

The combination of intersymbol interference with the quadrature components interference produce some unexpected results. For instance, it is demonstrated [3] that if we apply a complex frequency shift to $h(t)$, and define the signal

$$s_2(t) = \sum_{n=-\infty}^{\infty} u[n]h(t-nT)\ e^{j\,2\pi f_s(t-nT)}, \tag{3}$$

then we can get a PRS with a energy/bandwidth performance quite different from that of signal (1). (Of course, in the well-known case where there is no ISI, there is no change in performance.) We evaluate those effects by deriving the theoretical asymptotic error probability, and also measuring it via simulations, and it is shown that the energy efficiency can improve when $f_s \neq 0$.

Alternately, if we apply a phase shift to $u[n]$, and define

$$s_3(t) = \sum_{n=-\infty}^{\infty} u[n]h(t-nT)\ e^{j\,2\pi n\,f_s\,T}, \tag{4}$$

then it can be proved that the generalized PRS (4) has *exactly* (not only asymptotically) the same noise immunity as (3). At the same time, the spectrum used by (4) is the same used by (1), and it is not shifted as with (3).

Some important conclusions follow from the results above:

- When the pulse $h(t)$ is set by the channel ISI, the frequency/phase shifts of (3) or (4) allow us to improve performance by relieving the effect of the ISI.
- During the design of optimized complex-valued PRS, subject to a bandwidth constraint, it is necessary to consider a sliding bandwidth parameter in order to find the optimal signals. Alternatively, the signal can be designed for a bandwidth centered at $f = 0$, and be optimized for the generalized PRS (4),
- Better PRS coding schemes may be synthesized by frequency and phase shifts: in a given RF bandwidth schemes with better free distance exist, or at a fixed distance, schemes with better bandwidth exist.

References

[1] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. on Commun.*, vol. 23, pp. 921–934, Sept. 1975.

[2] A. Lender, "The duobinary technique for high-speed data transmission," *IEEE Trans. Commun. Electron.*, vol. 82, pp. 214–218, May 1963.

[3] A. Said, *Design of Optimal Signals for Bandwidth-Efficient Linear Coded Modulation,* Ph.D. Thesis, also in Communication, Information and Voice Processing Report Series, TR93–3, Electrical, Computer and Systems Department, Rensselaer Polytechnic Institute, Troy, NY, Oct. 1993.

Self-Training Adaptive Equalization for Multilevel Partial-Response Transmission Systems

G. Cherubini, S. Ölçer, and G. Ungerboeck

IBM Research Division, Zurich Research Laboratory,
CH–8803 Rüschlikon, Switzerland

Abstract — We present a new self-training method for adjusting the coefficients of a transversal equalizer with T-spaced taps in a multilevel partial response class-IV (PRIV) system. Self-training equalization from distorted random data signals is inherently more difficult to achieve for partial-response systems than for full-response systems. Also, because of the lack of excess bandwidth, traditional bandedge timing recovery schemes cannot be applied in PRIV systems. On the other hand, an equalizer with T-spaced taps is sufficient to obtain equalized output signals for arbitrary sampling phase. Convergence with the known self-training algorithm by Sato is too slow to ever reach satisfactory performance, e.g., for switching to decision-directed equalization, when no recovered clock is available and the phase of the local receiver clock drifts only slightly relative to the phase of the received signal. Following Sato, in the described self-training equalization algorithm we first transform the equalizer output into full-response form, then compute a pseudo-error signal, and finally translate the pseudo-error signal into an error signal for the desired partial-response equalizer output. This error signal is used to adjust the equalizer coefficients according to the LMS algorithm. The new method differs from the Sato algorithm in two ways. First, the channel inversion for obtaining full-response signals is accomplished exactly by mixed linear feedback and decision feedback equalization, whereas in the case of the Sato algorithm the inversion is only achieved approximately. Secondly, for the derivation of pseudo-error signals more knowledge of the statistical properties of ideal, but noisy full-response signals is exploited. Basically, the pseudo errors are obtained from knowledge of the largest positive and negative symbol values and the probability of the occurrence of equalized signals in the interval between these values and outside these values. In the absence of noise, the new pseudo-error signals vanish as equalization is achieved. We present simulation results illustrating the superior convergence properties of the new self-training method.

Summary

Self-training adaptive equalization has mainly been studied for full-response systems in the past, e.g., in [1]-[3]. Methods to achieve self-training equalization for partial-response systems have been proposed in [4] and [5] for linear and distributed-arithmetic equalizers, respectively.

We denote the output of the linear equalizer by y_n:

$$y_n = \boldsymbol{c}_n \boldsymbol{x}_n^T , \tag{1}$$

where $\boldsymbol{c}_n = \{c_{0,n}, ..., c_{N-1,n}\}$ represents the vector of equalizer coefficients and $\boldsymbol{x}_n = \{x_n, ..., x_{n-N+1}\}$ the vector of signals stored in the equalizer delay line at time n. The objective of an adaptive equalizer for a PRIV system is to provide an equalized signal of the form

$$y_n = (a_n - a_{n-2}) + e_n , \tag{2}$$

where a_n is the channel-input symbol and e_n is an error signal due to noise and residual signal distortion. We describe the algorithm for quaternary modulation. In this case, $a_n \in \{-3, -1, +1, +3\}$.

We first transform the equalizer output y_n signal into a full-response signal u_n by channel inversion via mixed linear feedback and decision feedback:

$$u_n = y_n + \varrho\, u_{n-2} + (1 - \varrho)\, \hat{a}_{n-2} , \tag{3}$$

where $\hat{a}_n$ is a tentative quaternary decision on the transmitted symbol a_n based on the signal u_n, and $0 < \varrho < 1$. We then define a pseudo-error ε_n by

$$\varepsilon_n = \begin{cases} u_n - \hat{a}_n & \text{if } |u_n| \geq 3 \\ -\delta_n \operatorname{sign}(u_n) & \text{otherwise,} \end{cases} \tag{4}$$

where δ_n is a non-negative value updated at each iteration as follows:

$$\delta_{n+1} = \begin{cases} \delta_n - \frac{3}{4}\Delta & \text{if } |u_n| \geq 3 \\ \delta_n + \frac{1}{4}\Delta & \text{otherwise,} \end{cases} \tag{5}$$

and Δ is a positive constant. The generation of the pseudo-error ε_n is based on *a priori* knowledge of the statistics of the signal u_n. In the case of accomplished equalization, u_n corresponds to the quaternary channel input symbol a_n embedded in noise. Therefore, whenever the event $|u_n| \geq 3$ is observed, we can use $u_n - \hat{a}_n$ as a trusted error to update the equalizer coefficients. If we observe the event $|u_n| < 3$, no trusted error is available. In this case, we choose to update the equalizer coefficients so that the probabilities of the events $|u_n| < 3$ and $|u_n| \geq 3$ assume the values 3/4 and 1/4, respectively, which are the probabilities of these events for an ideally equalized, noisy quaternary signal. This is achieved by setting the pseudo-error equal to $-\delta_n \operatorname{sign}(u_n)$ whenever $|u_n| < 3$ and updating the value of δ_n at each iteration so that δ_n becomes larger if the event $|u_n| < 3$ occurs more often than expected and smaller otherwise.

The LMS algorithm for self-training adaptive equalization is given by

$$\boldsymbol{c}_{n+1} = \boldsymbol{c}_n - \alpha\, (\varepsilon_n - \varrho\, \varepsilon_{n-2})\, \boldsymbol{x}_n , \tag{6}$$

where α is the adaptation gain.

We present simulation results which show that convergence is achieved even in the presence of significant initial clock drift. The new algorithm outperforms the known self-training technique for PRIV systems by Sato [4] in terms of speed of convergence and achievable mean-square error in the steady-state. The new approach has been realized in a prototype transceiver for full-duplex transmission at 125 Mbit/s over telephone-grade twisted-pair cables, which also employs adaptive near-end crosstalk cancellation.

References

[1] D. N. Godard, "Self recovering equalization and carrier tracking in two-dimensional data communications systems", *IEEE Trans. Commun.*, Vol. COM-28, pp. 1867-1875, Nov. 1980.

[2] S. Bellini, "Bussgang techniques for blind equalization", Proc. of *IEEE GLOBECOM* 1986, pp. 46.1.1–46.1.7, Dec. 1986.

[3] G. Picchi and G. Prati, "Blind equalization and carrier recovery using a "Stop-and-Go" decision directed algorithm", *IEEE Trans. Commun.*, Vol. COM-35, pp. 877-887, Sept. 1987.

[4] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems", *IEEE Trans. Commun.*, Vol. COM-23, pp. 679-682, June 1975.

[5] G. Cherubini, "Nonlinear self-training adaptive equalization for partial-response systems", *IEEE Trans. Commun.*, Vol. COM-42, pp. 367-376, Feb. 1994.

On the Existence and Uniqueness of Joint Channel and Data Estimates

Keith M. Chugg and Andreas Polydoros
Communication Sciences Institute, EE–System Dept.
University of Southern California
Los Angeles, California, 90089-2565

Abstract — **Results regarding two aspects of joint Maximum Likelihood (ML) data sequence and inter-symbol interference channel estimation are considered: (i) the joint-ML estimation problem is ill-posed when based on the continuous time observation, and (ii) processing based on a discrete time signal model yields equivalence classes of data sequences.**

I. Continuous Time Processing

We consider joint-ML estimation of a digital data sequence $\{a_k\}$ and a dispersive channel impulse response $h(t)$ from the complex baseband model

$$r(t) = y(t) + n(t) = \sum_i a_i h(t - iT) + n(t), \qquad (1)$$

where $n(t)$ is baseband equivalent of additive white Gaussian noise. The data sequence is assumed to be independent and uniformly distributed over a finite alphabet, while $h(t)$ is assumed to be a static, deterministic function with support contained in $[0, LT)$.

A "chipped" signal notation is introduced to convert an arbitrary function $m(t)$ into a vector of chip functions

$$\mathbf{m}_i(t) = \left[\begin{array}{cccc} m_i(t) & m_{i-1}(t) & \cdots & m_0(t) \end{array} \right]^T, \qquad (2)$$

where the i^{th} chip is $m_i(t) = m(t + iT)$ for $t \in [0, T)$ and zero otherwise. Applying this notation to the model in (1) for the observation interval $[0, kT)$ yields

$$\mathbf{r}_k(t) = \mathbf{y}_k(t) + \mathbf{n}_k(t) = \mathbf{A}_k \diamond \mathbf{h}(t) + \mathbf{n}_k(t), \qquad (3)$$

where the $((k+1) \times L)$ Toeplitz data matrix $\mathbf{A}_k$ has i^{th} row

$$\boldsymbol{\alpha}_i^H = \left[\begin{array}{cccc} a_{i-L+1} & a_{i-L+2} & \cdots & a_i \end{array} \right]. \qquad (4)$$

For a hypothesized $\tilde{\mathbf{A}}_k$, minimization of the joint-ML metric over $\tilde{\mathbf{h}}(t)$ results in the critical point $\hat{\mathbf{h}}(t; \tilde{\mathbf{A}}_k) = \tilde{\mathbf{A}}_k^I \diamond \mathbf{r}_k(t)$, where $\tilde{\mathbf{A}}_k^I$ is the pseudo-inverse of $\tilde{\mathbf{A}}_k$. Substitution yields a metric dependent only on the hypothesized data sequence

$$\Gamma_k(\tilde{\mathbf{A}}_k) \triangleq \Gamma_k(\tilde{\mathbf{A}}_k, \hat{\mathbf{h}}(t; \tilde{\mathbf{A}}_k)) = -\int_0^T \mathbf{r}_k^H(t) \diamond \tilde{\mathbf{P}}_k \diamond \mathbf{r}_k(t) dt, \qquad (5)$$

where $\tilde{\mathbf{P}}_k = \tilde{\mathbf{A}}_k \tilde{\mathbf{A}}_k^I$ is the matrix which projects onto the range of $\tilde{\mathbf{A}}_k$.

The metric suggested in (5) does not exist in the mean-square sense. To illustrate this, consider a fixed $t \in [0, T)$ so that

$$\mathbb{E}\left\{ \mathbf{n}_k^H(t) \tilde{\mathbf{P}}_k \mathbf{n}_k(t) \right\} = \mathrm{tr}\left(\tilde{\mathbf{P}}_k \mathbb{E}\left\{ \mathbf{n}_k(t) \mathbf{n}_k^H(t) \right\} \tilde{\mathbf{P}}_k \right), \qquad (6)$$

which is not well defined since $\mathbb{E}\{n(t+\tau) n^*(t)\} = N_0 \delta(\tau)$. The conclusion – i.e., that the joint-ML channel and sequence estimation problem for the model of (1) is ill-posed – holds even when the noise is colored [1].

II. Practical Processing

Front-end processing structures exist which neglect a small amount of high-frequency energy and circumvent the ill-posed problem [2]. The output of such a front-end is modeled by a discrete time version of (3): $\mathbf{z}_k = \mathbf{A}_k \diamond \mathbf{f} + \mathbf{w}_k$, with a metric function analogous to (5)

$$\Lambda_k(\tilde{\mathbf{A}}_k) = \| (\mathbf{I} - \tilde{\mathbf{P}}_k) \diamond \mathbf{z}_k \|^2. \qquad (7)$$

The residual least-squares error metric of (7) can be computed recursively with k, which allows the problem to be formulated as a tree-search with per-sequence channel estimation [1]. Practical recursive algorithms truncate this search, maintaining only a finite number of candidate paths.

III. Equivalent Sequences

The metric function in (7) implies that data sequences with matrices having the same range are indistinguishable. Thus, two data matrices $\mathbf{A}_k$ and $\mathbf{D}_k$ are equivalent if the associated projection matrices are equal: $\mathbf{P}_{\mathbf{A}_k} = \mathbf{P}_{\mathbf{D}_k}$. We use the notation $\mathbf{A}_k \equiv \mathbf{D}_k \in \mathcal{E}(\mathbf{A}_k)$, where $\mathcal{E}(\mathbf{A}_k)$ is the set of all admissible data matrices with the same range as $\mathbf{A}_k$. An equivalent characterization is that there exists an invertible $(L \times L)$ matrix $\mathbf{M}$ such that $\mathbf{A}_k = \mathbf{D}_k \mathbf{M}$.

We characterize these classes as either *memoryless equivalence classes* or *memory equivalence classes.* A memoryless equivalence class is one in which $\mathbf{M}$ is diagonal, and results from rotational invariance in the symbol constellation.

Theorem: For BPSK signals ($a_k \in \{-1, +1\}$)

$$\lim_{k \to \infty} P(\mathcal{E}(\mathbf{A}_k) = \{\mathbf{A}_k, -\mathbf{A}_k\}) = 1, \qquad (8)$$

where the probability is over all $\mathbf{A}_k$.

The proof follows from two facts: (i) the probability that all 2^L possible values of $\boldsymbol{\alpha}$ will appear in $\mathbf{A}_k$ goes to one, and (ii) for those $\mathbf{A}_k$ which contain all possible rows, there are only a finite number of $\mathbf{M}$ which yield admissible data matrices.

The result suggests that for asymptotically large k, the effect of the equivalence classes can be negated by differential encoding and decoding (i.e., one need only be concerned with memoryless equivalence classes). However, the effect of memory equivalence classes on the short term acquisition properties of practical algorithms is significant.

References

[1] K. M. Chugg and A. Polydoros, "MLSE for an Unknown Channel – Part I: Optimality Considerations," accepted for publication in *IEEE Trans. Communications.*

[2] K. M. Chugg and A. Polydoros, "Front-End Processing for Joint Maximum Likelihood Channel and Sequence Estimation," *Proc. CTMC-Globecom '94,* CTMC02.3.

Data Detection of Coded PSK in the Presence of Unknown Carrier Phase[1]

Carl R Nassar and M Reza Soleymani[2]
Department of Electrical Engineering, McGill University, Montreal, Quebec, Canada

Abstract — **Our work introduces a novel data detection scheme for coded PSK in the presence of unknown phase. This scheme offers a performance very close to coherent in cases of $n \geq 20$, and requires a low complexity.**

I. Introduction

Coded PSK demonstrates an extreme sensitivity to unknown channel phase. Without a careful effort to deal with this phase, the gain of coded PSK may be greatly diminished. In the case of slowly varying phase (e.g. constant over 500 symbols), several effective data detection strategies have been proposed. However, in communication over a channel with rapidly changing phase, these strategies are ineffective. Recently proposed schemes for data detection in a rapid phase change environment are based on extending the ideas of Multiple Symbol Differential Detection (MSDD) to coded modulation (e.g. [1]). These schemes offer some gains over coded DPSK, but they are still unable to match coherent performance. We introduce a novel coded PSK data detection scheme which offers a performance very close to coherent in cases of constant phase over 20 or more symbols. This scheme requires a low complexity, and it employs a Viterbi Algorithm (VA) implementation.

II. Receiver Design

The received signal is represented by $\underline{r} = (r_0, r_1, ..., r_{N-1})$, where $r_i = a_i e^{j\theta_i} + \eta_i$. Here, η_i's represent samples from an AWGN source; θ_i corresponds to the channel's phase rotation; and a_i corresponds to a differentially encoded MPSK symbol generated at sample time i by a trellis encoder. The differential encoding and trellis encoder are chosen to create $\frac{2\pi}{M}$ phase invariance [2].

Our data detection scheme is based on ML detection. According to ML detection, the best output sequence $\hat{\underline{a}}$, given a received $\underline{r}$, is $\hat{\underline{a}} = \arg\max_{\underline{a}\in A} p(\underline{r}|\underline{a})$, where A is the set of possible $\underline{a}$ sequences generated at the transmitter. Introducing the unknown phase, this becomes $\hat{\underline{a}} = \arg\max_{\underline{a}\in A} \int_N p(\underline{r}|\underline{a},\underline{\theta})p(\underline{\theta})d\underline{\theta}$, where $\int_N$ refers to N^{th} order integration (one integral per phase θ_i in $\underline{\theta}$).

Our derivation continues by introducing the information regarding phase. It is assumed that θ_i is constant over a block of n symbols, that is, $\theta_0 = \theta_1 = ... = \theta_{n-1}$, $\theta_n = \theta_{n+1} = ... = \theta_{2n-1}$, and so on. Using this, we simplify our integral equation and achieve an intermediate result.

We complete our derivation by approximating the continuous phase space by a discrete phase space. The continuous phase space is $\Phi = [0, 2\pi)$. However, because we have introduced a differential encoding and TCM code which create $\frac{2\pi}{M}$ phase invariance, we can map the output of our receiver into the correct $[\frac{2\pi i}{M}, \frac{2\pi(i+1)}{M})$ sector of space by following our receiver with a differential decoder. Hence, it suffices, for the purposes of our receiver, to represent the continuous phase space by $\Theta = [0, \frac{2\pi}{M})$. We approximate Θ using $\tilde{\Theta} = \{\frac{2\pi}{M}\frac{2j+1}{2m}, j = 0, 1, ..., m-1\}$. It can be shown that $m = 4$ is sufficient to achieve good results. Replacing the continuous phase space by $\tilde{\Theta}$ in our ML equation leads to our final result. Specifically, the discretizing of the phase space results in the integrals becoming summations. Additionally, it is easily shown that each sum is well approximated by the largest term in the sum. This results in: choose the $\hat{\underline{a}}$ from

$$\max_{\tilde{\theta}_0\in\tilde{\Theta},\underline{a}_0} \sum_{i=0}^{n-1} \ln p(r_i|a_i,\tilde{\theta}_0) + \max_{\tilde{\theta}_n\in\tilde{\Theta},\underline{a}_n,E(\underline{a}_0)} \sum_{i=n}^{2n-1} \ln p(r_i|a_i,\tilde{\theta}_n) + ... + \max_{\tilde{\theta}_{Ln}\in\tilde{\Theta},\underline{a}_{Ln},E(\underline{a}_{(L-1)n})} \sum_{i=Ln}^{N-1} \ln p(r_i|a_i,\tilde{\theta}_{Ln}), \quad (1)$$

where $\underline{a}_0 = (a_0, ..., a_{n-1})$, $\underline{a}_n = (a_n, ..., a_{2n-1})$, and $\underline{a}_{Ln} = (a_{Ln}, ..., a_{N-1})$; and $E(\underline{a}_0)$ refers to the end node of sequence $\underline{a}_0$.

III. Implementation

This equation is implemented as follows. Consider first the block of symbols $\underline{a}_0 = (a_0, ..., a_{n-1})$. We can choose the best $\underline{a}_0$ and $\tilde{\theta}_0$ to each end node $E(\underline{a}_0)$, since this is the only term future symbols depend on. By best, we mean the values which maximize the first sum in the above equation. This selection of the best $(\underline{a}_0, \tilde{\theta}_0)$ can be carried out by using the VA, with metric $\ln p(r_i|a_i, \tilde{\theta}_0)$, over the first block of n symbols. Specifically, four VA's are carried out, 1 for each possible $\tilde{\theta}_0$. Next, consider the second block of symbols, $\underline{a}_n$. Much like the previous set $\underline{a}_0$, the $\underline{a}_n$ and $\tilde{\theta}_n$ can be chosen to each end node $E(\underline{a}_n)$. Their selection can be carried out using 4 VA's over the block of n symbols, $\underline{a}_n$, each with path metric $\ln p(r_i|a_i, \tilde{\theta}_n)$ (and a unique $\tilde{\theta}_n \in \tilde{\Theta}$). Here, each path is weighted by the appropriate start node value. This continues, in an analogous fashion, over the remaining blocks of symbols. Putting this together, we essentially have 4 VA's running over the block of symbols.

IV. Performance

The performance of this scheme increases as n increases. Most notably, considering rate 2/3, 8-PSK TCM, the performance of this scheme is very close to coherent for all $n \geq 20$.

References

[1] D. Divsalar, M.K. Simon, M. Shahshahani, "The performance of trellis-coded MDPSK with multiple symbol detection," *IEEE Trans. Commun.*, Vol. 38, pp. 1391-1403, Sept. 1990.

[2] M. Oerder, "Rotationally invariant trellis codes for mPSK modulation," presented at *ICC'85*, Chicago, Ill., June 23-26, 1985, pp. 552-556.

[1] This work is supported by NSERC Grant OGP/N011 and NSERC Scholarship 106418

[2] also with SPAR Aerospace Limited, Satellite and Communications Systems Division, Ste-Anne-de-Bellevue, Quebec.

Delayed Decision Feedback Equalization[1]

MAHESH K. VARANASI

ECE Dept, University of Colorado, Boulder, Colorado 80309. *varanasi@spot.colorado.edu*

Summary– Decision feedback equalization (DFE) is generalized within the context of linearly modulated data transmission over intersymbol interference (ISI) channels. The main motivation for this new approach is that for channels with severe ISI, linear and decision feedback equalizers have a poor performance while the the Viterbi algorithm has a complexity that is exponential in the length of the ISI channel response.

The delayed decision feedback equalizer (DDFE) introduced in this work applies to FIR as well as IIR channel responses. It is parametrized by two integer design parameters M and L with $L \leq M$ and a subset $S \subseteq \{1, \cdots, M\} \equiv \Omega$ of indices with the cardinality of S being equal to L. This DDFE is denoted as (M, L, S)-DDFE. The parameter M is equal to the decision delay in units of symbol duration, L determines the computational complexity per symbol (CCS) of the DDFE algorithm which is $O(F^L)$ where F is the data symbol alphabet size. For a given channel, and fixed values of M and L, the subset S is chosen to optimize the performance of the DDFE. This optimized DDFE is denoted as the (M, L)-DDFE. The subset optimization adds only to the design complexity but not the implementation complexity for a fixed channel. Performance is defined as the SNR gain over the conventional DFE in the high SNR region.

The connections with previous results are as follows. In the degenerate case where $M = L = 1$, the DDFE reduces to the conventional DFE [1]. For a given L, when $M = L$, we have $S = \Omega$, so that the (L, L, Ω)-DDFE is equivalent to the (L, L)-DDFE, which can be shown to be equivalent to the $(L, 1)$-BDFE (block decision feedback equalizer) of [2]. For this case, our performance analysis sheds new light on the BDFE.

Example– Consider a binary PAM-ISI, monic, causal, min-phase channel $G(z) = \sum_{i=0}^{\infty} g(i) z^{-i}$ with $g(1) = \alpha$ and a antipodal symbol alphabet $\{+1, -1\}$. For the $(2, 1)$-BDFE which is also the $(2, 2)$-DDFE, it can be shown that the SNR gain is given as

$$\eta_{(2,2)-DDFE} = \begin{cases} 1 + \alpha^2 & \text{if } |\alpha| \leq 1/2 \ ; \\ 1 + (1 - |\alpha|)^2 & \text{else.} \end{cases}$$

This SNR gain is thus greater than unity implying a uniformly better performance than the conventional DFE. Applying this result for the case of the single-zero channel model $1 + \alpha z^{-1}$, we can deduce that

$$\eta_{(2,2)-DDFE} = \begin{cases} \eta_{VA} & \text{if } |\alpha| \leq 1/2 \ ; \\ \frac{1+(1-|\alpha|)^2}{1+\alpha^2} \eta_{VA} & \text{else} \end{cases}$$

where η_{VA} is the SNR gain over the conventional DFE of the Viterbi Algorithm so that when $|\alpha| \leq 1/2$, the $(2, 2)$-DDFE has a performance that is indistinguishable from the more complex Viterbi algorithm.

The (M, L)-BDFE of [2] when $L > 1$ is not a useful generalization of $(M, 1)$-BDFE. The only "block" size in the feedback loop that is meaningful in block decision feedback equalization is 1, the degenerate case. The reason is as follows. The CCS of the (M, L)-BDFE is $O(F^M)$ and is relatively independent of L (for sufficiently large values of these parameters so as to ignore polynomial dependencies). Furthermore, it is out-performed by the $(M, 1)$-BDFE. A stronger result is that the (M, L)-BDFE is outperformed by the $(N, 1)$-BDFE (or equivalently the (N, N)-DDFE) where $N = M - L + 1$. Therefore, among the (M, L)-BDFEs, those with $L > 1$ can be outperformed by the corresponding $(N, 1)$-BDFE which has a better performance *and* a lower complexity.

The (M, L)-DDFE with $M > L$ on the other hand, performs no worse than the $(L, 1)$-BDFE (or equivalently the (L, L)-DDFE). This is an appropriate comparison because the CCS of both these schemes is given by $O(F^L)$. Consequently, even the best candidates from the BDFEs can be improved for the same CCS by the DDFEs. Moreover, the (M, L_1)-DDFE uniformly outperforms the (M, L_2)-DDFE when $L_2 < L_1$ which is to be expected since the complexity of the former is greater than that of the latter. No surprises here. The following example illustrates the superiority of the DDFE over the BDFE of the same complexity.

Consider a binary PAM-ISI, causal, monic, min-phase channel $G(z) = \sum_{i=0}^{\infty} g(i) z^{-i}$ and let $g(1) = 1/8$ and $g(2) = -31/64$. It can be shown that the $(3, 2)$-DDFE has an SNR gain of 1.2462 relative to the conventional DFE whereas the $(2, 1)$-BDFE (or the $(2, 2)$-DDFE) has an SNR gain of 1.016 inspite of the CCS of the two algorithms being identical. Furthermore, the matched filter upper bound on the SNR gain relative to the DFE is met with equality by the Viterbi algorithm for the channel $G(z) = 1 + (1/8)z^{-1} - (31/64)z^{-2}$. It is given as $\eta_{VA} = 1.2502$. Notice that the $(3, 2)$-DDFE performs nearly as well as the Viterbi algorithm without involving any sort of trellis detection and it performs much better than the $(2, 1)$-BDFE.

[1] This work was supported by NSF Grant NCR-9406069.

References

[1] E. A. Lee and D. G. Messerschmitt, *Digital Communication*, 2nd Edition, Kluwer Academic Publishers: Boston 1994.

[2] D. Williamson, R. A. Kennedy and G. W. Pulford, "Block Decision Feedback Equalization," *IEEE Trans. Commun.* **40**:2, pp. 255-264 (February 1992).

Tree Search Algorithms for Self-Adaptive Maximum-Likelihood Sequence Estimation

Bernd-Peter Paris and Ali R. Shah[1]
Department of Elect. & Comp. Eng.
Center of Excellence in C3I,
George Mason University
Fairfax, VA 22030

Abstract — **The problem of implementing self-adaptive equalization algorithms in real-time is addressed. Self-adaptive equalization determines the transmitted sequence without using a training sequence. Simulation results for the self-adaptive tree search procedures based on Fano, stack and M-algorithm are presented.**

I. Introduction

Many problems in digital communications can be modeled by means of a discrete-time finite-state Markov process representing the signal which is observed in independent identically distributed noise. We are considering the case when the process parameters are unknown. We are investigating methods to exploit the structure and finiteness of the state space of the signal *to determine the most likely state sequence without resorting to a known training sequence.* We will refer to this approach as self-adaptive MLSE.

We will focus our attention on the special case of a discrete-time finite-state Markov process in which a sequence of equally likely symbols s_k drawn from an a discrete and finite alphabet $\mathbb{A}$ is input to a channel which introduces intersymbol interference in addition to white Gaussian noise. The coefficients θ_l, $l = 0, \ldots, L$ of the channel impulse response are assumed to be unknown but constant. The objective of our work is now to determine the most likely input sequence given the observed sequence v_k without knowledge of the channel coefficients.

II. The Self-Adaptive MLSE

In [4] we propose the metric for the self-adaptive MLSE

$$d(\mathbf{s}) = \|\mathbf{P}_s \mathbf{v}\|^2, \tag{1}$$

where $\mathbf{s}$ and $\mathbf{v}$ are vectors comprising the input symbols and observations, respectively. If $\mathbf{S}$ is an $(N+L) \times (L+1)$ matrix whose columns are shifted versions of $\mathbf{s}$ and P_s is projection matrix $P_s = \mathbf{S}'(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}$. Among all possible input sequences, we are looking for the one which maximizes the metric in (1). The optimal sequence is then the one which spans the signal sub-space containing the largest portion of the received signal.

This observation provides the basis for our adaptation of sequential tree search algorithms, originally developed for decoding of convolutional codes, to the problem of self-adaptive equalization. In particular, we consider adaptations of the Fano algorithm [2], the stack algorithm [3], and the M-algorithm [1].

As an illustrative example for our results, Figure 1 shows the results of a series of simulations with sequences of $N =$ 1000 antipodal bits and channels with $L = 3$ memory elements. The simulations indicate clearly that the proposed "self-adaptive" M-algorithm matches closely the performance of the optimum (Viterbi) search algorithm with known coefficients if the number of retained paths is chosen sufficiently large. It also demonstrates that at higher signal-to-noise ratios the required number of paths to be retained decreases. Similar results are obtained for the other sequential algorithms.

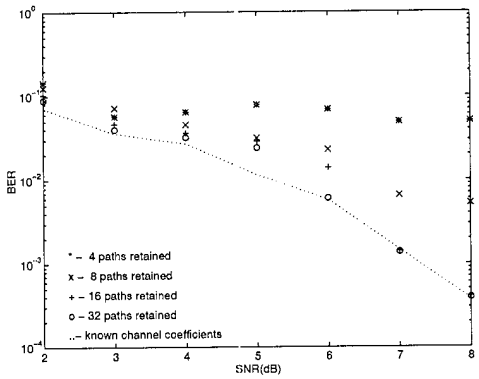


Figure 1: Simulation Results with M-algorithm ($N = 1000$, $L = 3$)

References

[1] J. B. Anderson, "Limited Search Trellis Decoding of Convolutional Codes," *IEEE Transactions on Information Theory*, vol. 35, September 1989.

[2] R. M. Fano, "A Heuristic Discussion of Probabilistic Decoding," *IEEE Trans. Inform. Theory*, vol. IT-9, pp. 64–67, April 1963.

[3] F. Jelinek, "A Fast Sequential-Decoding Algorithm Using a Stack," *IBM Journal of Res. and Dev.*, vol. 13, pp. 675–685, November 1969.

[4] B.-P. Paris and A. R. Shah, "Matched Subspace Detectors for Distinguishing between Two Signals with Extensions to Blind Maximum Likelihood Sequence Estimation," in *Proceedings of the 29-th Annual Conference on Information Science and Systems*, (Baltimore, MD), Johns Hopkins University, March 1995.

[1]Supported in part by the National Science Foundation under Grant NCR-9309044 and by Rome Laboratories under contract F30602-92-C-0053.

Modified/Quadrature Partial Response-Trellis Coded Modulation (M/QPR-TCM) Systems

Osman Nuri Uçan
Istanbul Technical University,Fac.of Elec.
Electr. Eng.,80626 Maslak,Istanbul,Turkey

Abstract:In this paper, to improve both bandwidth efficiency and error performance, partial response signalling (PRS) and trellis coded modulation (TCM) are combined together and denoted as Modified/ Quadrature Partial Response-Trellis Coded Modulation (M/QPR-TCM) for M-PSK. M/6QPR-TCM, M/9QPR-TCM and M/33QPR-TCM schemes are introduced for 4-PSK and 8-PSK respectively. In colored noise environment for negative noise correlation coefficient values M/QPR-TCM schemes outperform better than the classical structures. In fading channel, the proposed schemes are better than their counterparts for SNR values greater than a threshold. In terms of spectral efficiency and bit error rate with decreasing fading parameter K values, M/QPR-TCM systems appear to be the best choise in the literature.

Summary

The block diagram of the M/QPR-TCM scheme consists of k number unit memory precoders followed by $k/k+1$ rated convolutional encoder with ν units memory and $(1+D)$ PRS with $k+1$ units memory which represents the binary correspondent of the previous signal. K number appropriate precoders are included into the system to prevent the undesired catostrophic nature of the partial response block. The precoders do not increase the number of trellis states because of equivalence of the signals stored simultaneously in the delay cells of the coders and PRS. M/QPR-TCM scheme reduces the state number of the combined trellis structure from $2^k 2^\nu 2^{k+1}$ resulting from k-number precoder, ν unit convolutional encoder and (k+1) units (1+D) PRS memory to only $2^{\nu+k}$. In this paper, to give practical examples, M/6QPR-TCM, M/9QPR-TCM are introduced for 4-PSK with encoder memory $\nu = 1$ and $\nu = 2$ respectively and M/33QPR-TCM for 8-PSK with encoder memory $\nu = 3$.

For many practical trellis coded systems where the noise is not white, correlation between noise samples affects error performance [1]-[2]. M/QPR-TCM systems perform better than the related schemes for negative noise correlation coefficients.

Under the assumption of ideal channel state information and infinite interleaving/deinterleaving [3]-[4], analytical bit error probability upper bounds of the considered schemes are derived and compared to the related modulation systems in fading channels. M/QPR-TCM structures are better than their counterparts for SNR values greater than a threshold for small values of fading parameter K. In Rayleigh fading ($K = 0$) M/6QPR-TCM performs better after the SNR values of 9.2 dB . Similarly, error performance improvement of M/6QPR-TCM occurs at 9.5 dB for K=5 dB. As K increases, where AWGN starts to dominate the fading, the performance of the M/6QPR-TCM scheme tends to decrease. In Rayleigh fading, M/9QPR-TCM outperforms better at SNR values greater than 12 dB. This improvement begins at 15 dB for Rician (K=5 dB) fading and diminishes completely for AWGN as usual.

M/QPR-TCM systems appear to be the best choice in the literature in terms of spectral efficiency and bit error rate with decreasing K values.

References

[1] O.N.Uçan,"Performance analysis of quadrature partial response-trellis coded modulation (QPR-TCM) schemes",*Dissertation Thesis* 1995.

[2] O.N.Uçan, Ü.Aygölü and E.Panayırcı,"Performance and jitter analysis of quadrature partial response/trellis coded modulation (QPR-TCM) signals in the presence of intersymbol interference and colored noise",*IEEE JSAC.* 1992 Vol.10, No.8, pp.1264-1270.

[3] O.N.Uçan, Ü.Aygölü and E.Panayırcı,"Error performance analysis of quadrature partial response trellis modulation (QPR-TCM) in fading mobile satellite channel", In Proc. 26 th Ann. Conf. Inform. Sc. and Syst. NJ.Mar.1992,pp.517-521.

[4] E.Panayırcı,Ü.Aygölü and O.N.Uçan,"Error performance analysis of quadrature partial response trellis modulation (QPR-TCM) in fading mobile satellite channels",*IEEE Trans. on Comm.* to be appeared in 1995.

Soft Decoding Employing Algebraic Decoding Beyond e_{BCH}

Jan E.M. Nilsson
National Defence Research Establishment,
S-581 11 Linköping, Sweden, E-mail: jann@lin.foa.se

Abstract — **Soft decoding of binary codes based on algebraic decoding is treated. The algebraic decoder generates all error patterns up to a given weight higher than the designed error correcting capability e_{BCH}. The performance of different soft decoders employing such algebraic decoding is investigated and compared with hard decoding, the Chase second algorithm and soft maximum likelihood decoding. Furthermore, we propose an iterative decoder using an acceptance criteria to determine if we have found the maximum likelihood decision estimate. The acceptance criteria ensures a low average decoding complexity and by iterative decoding performance close to that of maximum likelihood decoding is obtained.**

I. Introduction

The type of soft decoders we consider can be described in the following way. In the first step, the demodulator outputs an estimate on what was received and it may also output reliability information on that estimate. In the second step, the estimate is decoded with an algebraic decoder (with or without help of reliability information) into a set of tentative codewords. Finally, the decoder selects as a decision the codeword "closest" to the received sequence with respect to Euclidean metric. Two well-known decoders of this type are proposed in [1] and [2]. For a given code the performance of the soft decoder depends on the reliability information used, the algebraic decoders efficiency in finding tentative codewords and the decision strategy.

The central problem is how to efficiently generate a set of code words such that it contains the maximum likelihood decision (MLD) estimate of the transmitted codeword with high probability. Also, it is desirable to find the MLD estimate of the transmitted codeword as soon as possible. When can the generation of tentative codewords be stooped? That is, when is the codeword corresponding to the maximum likelihood decision in the set of codewords already found?

II. The decoder

For a given code let d_{min} and e_{BCH} denote the minimum Hamming distance and the designed error correcting capability respectively. The algebraic decoder we use finds all error patterns of weight at most $t+\epsilon$ ($\epsilon > 0$), where $2t+1 = d_{min}$ and $t = e_{BCH}$; see [3]. Our soft algebraic decoder selects as decision the "best" codeword, in terms of Euclidean metric, among all tentative codewords found. We note that as long as the covering radius is less or equal to $t+\epsilon$ at least one codeword is found.

III. Results

We have compared different strong versions of our decoder with hard decoding (t error correction), the Chase second algorithm and a lower bound for soft maximum likelihood decision (MLD) decoding. In the evaluation (simulations) we consider: binary BCH codes, at most $(t+2)$-error correction, transmission over the additive white Gaussian noise channel (AWGN), and binary antipodal modulation. Our results show that decoding up to the covering radius is important, i.e., such that at least one codeword is found. Then, at least for the cases we have considered, the soft algebraic decoder performs better than the Chase second algorithm.

IV. Further improvements

If performance close to that of soft MLD decoding is desired we propose to use an iterative decoder employing MLD estimate tests. That is, a test which can determine if we have found the MLD estimate. From a practical point of view iterative decoding is probably a better option than generating error patterns of weight much higher than t. That is, generating such error patterns is complicated, very many may exist and the decoder has to be designed for the worst case. On the other hand, such error patterns seldom have to be considered if an MLD estimate test is used.

The proposed MLD tests are based on comparing with a competing word which is "close" to the received word. Related tests for $t-$error correction can be found in the literature. However, our tests are developed for a decoder correcting more than t errors. This makes the test more efficient.

When the codeword tested is not the MLD estimate the MLD estimate will hopefully be close to the competing word. We show that this often is the case. Then as a second decoding attempt, when the MLD estimate test fails, we decode the competing word. We can continue and generate a second competing word and perform a third decoding attempt and so on. Important is that the algebraic decoder corrects up to the covering radius in Hamming metric. Such an algebraic decoder ensures that at least a fairly good estimate of the transmitted codeword is found already in the first decoding attempt.

We have investigated two versions of iterative decoding, at most two decoding attempts and at most three decoding attempts. In both versions, however, due to the MLD estimate tests, the average number of decoding attempts is close to one. For the cases we have investigated the iterative decoder is much more powerful than the Chase second algorithm and its performance is close to that of soft MLD decoding.

References

[1] C. D. Forney, "Concatenated Codes," *The M.I.T. Press.*, Cambridge, Massachusetts 1966.

[2] D. Chase, "A Class of Algorithms for Decoding Block Codes With Channel Measurement Information," *IEEE Trans. Info. Theory*, IT-18, No.1, pp.170-182, January, 1972.

[3] J.E.M. Nilsson, "On Hard and Soft Decoding of Block Codes," *PhD Thesis (dissertation No. 333)*, Linköping Studies in Science and Technology, Linköping, Sweden, 1994.

The Algebraic Decoding of the Z_4-Linear Goethals Code

Tor Helleseth and P. Vijay Kumar[1]

Department of Informatics, University of Bergen, Høyteknologisenteret, N-5020 Bergen, Norway and Communication Sciences Institute, EE-Systems, EEB 534, University of Southern California, Los Angeles, CA 90089-2565, USA

Abstract — **The quaternary Goethals code is a Z_4-linear code of length 2^m which has $2^{2^{m+1}-3m-2}$ codewords and minimum Lee distance 8 for any odd $m \geq 3$. The Gray map of this code is known to be a nonlinear binary $(2^{m+1}, 2^{2^{m+1}-3m-2}, 8)$ code. The covering radius of the Z_4-linear Goethals code is 6 and we present a complete decoding algorithm for the code.**

I. Introduction

Let Z_4 denote the ring of integers modulo 4 and let R be a Galois ring of characteristic 4 with 4^m elements. The multiplicative group of units in R contains a unique cyclic subgroup of order $2^m - 1$. Let β be a generator of this subgroup and let $\mathcal{T} = \{0, 1, \beta, \cdots, \beta^{2^m-2}\}$. Let $\mu : Z_4 \to Z_2$ denote the modulo 2 reduction map. We can extend μ to R in a natural way and it can be shown that $\mu(\mathcal{T}) = F$, where F is a finite field of order 2^m.

The Gray map ϕ is defined by $\phi(0) = 00$, $\phi(1) = 01$, $\phi(2) = 11$ and $\phi(3) = 10$. Let C be the binary code defined by $C = \phi(\mathcal{C})$, where $\mathcal{C}$ is the quaternary code with parity-check matrix given by

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & \beta & \beta^2 & \cdots & \beta^{2^m-2} \\ 0 & 2 & 2\beta^3 & 2\beta^6 & \cdots & 2\beta^{3(2^m-2)} \end{bmatrix}.$$

In Hammons, Kumar, Calderbank, Sloane and Solé [1], it is shown that if m is odd, then $\mathcal{C}$ has minimum Lee distance 8 which is equal to the minimum Hamming distance of C. The binary $(2^{m+1}, 2^{2^{m+1}-3m-2}, 8)$ code C has parameters that are identical to the (extended) binary Goethals code.

The purpose of this paper is to give a complete decoding algorithm for the triple error-correcting Z_4-linear Goethals code $\mathcal{C}$, i.e., an algorithm that for any received vector finds the closest codeword.

II. Decoding of the Goethals code

Let $\mathbf{r} \in Z_4^{2^m}$ be the received vector and let $\mathbf{e} \in Z_4^{2^m}$ be the error vector. The syndrome of the received vector is $\mathbf{S} = \mathbf{r}H^{tr} = \mathbf{e}H^{tr} = (t, A+2B, 2C)$ where $t \in Z_4$, $A, B, C \in \mathcal{T}$ and H^{tr} denotes the transpose of H. We index the components of a vector $\mathbf{e} \in Z_4^{2^m}$ by the elements of $\mathcal{T}$, i.e. $\mathbf{e} = (e_X)_{X \in \mathcal{T}}$. The syndrome equations that have to be solved are

$$\begin{aligned} \sum_{X \in \mathcal{T}} e_X &= t, \quad t \in Z_4 \\ \sum_{X \in \mathcal{T}} e_X X &= A + 2B, \quad A, B \in \mathcal{T} \\ 2\sum_{X \in \mathcal{T}} e_X X^3 &= 2C, \quad C \in \mathcal{T}. \end{aligned}$$

Let X, Y, A, B, etc. denote elements in $\mathcal{T}$ and x, y, a, b their respective projections modulo 2 in F. For any coset it is sufficient to find the projections x, y, and z in F of the error locations of a coset leader and the corresponding error values e_X, e_Y, and e_Z in Z_4 which satisfy the syndrome equations.

We first find the unique coset leader (i.e., a vector of smallest Lee weight) of each coset which contains a vector of Lee weight ≤ 3. As an example the decoding of cosets corresponding to syndromes with $t = 1$ are given below. The cases $t = 0, 2$ and 3 are similar.

Theorem 1 Let $\mathbf{S} = (1, A + 2B, 2C)$ denote the syndrome of a coset.

(i) If $b = 0$ and $c = a^3$, then the coset leader has Lee weight 1 and is uniquely determined by $x = a$ and $e_X = 1$.

(ii) If $b \neq 0$ and $c = a^3$, then the coset leader has Lee weight 3 and is uniquely determined by $x = a + b$, $e_X = 2$, $y = a$ and $e_Y = -1$.

(iii) If $b \neq 0$, $c \neq a^3$ and $Tr(b^3/(a^3 + c)) = 0$, then the coset leader has Lee weight 3. The coset leader is uniquely determined such that x and y are solutions of $b^2u^2 + (a^3 + c)u + a^4 + a^2b^2 + ac + b^4 = 0$, $e_X = e_Y = 1$, $z = a + \frac{a^3+c}{b^2}$ and $e_Z = -1$.

(iv) If $\sigma(u) = u^3 + au^2 + (a^2 + b^2)u + ab^2 + c$ has three distinct zeros in F then a coset leader has Lee weight 3 and is uniquely determined such that x, y, z are the three distinct zeros in F of $\sigma(u)$ and $e_X = e_Y = e_Z = -1$.

(v) If none of (i)-(iv) hold, then any coset leader has Lee weight ≥ 5.

III. Complete decoding

In the considerably more complicated cases when more than 3 errors occur we show how to construct a coset leader in any coset. In addition we proved the following results.

Theorem 2 (i) For any coset with syndrome $\mathbf{S} = (0, A + 2B, 2C)$, there exists a coset leader of Lee weight ≤ 6.

(ii) For any coset with syndrome $\mathbf{S} = (t, A+2B, 2C)$ where $t = 1$ or $t = 3$, there exists a coset leader of weight ≤ 5.

(iii) Let $m \geq 5$, then for any coset with syndrome $\mathbf{S} = (2, A + 2B, 2C)$ there exists a coset leader of weight ≤ 4.

Theorem 3 Let D_i denote the number of cosets with a coset leader of weight i in the Z_4-linear Goethals code.

(i) If $m \geq 5$ then $D_0 = 1$, $D_1 = \binom{2^{m+1}}{1}$, $D_2 = \binom{2^{m+1}}{2}$, $D_3 = \binom{2^{m+1}}{3}$, $D_4 = 2^{3m+1} - 1 - \binom{2^{m+1}}{2} - (2^m - 1)\frac{2^m+4}{3}$, $D_5 = 2^{3m+1} - \binom{2^{m+1}}{1} - \binom{2^{m+1}}{3}$ and $D_6 = (2^m - 1)\frac{2^m+4}{3}$.

(ii) If $m = 3$ then $D_0 = 1$, $D_1 = 16$, $D_2 = 120$, $D_3 = 480$, $D_4 = 823$, $D_5 = 528$ and $D_6 = 80$.

References

[1] R. Hammons, P.V. Kumar, N.J.A. Sloane, R. Calderbank and P.Solé, The Z_4-Linearity of Kerdock, Preparata, Goethals, and Related Codes, IEEE Trans. on Inform. Theory, vol. 40, pp. 301-319, 1994.

[1]This work was supported in part by The Norwegian Research Council under Grant Numbers 107542/410 and 107623/420 and the National Science Foundation under Grant Number NCR-9016077

The Welch–Berlekamp and Berlekamp–Massey Algorithms

Simon R. Blackburn[1]

Department of Mathematics, Royal Holloway,
University of London, Egham, Surrey TW20 0EX, U.K.

Abstract — **We show that the problems solved by the Berlekamp–Massey and Welch–Berlekamp algorithms are special instances of a more general problem which has been studied (in the characteristic zero case) by control theorists. We present an algorithm to solve this general problem which can be used to find the solutions to both the classical Key Equation and the Welch–Berlekamp interpolation problem.**

Summary

Classically, the decoding of a Reed–Solomon code is carried out by calculating power sum syndromes and then using the Berlekamp–Massey algorithm to solve the resulting linear recurrence problem [2, 4]. A new approach, taken by Welch and Berlekamp [5], is to convert the decoding problem into a rational interpolation problem which can then be solved by the Welch–Berlekamp algorithm. One of the advantages of this second approach is that the syndromes do not have to be calculated, thus saving decoder computations.

Both the Berlekamp–Massey and Welch–Berlekamp algorithms can be thought of as solving special instances of the following problem.

The Problem: Let F be a field. If $f(X) = P(X)/Q(X)$ is a rational function of two polynomials with coefficients in F, we define the complexity $\lambda(f)$ of f to be the integer $\max\{(\deg P(X)) + 1, \deg Q(X)\}$. Let $x_0, x_1, \ldots, x_{m-1} \in F$ be distinct. For each $i \in \{0, 1, \ldots, m-1\}$, let l_i be a nonnegative integer and let $y_{i,0}, y_{i,1}, \ldots, y_{i,l_i} \in F$. We say that the function $f(X) := P(X)/Q(X)$ is a generalised rational interpolation if, for all $i \in \{0, 1, \ldots, m-1\}$, the formal power series of f at x_i is defined and is of the form

$$\sum_{j=0}^{l_i} y_{i,j}(X - x_i)^j + \text{higher terms.}$$

The **generalised rational interpolation problem** asks for the generalised rational interpolation f of lowest complexity $\lambda(f)$.

Thus the generalised rational interpolation problem asks for the 'smallest' rational function which has specified low order terms in its power series expansion at certain points. The Welch–Berlekamp interpolation problem is the special case of this problem when $l_i = 0$ for all i (since $y_{i,0}$ is simply the value of f at x_i). The problem solved by the Berlekamp–Massey algorithm can be thought of as the case when $m = 1$ and $x_0 = 0$.

When F has characteristic zero, there is a close relationship between formal power series and Taylor series: We may regard the generalised rational interpolation problem as asking for the lowest complexity rational function with specified low order derivatives at certain points. This is a problem in control theory studied by Antoulas and Anderson [1]. So we can regard the problem above as generalising their problem to fields of arbitrary characteristic.

We present a new algorithm (a close analogue of the 'Welch–Berlekamp' algorithm of Chambers *et al* [3]) which solves the generalised rational interpolation problem. Like the Berlekamp–Massey algorithm, the data can be fed into our algorithm serially. The algorithm uses $O(n^2)$ field operations, where $n = \sum_{i=0}^{m-1}(l_i + 1)$. The algorithm can be used in place of the Berlekamp–Massey or Welch–Berlekamp algorithms, since both problems are special cases of the generalised rational interpolation problem.

References

[1] A.C. Antoulas, B.D.Q. Anderson, 'On the Scalar Rational Interpolation Problem', *IMA J. Math. Control and Inform.* Vol 3 (1986), pp. 61-88.

[2] E.R. Berlekamp, *Algebraic Coding Theory*, (McGraw–Hill, New York, 1968).

[3] W.G. Chambers, R.E. Peile, K.Y.Tsie, N.Zein, 'Algorithm for Solving the Welch–Berlekamp Key Equation, with a Simplified Proof', *Elect. Letters* Vol 29 No 18 (September 1993), pp. 1620-1621.

[4] J.L. Massey, 'Shift Register Synthesis and BCH Decoding', *IEEE Trans. Inform. Theory* Vol. IT-15 (January 1969), pp. 122-127.

[5] L. Welch, E.R. Berlekamp, 'Error Correction for Algebraic Block Codes', U.S. Patent 4 633 470, September 1983.

[1]The author was supported by E.P.S.R.C. research grant GR/H23719.

The Optimal Erasing Strategy for Concatenated Codes

Yi Hsuan[1], John T. Coffey[1], and Oliver M. Collins[2]

[1]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, U.S.A.
[2]School of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland, U.S.A.

Abstract — **We consider the optimal strategy for erasing symbols in concatenated coding schemes. This erasing strategy uses *a posteriori* likelihoods of RS symbols to determine erasures which maximize the probability of decoding correctly. Some properties of performance of this strategy are presented. Erasing rules for decoding the same received word more than once are also examined.**

I. Introduction

Various kinds of erasing strategies [1] have been explored by many researchers since Forney first presented concatenated coding schemes. We are interested in the erasing strategy which maximizes the probability of decoding an outer word correctly given *a posteriori* likelihoods of all RS symbols. The first stage of the optimal erasing rule [2] is to erase the most unreliable symbol if D, the minimum distance of the RS code, is even and erase nothing if D is odd. Then symbols should be erased in pairs in order of ascending reliability, thus keeping the difference between D and the number of erasures odd. Let $p_1, p_2, \cdots, p_N$ be the error probabilities of the symbols provided by the inner decoder in a RS word, and $p_1 \le p_2 \le \cdots \le p_N$. $P_h(\mathbf{p}_i)$ denotes the probability that h errors occur in the i most reliable symbols. Given that the most unreliable e symbols have been erased in a RS word, erasing the next two symbols, with error probabilities p_{N-e} and p_{N-e-1}, can increase the probability of decoding correctly if and only if

$$\frac{P_{h-1}(\mathbf{p}_j)}{P_h(\mathbf{p}_j)} > \frac{q_{j+1}q_{j+2}}{p_{j+1}p_{j+2}} \qquad (1)$$

where $j = N-e-2$, $h = (D-e-1)/2$, and $q_i = 1-p_i$ for $1 \le i \le N$.

However if a decoder erases symbols in pairs until (1) fails to hold, the resulting probability of decoding correctly is not necessarily the maximum obtainable. Here we present different approaches to simplify the search for the optimal number of erasures by exploring bounds on $P_{h-1}(\mathbf{p}_j)/P_h(\mathbf{p}_j)$.

II. The Optimal Number of Erasures

This first problem encountered is how to evaluate $P_h(\mathbf{p}_j)$. Although it can be calculated exactly, an easily-calculated estimate is preferred. Barbour [3] derived an asymptotic expansion for $P_h(\mathbf{p}_j)$. This allows us to have an approximation to the left hand side of (1) and a bound on the error of approximation . However as more accuracy is required, more terms in the expansion should be included in the approximation and the complexity increases dramatically.

The RHS of (1) decreases with j while the LHS of (1) is not necessarily increasing with j. If the LHS of (1) is increasing with j, apparently the probability of decoding correctly has only one local maximum with respect to numbers of erasures. We show that the LHS of (1) increases with j if $p_{N-D+1} = p_{N-D+2} = \cdots = p_N$. Given that there are two different symbol error probabilities, p_{low} and p_{high}, among all N symbols, we also show that the LHS of (1) increases with j if p_{low} and p_{high} satisfy an inequality. Basically this inequality gives an upper bound on p_{high} in terms of p_{low}.

The derivative of $P_{h-1}(\mathbf{p}_j)/P_h(\mathbf{p}_j)$ with respect to p_i is always non-negative, $1 \le i \le j$. This observation enables us to find several upper and lower bounds on the optimal number of erasures. An upper bound on $P_{h-1}(\mathbf{p}_j)/P_h(\mathbf{p}_j)$ is $\frac{q_1}{p_1}\frac{h}{j-h+1}$. Since this bound increases with j, given a fixed integer $m \in S = \{N-D+1, N-D+3, \cdots, N-2\}$, the optimal number of erasures of the RS word is not more than $N-m-2$ if $\frac{q_1}{p_1}\frac{h}{m-h+1} \le \frac{q_{m+1}q_{m+2}}{p_{m+1}p_{m+2}}$. Note that if the inequality holds for some m, changing $p_2, \cdots, p_m$ arbitrarily can not increase the optimal number of erasures to more than $N-m-2$. A drawback of this bound is that it depends solely on p_1 and becomes very loose when p_1 is small compared to other p_i's. Two more upper bounds can be obtained to fix this problem. One bound is based on $p_{j-h+1}, \cdots, p_j$ instead of p_1. The other bound is based on p_1 and p_a, the average of $p_1, \cdots, p_j$. Examples show that applying these three upper bounds of $P_{h-1}(\mathbf{p}_j)/P_h(\mathbf{p}_j)$ often gives very tight upper bound on the optimal number of erasures. Similar approaches can be used to find lower bounds also. Experiments show that upper and lower bounds meet very often.

III. Results for Multiple Decodings

Assume that the first decoding uses the optimal erasing strategy described above. If the first decoding fails to decode, we discuss the best erasing rule that the second decoding should use. Here we show that the probability of decoding correctly when $e-i$ symbols are erased is always larger than that when $e+i$ symbols are erased, where e is the number of erasures of the first decoding and all p_i's are less than one half. If the second decoding erases less symbols than the first decoding and all symbols erased by the first decoding have the same error probability, we show that the optimal number of erasures for the second decoding is either one or zero no matter what e is. Erasing rules for decoding more than twice are also discussed.

References

[1] T. Hashimoto, "Further results on the performance of erasure-and-error decoding rules," *Proc. 1994 IEEE Int. Symp. Inform. Theory*, Trondheim, Norway, 1994.

[2] Y. Hsuan, J. T. Coffey, and O. M. Collins, "Erasing gains for concatenated codes," to appear in *Proceedings of the IEE, Pt. I (Communications)*.

[3] A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, New York: Oxford University Press, 1992.

Error and Erasure Decoding of Binary Cyclic Code up to Actual Minimum Distance[1]

H. Lee
AT&T Bell Labs,
Allentown, PA. 18103

K. K. Tzeng
Dept. of EECS, Lehigh University,
Bethlehem, PA. 18105

C. J. Chen
Dept. of Comm., Northern Jiao-tong University,
Beijing, China 100044

Abstract - **A new error-and-erasure decoding procedure that decodes cyclic codes up to the actual minimum distance is presented. This procedure annihilates erasure effects from a syndrome matrices and produces modified syndrome matrices that can be used to obtain error locations with an error-only decoding algorithm.**

I. Introduction

This paper presents a new error-and-erasure decoding procedure that produces erasure masking matrices to annihilate erasure effects from original syndrome matrices.

In general, an error-and-erasure decoding procedure is based on error-only decoding algorithms. In [2], Forney, based on Peterson, Gorenstein and Zierler's earlier work [8], introduced an error-and-erasure decoding procedure that can decode up to the BCH bound. Later, in [6], Shahri and Tzeng developed an error-and-erasure decoding algorithm to decode cyclic codes up to the HT bound. The procedure in [6] uses Feng and Tzeng's *algorithm for Multi-sequence Shift-Register Synthesis* [12]. Then, an error-and-erasure decoding procedure up to special cases of the Roos bound was given by Shahri, Tzeng and Jensen[10].

Recently, Feng and Tzeng introduced algorithms for error-only decoding of cyclic codes up to the actual minimum distance[1,9]. The algorithm in [1] uses the *nonrecurrent syndrome dependence relations* among the known syndromes. In [9], they determined the unknown syndromes by employing a (2t+1) x (2t+1) syndrome matrix and *majority voting method*. The error-and-erasure decoding procedure presented in this paper is based on Feng and Tzeng's recent work [1,9].

II. Decoding Procedure

The procedure presented in this paper generates erasure masking matrices which annihilate all erasure effects in a syndrome matrix. Thus, it converts an error-and-erasure decoding problem to an error-only decoding problem. Furthermore, since it produces modified syndrome matrices which are homomorphic images of the original syndrome matrices, error-only decoding algorithms, ex. Feng and Tzeng's algorithms [1,9], can be applied.

A brief description of our decoding procedure is given below:

Step 1. Construct a syndrome matrix S just as for any error-only decoding case.

Step 2. Partition the ρ erasure locations into two arbitrary groups, say G_1 and G_2, where $G_1 = (\alpha^{l_1}, \alpha^{l_2}, \ldots \alpha^{l_k}) = (F_1, F_2, \ldots, F_k)$ and $G_2 = (\alpha^{l_{k+1}}, \alpha^{l_{k+2}} \ldots \alpha^{l_\rho}) = (F_{k+1}, F_{k+2}, \ldots, F_\rho)$, then F_i are the erasure locations and $k = \lfloor (\rho + 1)/2 \rfloor$

Step 3. From G_1 and G_2, construct erasure masking matrices, μ and Λ such that μ masks erasures in G_1, and Λ masks erasures in G_2

Step 4. Compute a modified syndrome, $U = \mu S\Lambda = \mu E\Lambda + \mu F\Lambda$, where E is the error portion of a syndrome matrix and F is the erasure portion of a syndrome matrix. Since μ and Λ matrices mask all erasures, $\mu F\Lambda = 0$ and $U = \mu S\Lambda = \mu E\Lambda$.

Step 5. Use an error-only decoding algorithm to find an modified error locator polynomial γ, such that $U\gamma = 0$.

Step 6. Obtain the coefficients for an error locating polynomial f(z) from $\Lambda\gamma = f$.

Step 7. Use the Chien search to fine the roots of f(z).
If the number of *n*th root of unity roots in f(z) is less than (d -1)/2, then all the error locations are found.
If not, go to step 8.

Step 8. Compute modified unknown syndromes using error-only decoding algorithms presented in [1] or [9]. Then, find the values of unknown syndromes from the computed modified unknown syndromes.

Step 9. If all unknown syndromes can be found, obtain a codeword by means of *Inverse Fourier Transformation.*

Step 4 yields a modified syndrome matrix U which is a homomorphic image of the syndrome matrix in step 1. Thus, error-only decoding algorithm in step 1 can be applied to matrix U to solve for error locations.

In summary, we developed an efficient systematic error-and-erasure decoding procedure using erasure masking matrices μ and Λ that can be applied to any type of syndrome matrix. Therefore, our procedure can be used with any error-only decoding algorithm as long as it uses a syndrome matrix.

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. 37, pp. 1716-1723, Nov. 1991.

[2] G. D. Forney, Jr., "On decoding BCH codes," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 549-557, Oct. 1965.

[3] C. R. P. Hartmann and K. K. Tzeng, "Generalizations of the BCH bound," *Inform. Contr.*, vol. 20, pp. 489-498, 1972.

[4] C. Roos, "A new lower bound for the minimum distance of a cyclic code," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 330-332, May 1983.

[5] P. Stevens, "Error-erasure decoding of binary cyclic code up to a particular instance of the Hartmann-Tzeng bound," *IEEE Trans. Inform. Theory*, vol. 36, Sept. 1990.

[6] H. Shahri and K. K. Tzeng, "On error-and- erasure decoding of cyclic code," *IEEE Trans. Inform. Theory*, vol. 38, pp. 489-496 March 1992.

[7] R. E. Blahut, *Theory and Practice of Error Control Codes*. New York: Addison-Wesley, 1983.

[8] W. W. Peterson and E. J. Weldon, Jr., *Error Correcting codes*, 2nd ed. MIT press, 1971.

[9] G. L.Feng and K. K. Tzeng, "A new procedure for decoding cyclic and BCH codes up to actual minimum distance," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1364-1374 March 1992.

[10] H. Shahri, K. K. Tzeng and J. Janssen, "On error-and-erasure decoding of cyclic codes up to the Roos Bound," 1991 *IEEE International Symposium on Information Theory*, Budapest, Hungary, June 1991.

[11] I. Duursma and R. Kotter, "Error-locating pairs for cyclic codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1108-1121 July 1994.

[12] G. L.Feng and K. K. Tzeng, "A generalization of the Berlekamp-Massey algorithm for multi-sequence shift-register synthesis with applications to decoding cyclic codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1274-1287 Sept. 1991.

[1]This work was supported in part by the National Science Foundation under Grant NCR-9016095 and 9406043.

Enhanced Decoding of Interleaved Error Correcting Codes

Mario Blaum†
Henk C. A. van Tilborg‡

†IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA
‡Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract — **A common way to deal with bursts in data storage systems is to interleave byte-error correcting codes. In the decoding of each of these byte-error correcting codes, one normally does not make use of the information obtained from the previous or subsequent code, while the bursty character of the channel indicates a dependency.**

In [1] such a dependency is exploited to enhance the decoding performance. Here a different, but similar approach is proposed.

I. Introduction

In order to correct burst errors in a data storage channel, the most common procedure is to interleave an error-correcting code to a certain depth. This error correcting code is normally a byte-error correcting code, like a Reed-Solomon code. The depth of interleaving determines the burst correcting power of the interleaved scheme. In this way, the bursts are "randomized" into different codewords. Each codeword sees a random error event.

Although interleaving is an efficient approach, it throws away information, since it ignores the fact that in a bursty channel errors are usually correlated. Ways of exploiting this correlation in order to forecast errors were studied in the literature [1] with the introduction of the so called "helical" interleavers. Here, we introduce a different, but somewhat similar technique. Even when the error-correcting capability of the error correcting code has been exceeded, one can still, by making use of these methods, retrieve the data in many cases.

II. A general description

In its most basic form, the procedure works as follows: the decoder decodes normally using the interleaved scheme. However, if a codeword is uncorrectable due to too many errors, it is flagged. Then, an attempt to decode it again using the previous and/or following codeword is made. To this end, the decoder declares erasures in the locations corresponding to errors in the previous and/or following codeword. If errors have occurred in bursts, it is likely that the decoding power will be enhanced, since a code can correct roughly twice as many erasures as errors. We will present several variations of this strategy, that trade reliability with decoding power. We will show how to adapt the method to channels that suffer from bursts as well as from random errors at the same time. We will also introduce a toroidal interleaving method that eliminates the lack of symmetry between the first and the last codeword in a regular interleaving scheme.

The toroidal scheme works as follows: if λ is the depth of interleaving and n is the length of a codeword, such that λ and n are relatively prime, then symbol $a_{i,j}$ is followed by symbol $a_{i+1,j+1}$, where $i+1$ is taken modulo λ and $j+1$ is taken modulo n (in normal interleaving, symbol $a_{i,j}$ is followed by symbol $a_{i+1,j}$ when $0 \leq i \leq \lambda-2$ and symbol $a_{\lambda-1,j}$ is followed by symbol $a_{0,j+1}$).

III. An example

Below is a simple example of the enhanced interleaved scheme when $\lambda = 5$ and $n = 11$. Assume that each row implements a code with minimum distance $d = 4$, therefore it can correct one error and detect two, as well as an error and an erasure. Also, assume the toroidal interleaving described above.

		x								
			x			x				
				x						
					x					

The x's represent errors in the corresponding symbols. As we can see, a burst of length 4 and a random error (in row 2) have occurred. The decoder detects an uncorrectable error pattern in row 2, so it flags that row. By examining row 1, it finds that entry (1,2) is in error, and similarly, by examining row 3, it finds that entry (3,4) is in error. Therefore, the decoder will predict that there was an error in entry (2,3), so it will declare an erasure there. Now, row 2 has an error and an erasure, which is within its error-correcting capability. Finally, the decoder corrects row 2. Notice that this was not possible with the traditional scheme.

Details of the implementation can be found in [2].

References

[1] E. Berlekamp and P. Tong, US Patent 4,559,625, Dec. 1985.

[2] M. Blaum and H. C. A. van Tilborg, US Patent 5,299,208, March 1994.

A decoding algorithm for linear codes overs Z_4

Manish Goel and B.Sundar Rajan
EE Department, Indian Institute of Technology, New Delhi, India. E-mail: bsrajan@ee.iitd.ernet.in

Abstract — **A decoding algorithm for linear codes over $Z_4 = \{0, 1, 2, 3\}$, the ring of integers modulo 4, is given which gives the codewords that is closest to the received vector in Lee distance.**

I. Introduction

A linear code over Z_4 is same as a group code over the 4-element cyclic group and can be defined by a check-matrix [1]. The algorithm proposed is similar to the one given in [2] for soft-decision decoding of binary linear codes. First a trellis is cons tructed using the check matrix of the linear code over Z_4 under consideration using Wolf's trellis construction[3]. There is one to one correspondence between the set of paths from the start node to the goal node and the set of codewords. Hence, the pr oblem of decoding is same as finding the path in the trellis which is closest to the received vector in Lee distance. The search is guided by an evaluating function $f = g + h$ defined on each node, where g depends only on the past and h(called heuristic func tion) is an estimate on the set of possible futures. The nodes with minimum value of f is given the first priority for expanding. The most important factor in the efficeincy of the algorithm depends on the complexity of the heuristic function. We define a heuristic function which can be easily computed with the worst case complexity of 4n searches over Lee weight distribution , where n is the length of the code.

II. Heuristic function 'h' and cost function 'f'

Let $r = (r_0, r_1, \ldots, r_{n-1})$ be the received vector. A cost function $f(m,t)$ for any node m at level t, $(0 \leq t \leq n-1)$ is defined by

$$f(m,t) = g(m,t) + h(m,t) \tag{1}$$

where $g(m,t)$ and $h(m,t)$ are defined as follows

$$g(m,t) = \sum_{i=0}^{t} LW(r_i - c_i) \tag{2}$$

where $LW(x)$ = Lee weight of x and $c_p(t) = (c_0, c_1, \ldots, c_t)$ is the path leading to that node and

$$h(m,t) = \min_{x \in X(t)} \left(\sum_{i=t+1}^{n-1} LW(r_i - x_i) \right) \tag{3}$$

where
X(t)=$\{\underline{x} = (c_0, c_1, \ldots, c_t, x_{t+1}, \ldots, x_{n-1})/x_{t+1}, \ldots, x_{n-1} \in Z_4, LW(\underline{x}) \in L_s\}$ and L_s is the set of all Lee weights of the codewords. The decoding algorithm given below gives the codeword which is closest to the received vector in Lee distance.

III. The Decoding Algorithm

Step 1: Create a list called OPEN and let start node be the only element in OPEN.

Step 2: Select and remove the first node from OPEN and call it node m. If m is the goal node exit successfully, and the path history of node m is the output of the decoder.

Step 3: Expand node m, generating next level nodes which are successors of the node m. This expanding operation consists of

- **(a)** Obtaining all successor nodes and computing g and h values of all the successor nodes.
- **(b)** For each of the successor storing the path followed so far (called path history) from the start node.
- **(c)** Storing all successors in OPEN. *em[(d)] Arranging the nodes in OPEN in the increasing order of their f value. (For nodes with equal value of f arrange them in the decreasing order of the levels of the nodes. For nodes with equal values of f and in the same level arrange in the increasin g order of their g value.)*

Step 4: Go to Step 2.

IV. A simple procedure to calculate $h(m,t)$

The following theorem leads to a simple procedure which gives the value of h(m,t) without actually carrying out the minimization.

Theorem 1 *For a chosen node m, which is say at level t, let* $\underline{c_t} = (c_0, c_1, \ldots, c_t)$, *and* $l_c = LW(\underline{c_t})$. *Also let* $\underline{r_t} = (r_{t+1}, \ldots, r_{n-1})$, *and* $l_r = LW(\underline{r_t})$. *Then,* $h(m,t) = |h^*|$, *(absolute value of* h^**) wh ere* h^* *is the least integer such that* $l_c + l_r \pm h^* \in L_s$

For any node m, the possible values for h(m,t) are 0,1,2,...,2(n-t-1). From Theorem 1, it follows that one can find h(m,t) for each node, by successively assuming values from 0 to 2(n-t-1) and matching with elements of L_s to check whether $l_c + l_r$ $pmh(m,t) \in L_s$ and stopping at the first value for which $l_c + l_r \pm h(m,t) \in L_s$ Clearly, the worst case for matching effort is for the start node for which the number of matching efforts may be 2n. The complexity of each search for matching depe nds on the Lee weight distribution of the code. If the code has codewords of specific weights only then the search becomes simple. For instance, for constant Lee weight codes the search is to test for that constant weight or zero. If minimum Lee weight is known then one checks only for zero and all weights starting from minimum Lee weight to four times the length of the code. In the absence of any knowledge of Lee weight distribution one is compelled to check for all weights from zero to twice the length of the code.

References

[1] G.Caire and E.Biglieri, "Linear Codes over Cyclic Groups", (Private Communication).

[2] Y.S.Hahn, C.R.P. Hartmann and C.C. Chen, "Efficient priority-first search m aximum-likelihood soft-decision decoding of linear block codes", IEEE Transactions on Information Theory, Vol. IT-35, No. 5, 1514-1523, 1993.

[3] J.K. Wolf, "Efficient maximum-likelihood decoding of linear block codes using a trellis", IEEE Transactions on Information Theory, Vol. IT-24, 76-80, 1978.

Decoding Linear Block Codes Using Optimization Techniques[1]

Ching-Cheng Shih, Christopher R. Wulff, Carlos R. P. Hartmann and Chilukuri K. Mohan
School of Computer and Information Science, Syracuse University, Syracuse, NY 13244-4100, USA

Abstract — **We present a new soft-decision decoding algorithm, *Modified A** (MA*), that conducts heuristic search through a code tree for a binary (n, k) linear code. MA* improves on the results obtained earlier using Algorithm A*. We also describe the application of the *simulated annealing* (SA) algorithm to the decoding problem, transformed into a continuous optimization problem.**

Summary

In MA*, 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. The algorithm maintains a list $\mathcal{L}$ of nodes of the code tree that are candidates to be expanded. The algorithm selects for expansion the node in $\mathcal{L}$ with minimum values of function f. If it selects a goal node for expansion, it has found an "optimal" path from the start node to the goal node whose labels correspond to a codeword that minimizes the error probability when we assume all codewords have equal probability of being transmitted. For every node m of the code tree visited by the algorithm, MA* keeps two values, $f(m)$ and $low_l(m)$, where l is a fixed non-negative integer, whereas A* keeps only one value $f(m)$ [1]; $low_l(m)$ is a new lower bound on the cost of an optimal path that goes through node m. This algorithm keeps an upper bound, UB, on the value of low_l for every node in an optimal path. If the value of low_l for a node is larger than or equal to UB, no further search through this node is necessary and the node can be discarded.

If no restriction is placed on the size of list $\mathcal{L}$, then the MA* decoding algorithm is an *maximum-likelihood soft-decision* (MLSD) decoding algorithm. In our *sub-optimal soft-decision* (SOSD) decoding algorithm, we limit the size of list $\mathcal{L}$ according to the following criterion. If a node m needs to be stored in list $\mathcal{L}$ when the size of list $\mathcal{L}$ has reached a given upper bound M_B, then we discard the node with larger f value between node m and the node in list $\mathcal{L}$ with the maximum f value.

To verify the performance of our SOSD decoding algorithm, we show simulation results for the $(104, 52)$ code and for the $(256, 131)$ code when these codes are transmitted over AWGN channels, with $M_B = 1000$ and $l = 4$. From Figure 1, for the $(104, 52)$ code the performance of our SOSD decoding algorithm is within 0.15 dB of the lower bound of the performance of the MLSD decoding algorithm. Thus, for the samples tried, limiting M_B to 1000 introduced only a small degradation on the performance of the algorithm. In Table 1, for the $(256, 131)$ code the results were obtained by simulating 35,000 samples. No decoding error occurred during simulation. For the examples tried, the average number of codewords constructed is insignificant compared with the total number of codewords. In Table 1, N(r) = number of nodes visited, C(r) = number of codewords constructed, M(r) = number of nodes stored in list $\mathcal{L}$, max = maximum value among samples tried, ave = average value among samples tried, and γ_b = SNR per transmitted information bit.

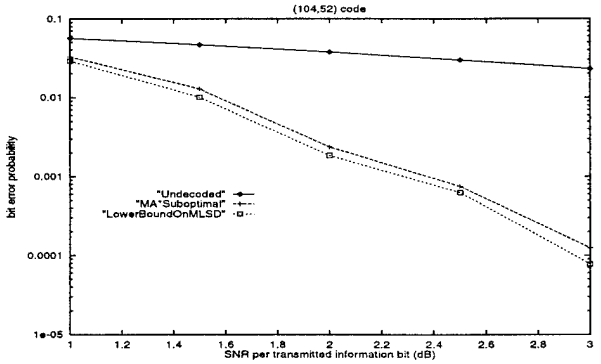


Figure 1: Performance of the MA* SOSD decoding algorithm for the $(104, 52)$ binary extended quadratic residue code

γ_b	6 dB		7 dB		8 dB	
	max	ave	max	ave	max	ave
N(r)	32275	11	3033	2	112	1
C(r)	1639	17	135	17	17	17
M(r)	1000	5	1000	1	112	1

Table 1: Performance of the MA* SOSD decoding algorithm for the $(256, 131)$ binary extended BCH code

When the decoding problem is transformed into a continuous optimization problem [2], it becomes finding a k dimensional real vector that minimizes the cost function g. SA, a technique that statistically guarantees finding global optima for optimization problems, could be applied to solve this problem. SA uses a control parameter called temperature (T), which is initially high and decreased steadily. At each temperature, a large number of possible "moves" are generated, evaluated, and possibly accepted. Each move effects a small change in the current "configuration" (a real vector), and may be obtained by perturbing one component of the current vector by a small quantity. This move is accepted if it decreases cost g. Also, this move is accepted with probability $e^{-\frac{\Delta g}{T}}$ even if the move results in an increase of Δg in g. This provides a mechanism to escape from local (non-global) optima, with higher probability at higher temperatures. There is a high likelihood that the system state moves to the region of the global optimum before the temperature becomes too low. When $T \approx 0$, the algorithm settles into the current local optimum. Simulation results for the SA decoding algorithm will be presented.

References

[1] Y. S. Han, C. R. P. Hartmann, and C.-C. Chen, "Efficient Priority-First Search Maximum-Likelihood Soft-Decision Decoding of Linear Block Codes," *IEEE Trans. on Information Theory*, pp. 1514–1523, September 1993.

[2] K. H. Farrell, L. D. Rudolph, Carlos R. P. Hartmann and Louise D. Nielsen, "Decoding by Local Optimization," *IEEE Trans. on Information Theory*, vol. IT-29, No. 5, pp. 740–743, September 1983.

[1]This work was partially supported by the NSF under Grant NCR-9205422. C. R. Wulff was supported by a Research Experience for Undergraduates Supplement of Grant NCR-9205422.

On Maximum Likelihood Soft Decision Syndrome Decoding

Marc P.C. Fossorier and Shu Lin[1], Jakov Snyders[2]
Dept. of Electrical Engineering, University of Hawaii, Honolulu, HI 96822, USA.
Dept. of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, Israel.

Abstract — **Maximum likelihood decoding (MLD) of binary linear block codes is addressed by combining the approaches of processing the generator matrix G and parity-check matrix H.**

I. MLD BASED ON ORDERING IN THE DUAL SPACE

Consider a binary linear (N, K, d_H) code with generator matrix G and check matrix H. For a given received sequence, let B_K be the most reliable basis (MRB) [2], [3], [4] for the column space $\mathrm{cs}(G) = \mathrm{GF}(2^K)$ of G and let Ω_{N-K} be the least reliable basis (LRB) [1] for $\mathrm{cs}(H) = \mathrm{GF}(2^{N-K})$ of H, consisting of the columns of the respective matrices.

Theorem 1: *The complement of the location set of B_K is the location set of Ω_{N-K}.* □

An efficient way to perform (nearly) MLD starts with forming $\bar{c}_0$, the codeword that agrees with bit-by-bit hard detection at the positions of the MRB. Thereafter, search procedures [2], [3], [4] examine alternatives to $\bar{c}_0$. In [2], the alternatives to $\bar{c}_0$ are considered in successive stages. At each order i of reprocessing, $\binom{K}{i}$ codewords are processed. A resource test tightly related to the reprocessing strategy reduces the number of computations at each decoding stage. A similar approach [4] utilizes a partial ordering of the information vectors. Syndrome decoding [5] is an alternative approach to accomplish MLD.

By Theorem 1, the resource tests can be related to the LRB. Also, those syndrome decoding aspects that are based on the LRB may conveniently be incorporated into the decoding procedure.

II. SYNDROME DECODING ASPECTS

Let $\bar{y}_i \in \Omega_{N-K}$; $i = 1, 2, \cdots, N-K$ be indexed in nondecreasing order of reliability. Let $\bar{s} = H^T \bar{c}_0$ be the syndrome corresponding to $\bar{c}_0$. Assuming $\bar{s} \neq 0$, expand $\bar{s}$ in terms of the LRB, i.e., $\bar{s} = \sum_{j=1}^{j=w} \bar{y}_{p_j}$ where $p_1 > p_2 > \cdots > p_w$. Setting $H = [A\ I_{N-K}]$, with the $N-K$ rightmost positions corresponding to the LRB, w is the Hamming weight of $\bar{s}$.

By [1], if either a) $w = 1$ or b) $p_1 + w \leq d_H$ then $\bar{c}_0$ is the most likely codeword. A stopping rule stronger than b) now follows.

Theorem 2: *If $w \geq 2$ and*

$$\max_{l \in [2, w_s]} \{p_l + 2(l-1) + 1\} \leq d_H, \qquad (1)$$

then order-0 reprocessing is optimum. □

Generalization of Theorem 2 to higher orders i of reprocessings is also presented. By such extension, we associate to either each $\bar{s}$ or the most likely syndromes a set of columns of H to be searched, as in [1]. We provide an efficient algorithm to preprocess the corresponding table look-up. The size of this table can be limited to the most likely error patterns using the statistical approach of [2]. Finally, we present an algorithm which iteratively evaluates the syndrome $\bar{s}$ each time a dimension is added when constructing the LRB. With this algorithm, the most likely error patterns are tested without completing the construction of the LRB.

Syndrome-based tests stop the search more effectively for some received words (typically when the signal to noise ratio (SNR) is low). However, most of the syndrome tests are code-dedicated, whereas resource tests are more universal.

III. SIMULATION RESULTS

For extended Hamming codes of length 2^m, $m \leq 7$, with order-1 reprocessing and table look-up, the maximum number of computations N_{tot} is compared in Table 1 with the worst case results of both [2] and [1] (the latter is MLD). We also indicate the partial ordering maximum cost N_{ord}. The average number of computations N_{ave} rapidly converges to N_{ord} as the SNR increases. For the (24,12,8) Golay code our decoding method requires on average 50 and 15 real operations to achieve practically optimum error performance at the respective BER 10^{-3} and 10^{-6}.

Finally, a new reprocessing algorithm is analyzed. After the ordering has been completed, this algorithm no longer requires real value operations. For all simulated codes, a performance within 1.5 dB of the optimum bit error performance has been achieved, even for long codes. For example, at the BER 10^{-5}, with an $o(K^3)$ syndrome computations, a degradation of less than a dB with respect to the ML performance is achieved for the (128,64,22) extended BCH code.

Table 1: Computation cost for extended Hamming codes.

m	code	N_{ord}	N_{tot}	[1]	order-1
3	(8,4,4)	15	27	17	36
4	(16,11,4)	33	68	60	108
5	(32,26,4)	63	153	188	290
6	(64,57,4)	113	330	-	726
7	(128,120,4)	199	703	-	1,736

REFERENCES

[1] J. Snyders, "Reduced Lists of Error Patterns for Maximum Likelihood Soft Decoding," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1194-1200, July 1991.

[2] M. P. C. Fossorier and S. Lin, "Soft-Decision Decoding of Linear Block Codes based on Ordered Statistics," *IEEE Trans. Inform. Theory*, to appear.

[3] Y.S. Han, C.R.P. Hartmann and C.C. Chen, "Efficient priority first search maximum-likelihood soft decision decoding of linear block codes," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 1514-1523, September 1993.

[4] D. Gazelle and J. Snyders, "Reliability-Based Code-Search Algorithms for Maximum Likelihood Decoding of Block Codes," submitted for publication.

[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, September 1989.

[1]Supported in part by NSF Grant NCR-94-15374.

[2]Supported in part by the Israel Science Foundation administered by the Israel Academy of Sciences and Humanities.

Fast Error Magnitude Evaluations for Reed-Solomon Codes

John J. Komo and Laurie L. Joiner
Electrical and Computer Engineering, Clemson University, Clemson, South Carolina 29634-0915 USA

Abstract — A fast algorithm for the evaluation of error magnitudes for Reed-Solomon codes is obtained here in terms of the error locations and syndromes. This fast algorithm is compared to the Forney algorithm in terms of required additions and multiplications and implementation speed.

I. INTRODUCTION

Assume a t error correcting Reed-Solomon code and assume that the error locations have been determined using the Berlekamp-Massey algorithm or some other procedure. The Forney algorithm [1] is the common algorithm used for obtaining the error magnitudes in Reed-Solomon decoding. For a codeword with $\nu \le t$ received errors, the Forney algorithm calculates the error magnitudes from the error locations β_i, $i=1,2,\cdots,\nu$ and the syndromes S_i, $i=1,2,\cdots,\nu$ as [2,3]

$$\gamma_i=\Omega(\beta_i^{-1})\Big[\prod_{j=1,j\neq i}^{\nu}(1+\beta_j\beta_i^{-1})\Big],\ i=1,2,\ldots,\nu \qquad (1)$$

where $\Lambda(X)[1+S(X)]\equiv\Omega(X) \bmod (X^{2\nu+1})$, $S(X)=S_1X+S_2X^2+\cdots+S_tX^t$, and $\Lambda(X)=(1+\beta_1X)(1+\beta_2X)\cdots(1+\beta_\nu X)=1+\Lambda_1X+\Lambda_2X^2+\cdots+\Lambda_\nu X^\nu$. Now $\Omega(X)$ can be expressed as [4] $\Omega(X)=1+(S_1+\Lambda_1)X+(S_2+\Lambda_1S_1+\Lambda_2)X^2+\cdots+(S_\nu+\Lambda_1S_{\nu-1}+\Lambda_2S_{\nu-2}+\cdots+\Lambda_{\nu-1}S_1+\Lambda_\nu)X^\nu$. The number of required additions is $(5\nu^2-\nu)/2$ and the required multiplies is $(7\nu^2-5\nu)/2$. In addition, there are $\nu(\nu-1)$ exponentiations.

In general once the error locations have been determined, the error magnitudes, γ_i, $i=1,2,\cdots,\nu$, are obtained by solving

$$\begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_\nu \\ \beta_1^2 & \beta_2^2 & \cdots & \beta_\nu^2 \\ \vdots & \vdots & & \vdots \\ \beta_1^\nu & \beta_2^\nu & \cdots & \beta_\nu^\nu \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_\nu \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_\nu \end{bmatrix} \qquad (2)$$

II. FAST ALGORITHM

Since the ß matrix is a Vandermonde matrix it is of full rank and standard techniques can be used to diagonalize it. Then using back substitution and making full use of the structure of a Vandermonde matrix, we developed the following iterative algorithm for obtaining the error magnitudes

$$\begin{aligned}
&S_{\nu-j}=S_{\nu-j}-\beta_iS_{\nu-j-1} && j=0,\cdots,\nu-i-1,\ i=1,\cdots,\nu-2\\
&S_\nu=(S_\nu-\beta_{\nu-1}S_{\nu-1})/(\beta_\nu-\beta_{\nu-1})\\
&S_{\nu-i}=S_{\nu-i}-S_{\nu-i+j} && j=1,\cdots,i \ \Big\}\\
&S_{\nu-i}=S_{\nu-i}/(\beta_{\nu-i}-\beta_{\nu-i-1}) && \Big\} \quad i=1,\cdots,\nu-2\\
&S_{\nu-i+j}=S_{\nu-i+j}/(\beta_{\nu-i+j}-\beta_{\nu-i-1}) && j=1,\cdots,i \ \Big\} \qquad (3)\\
&S_1=S_1-S_{j+1} && j=1,\cdots,\nu-1\\
&S_j=S_j/\beta_j && j=1,\cdots,\nu
\end{aligned}$$

where the error magnitudes γ_i's are contained in the S_j's. The number of required additions is $3\nu(\nu-1)/2$ and the required multiplies is ν^2.

This fast algorithm for evaluating the error magnitudes for Reed-Solomon decoding requires approximately 5/3 fewer additions and 7/2 fewer multiplications than the Forney algorithm without any exponentiations. The total number of operations including additions, multiplications, and exponentiations for the Forney algorithm is $7\nu^2-4\nu$ and for the fast algorithm $(5\nu^2-3\nu)/2$. Also, the memory required for the Forney algorithm and the fast algorithm are both small and essentially equal to the number of error magnitudes ν. Thus, the fast algorithm calculates the error magnitudes faster than the Forney algorithm by a factor ranging from approximately 1.67 to 3.5. If the operations require the same time the speedup factor is approximately 2.8.

A comparison of the execution times for calculating the error magnitudes using the Forney algorithm and the fast algorithm was performed for a length 1023 Reed-Solomon code with $\nu=t=1,2,\cdots,10$. It was shown that for this case the execution times for the fast algorithm are at least a factor of two faster than the Forney algorithm.

This work was supported in part by a grant from the US Army Research Office under the Focused Research Initiative.

III. REFERENCES

[1] G.D. Forney, "On Decoding BCH Codes," *IEEE Transactions on Information Theory*, Vol. IT-11, pp. 549-547, October 1965.

[2] R. E. Blahut, *Theory and Practice of Error Control Codes*, Reading, MA: Addison-Wesley, 1988.

[3] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, Englewood Cliffs, NJ: Prentice-Hall, 1995.

[4] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1983.

On Neural Decoding for Some Cyclic Codes

Yu Jian-ping, Zhao Yu-biao, and Wang Xin-mei
National Key Lab. of ISN, Xidian University, Xi'an 710071, P.R.China

Abstract — Based on analysis of the property of a class of cyclic codes, an algorithm for neural soft decision decoding of those codes is presented. The complexity of the new algorithm is much less than that of the available algorithms for decoding general linear block codes, and its performance is approached to that of the maximum likelihood decoding.

SUMMARY

The adaptability and parallel computing capability of neural networks make them be specially adequate for error correcting tasks. Several neural decoding schemes have been proposed. Now, it is well known that neural networks can be employed in soft-decision(SD) decoding, however, it calls for further study on the decoding complexity.

For an ordinary (n,k) binary linear block codes, most available neural decoding implementations perform SD decoding by searching a codeword with a minimal distance apart from the received vector $\mathbf{r}=(r_1,r_2,\cdots,r_n)$ in the whole code space C consisting of 2^k elements, the decoding complexity becomes large as 2^k increases. We can define the decoding complexity as the number of elements in the decoder searching set. So the complexity of an ordinary neural decoder for (n,k) code is 2^k. To a class of cyclic codes, we trade a slight degradation in performance for reducing decoding complexity by using a property of these codes.

Consider a systematic cyclic code $C(n,k,d)$, whose error-correcting capability is $t \le \lfloor \frac{d-1}{2} \rfloor$ for hard-decision(HD) decoder. Encoding is described in the group $(\{1,-1\},\times)$, the encoding equation is $c_i = \prod_{i=1}^{k} b_i^{g_{ij}}$, where $\{g_{ij}\}$ is the generator matrix. In each codeword, the first k bits are the information bits, which correspond to a set $I=\{1,2,\cdots,k\}$ and the other $n-k$ bits correspond to a set $Q=\{k+1,k+2,\cdots,n\}$. Define a weight $W_A(\mathbf{e})=\sum_{i\in A} e_i$, where A is a subset of $\{1,2,\cdots,n\}$. We can prove the following theorem:

Theorem 1 Let $\mathbf{r}$ be a received vector of the systematic cyclic code, $\mathbf{e}=\mathbf{c}\oplus\mathbf{b}$ be the error vector can be corrected by SD decoding, we get $W(\mathbf{e})=\sum_{i=1}^{n} e_i \le d-1$. Then, the number of error bits in I can be always reduced to

$$W_I(\mathbf{e})=\sum_{i=1}^{k} e_i \le t \tag{1}$$

by cyclically shifting the vector $\mathbf{r}$, if and only if

$$\frac{k}{n} < \frac{t+1}{d-1} \tag{2}$$

Using the above property, we lead a simplified decoding implementation for those codes. The new implementation is described as the following:

1) Cyclically shift $\mathbf{r}$ m times to get $\mathbf{r}^*$, such that $W_I(\mathbf{r}^*)=\sum_{i=1}^{k} r_i^*$ is minimized ;

2) Determine hard-decision vector $\mathbf{b}^*$ of $\mathbf{r}^*$;

3) Encode $\mathbf{b}_I^*=\{b_1^*,b_2^*,\cdots,b_k^*\}$,get a codeword $\mathbf{c}^*$, where $c_i^*=\prod_{i=0}^{k} b_i^{g_{ij}}$;

4) $\mathbf{r}'=\mathbf{c}^*\odot\mathbf{r}^*$, where $r_i'=r_i^*c_i^*$;

5) Decode $\mathbf{r}'$ to obtain a codeword $\mathbf{c}'$ using a neural decoder;

6) $\hat{\mathbf{c}}^*=\mathbf{c}'\odot\mathbf{c}^*$;

7) Cyclically shift $\hat{\mathbf{c}}^*$ $n-m$ times, get a codeword $\hat{\mathbf{c}}$, which is the result of decoding.

In the first step, we get $\mathbf{r}^*$ with a minimum weight $W_I(\mathbf{r}^*)$. An approximate assumption is that $\mathbf{r}^*$ satisfies (1). Based on this assumption, the number of error bits in I of $\mathbf{r}^*$ is no more than t, then we get $\sum_{i=1}^{k} c_i' \le t$. So, the decoder of the fifth step searches the desired codeword only in a subset S_u of C consisting of the codewords c' such that $W_I(\mathbf{c}') \le t$ rather than in the whole codeword space C. This decoder is called "narrow sense decoder(NSD)". The number of codewords in S_u is $N_u=\sum_{i=1}^{t}\binom{k}{i}$, so the complexity of NSD is N_u. The complexity of new decoder is little larger than N_u, so the proposed decoder is much simpler than the ordinary decoder. Simulation results indicate that the performance is close to that of maximum likelihood decoder.

Poisson Approximation for Excursions of Adaptive Algorithms

Adel A. Zerai
Electronics Eng. Tech. Dept.
College of Technological Studies
P.O.Box: 42325 Shuwaikh, Kuwait 70654

James A. Bucklew
Dept. Elect. & Comp. Eng.
University of Wisconsin-Madison
1415 Johnson Drive, Madison, WI 53706

Abstract — **This paper analyzes excursions of adaptive algorithms. The distribution of the number of excursions in n units of time is approximated by a Poisson distribution. The mean and distribution of the time of the occurrence of the first excursion are approximated by those of an exponential distribution. Expressions for the error in the approximations are derived. The approximations are shown to hold asymptotically as the excursion defining set converges to the empty set and as the algorithm's step size μ converges to zero. The validity of the approximations is tested on a variety of examples.**

I. Introduction

We study excursions of adaptive algorithms of the form

$$W_{k+1} = W_k - \mu\, h(W_k, X_k, D_k), \tag{1}$$

where X_k and D_k are real valued random variables, μ is a constant known as the algorithm's step size, and h is a measurable function.

The updates of the error between estimated and optimal weights for many adaptive filters (for example The Least Mean Square (LMS) algorithm and its "signed" variants) are of the form of Eq. (1). When one of these filters is driven by an i.i.d. sequence of inputs $\{X_k\}$ and an independent i.i.d. sequence of disturbances $\{D_k\}$, then Eq. (1) defines a discrete time Markov chain. The performance of an algorithm is acceptable if its corresponding Markov chain spends most of its time in a neighborhood of the equilibrium 0 (or preferably at 0). However, on rare occasions, an excursion (a visit or a cluster of visits to the set $B = [b, \infty)$ or the set $B = [-b, b]^c$) will occur.

Denote the time of the beginning of the first excursion by τ_B and the number of excursions in n units of time by S_n. We approximate the expectation of τ_B, the distribution of τ_B, and the distribution of S_n. The distribution and the mean of τ_B are approximated by those of an exponential distribution with mean $1/\pi(B)\theta$ and the distribution of S_n by a Poisson distribution with mean $\lambda\theta$, where $1/\theta$ is the mean clump size given that there is an excursion, $\lambda = n\pi(B)$, and π is the stationary distribution of the chain.

II. Excursion analysis as $B \to \phi$

Lattice state space case: Let $\{W_k\}_{k\geq 0}$ be an irreducible, positive recurrent, aperiodic Markov chain with a countable state space $\mathcal{S}$ (e.g. the even steps of the sgn-sgn variant of the LMS algorithm). Define an excursion to be a cluster of visits to the set B that is separated by the previous cluster by a visit to state 0 or by r visits to B^c for some integer r. Dividing the n steps of the chain into independent cycles that start from state 0 and end at state 0, calculating an upper bound for the probability that a cycle contains more than one cluster, and using the law of rare events [1, page 117] produces the desired approximation for S_n [2, theorems 3.1 and 3.2].

The approximation for the distribution of τ_B can be derived from the fact that $P(\tau_B > x) = P(S_{\lfloor x \rfloor} = 0)$ and the approximation derived for S_n. The approximation for the $E\tau_B$ is given in [2, lemma 3.3].

The sequence $\{W_k\}$ considered so far is a scalar sequence. However, the approximations are valid even where $\{W_k\}$ is a sequence of vectors of size m. All that is needed is to map the state space μ times Z^m into μ times Z and choosing a sequence of sets B_n that converges to ϕ, for example the sets $([-b\mu, b\mu] \times [-b\mu, b\mu] \times \cdots \times [-b\mu, b\mu])^c$.

Continuous state space case: The three approximations are extended to algorithms with an uncountable state space under the assumption that the resulting Markov chain is Harris recurrent. This assumption will be required in order to attach to the chain a generic atom that is visited infinitely often and hence may be used as a regeneration state.

Examples of algorithms with both continuous and lattice state space are given to demonstrate these results. One of the examples demonstrates the different behavior of clusters for the LMS algorithm and three of its signed variants. Another example demonstrates the applicability of the approximations in the vector case.

III. Excursion analysis as $\mu \to 0$

Here we assume that the set B that defines excursions is a fixed subset of the real line, for example $[b, \infty)$, and examine excursions as the step size $\mu \to 0$. The results of Section II, where $B \to \phi$, analyze excursions of a single Markov chain and increase the rarity of excursions by decreasing the set B. In contrast, in this section, each value of μ in Eq. (1) defines a Markov chain with some stationary distribution, say π_μ, and the rarity of excursions is increased by decreasing $\pi_\mu(B)$.

Our approximations for the mean and distribution of τ_B are still valid in this setting . All that remains to be answered is whether these bounds converge to 0 or not as $\mu \to 0$. We show that the approximations for the mean and distribution of τ_B hold. The approximation for the distribution of S_n is shown to hold after some modification to the definition of an excursion.

References

[1] R. Durrett, *Probability: Theory and Applications,* Belmont, California, Wadsworth, 1991.

[2] A. Zerai and J. Bucklew, "Poisson approximation for excursions of adaptive algorithms with a lattice state space," *IEEE Trans. Info. Theory* submitted 1994.

Additive random sampling and exact reconstruction

B.Lacaze and A.Duverdier
ENSEEIHT/GAPSE, 2 rue Camichel, 31071 Toulouse, France

Abstract — **This paper treats the case where $Z(t)$ is a continuous wide sense stationary process which is sampled at instants $t_n = n + A_n$, $n \in Z$. The series of random gaps A_n is stationary in the sense weaker than strict second order. We present a necessary and sufficient condition (NSC) for the exact (mean square) linear reconstruction of $Z(t)$.**

I. Introduction

Chronological series often stem from random continuous time processes ($t \in R$) that we wish to reconstruct. The linear reconstruction of the underlying process depends on the sampling technique and on the information we have on this latter.

In the framework of wide sense stationary processes, the case $t_n = n\theta$ has been completely resolved by Lloyd [1]. Concerning the random sampling, the model $t_n = n\theta + A_n$ is the most frequently used. When the gaps A_n are known or observed, numerous reconstruction formulas exist [2] and new ones still appear [3].

On the other hand, few attemps have been made in the case where the A_n are not observed and characterized only by their statistical properties [4]. In what follows we give a NSC for linear reconstruction without (mean square) error, of the underlying process, in the case where the A_n sequence is stationary in some sense.

II. Hypothesis

The taken hypotheses are marked H_0 for the sampled process $Z = \{Z(t), t \in R\}$ (wide sense stationary, mean square continuous) and H_1 for the sequence $A = \{A_n, n \in Z\}$ of non degenerated r.v. (Z and A being supposed independent):

$$H_0 \quad \begin{cases} E[Z(t)] = 0 \\ E[Z(t)\,Z^*(t-\tau)] = \int_{-\infty}^{+\infty} e^{i\omega\tau} dS_Z(\omega) \\ Z(t) = \int_{-\infty}^{+\infty} e^{i\omega t} d\Phi_Z(\omega) \end{cases} \tag{1}$$

$$H_1 \quad \begin{cases} \psi(\omega) = E\left[e^{i\omega A_n}\right] & (i) \\ \forall q \in Z, \varphi_q(\omega) = E\left[e^{i\omega(A_n - A_{n-q})}\right] & (ii) \\ (i) \text{ and } (ii) \text{ do not depend on } n \end{cases} \tag{2}$$

S_Z and Φ_Z are the power spectrum (spectral measure) and the Cramer-Loève representation of Z respectively [5].

The sequence $U = \{U_n, n \in Z\}$ where $U_n = Z(t_n)$, spans a Hilbert space $H(U)$. The problem is to know if, for any t, $Z(t) \in H(U)$ or equivalently $H(U) = H(Z)$ ($H(Z)$ is engendered linearly by Z). In this case, the observation of the randomly sampled sequence is enough to construct the original process Z.

III. Theorem

Let:

$$\begin{cases} G_n = \int_{-\infty}^{+\infty} e^{i\omega n} \psi(\omega)\, d\Phi_Z(\omega) \\ V_n = U_n - G_n \end{cases}$$

$Z(t)$ can be reconstructed linearly without (mean square) error from the observation of the series $U = \{U_n, n \in Z\}$ where $U_n = Z(t_n)$, if and only if:

a) The spectral measures (on $[-\pi, \pi]$) of the two sequences $G = \{G_n, n \in Z\}$ and $V = \{V_n, n \in Z\}$ are mutually singular.

b) The translated measures $S_Z^k(\omega) = S_Z(\omega + 2\pi k)$ are mutually singular for any $k \in Z$.

c) If $\Delta = \{\omega; \psi(\omega) \neq 0\}$ then $\int_\Delta dS_Z(\omega) = E\left[|Z(t)|^2\right]$.

Remark:

Condition a) is easily verified in the case where Z has a line spectrum and A is a continuous r.v. process.

The second condition is due to Lloyd [1].

Condition c) signifies that $\psi(\omega)$ is not nul on the support of $S_Z(\omega)$.

IV. Conclusion

In this paper, we obtained a condition necessary and sufficient for the exact linear reconstruction of a stationary stochastic process subjected to a random additive sampling.

References

[1] S.P. LLOYD, *A Sampling Theorem for (wide sense) Stochastic Processes*, Trans. Ann. Math. Soc., Vol 92, pp.1-12, 1959

[2] A.J. JERRI, *The Shannon Sampling Theorem. Its Various Extensions and Applications*, Proc. of IEEE, Vol. 65, pp. 1565-1595, 1977

[3] Y.M. ZHU, *Generalized Sampling Theorem*, IEEE Trans. on Circuits and Systems II, ADSP, Vol. 39, pp. 587-588, 1992

[4] A.V. BALAKRISHNAN, *On the Problem of Time Jitter in Sampling*, IRE Trans. on Inf. Th., pp226-236, 1962

[5] J.L. DOOB, *Stochastic Processes*, Wiley, 1952

Distortion Measures via Parametric Filtering

Ta-Hsin Li* **and Jerry D. Gibson†**

Texas A&M University, College Station, Texas 77843

Abstract — **Distortion measures are proposed on the basis of parametric filtering, a technique of signal characterization that combines a parametric filter bank with an analysis of first-order autocorrelation. Robustness of the distortion measures against narrow-band interference and spectral notch filtering is investigated.**

I. Introduction

Given a zero-mean stationary signal X_t, consider the demodulated first-order autocorrelation as defined by

$$\gamma_\theta(\eta) := \Re\{e^{-i\theta}\rho(\alpha)\} \quad (-1 < \eta < 1),$$

where $\rho(\alpha)$ is the (ordinary) first-order autocorrelation of the filtered signal $Y_t(\alpha) := \overline{\alpha}\, Y_{t-1}(\alpha) + X_t$ with $\alpha := \eta e^{-i\theta}$. It can be shown [1] [2] that for almost any θ the function $\gamma_\theta(\eta)$ uniquely determines the correlation structure of X_t and hence forms a *characterization function* of the signal. The *parametric filtering* (PF) method is one that utilizes this characterization property of $\gamma_\theta(\eta)$ for signal discrimination [1]. In particular, distortion measures can be derived from $\gamma_\theta(\eta)$.

II. PF-Based Distortion Measures

For any $-1 \le \eta_a < \eta_b \le 1$, consider the function

$$p_\theta(\eta) := \tfrac{1}{2}\big[\gamma'_\theta(\eta) + \big(\gamma_\theta(\eta_a^+) + 1\big)\,\delta(\eta - \eta_a) + \big(1 - \gamma_\theta(\eta_b^-)\big)\,\delta(\eta - \eta_b)\big],$$

where $\gamma'_\theta(\eta)$ is the derivative of $\gamma_\theta(\eta)$ w.r.t. η and $\delta(\eta)$ is the Dirac delta. Using the results in [1], it can be shown that $p_\theta(\eta)$ not only is equivalent to $\gamma_\theta(\eta)$ but also forms a *generalized pdf* in the interval $[\eta_a, \eta_b]$. This latter property gives rise to many possibilities of defining distortion measures. For instance, one may define the Kullback-Leibler information divergence by

$$\kappa(p_\theta^0 \| p_\theta^1) := \int_{\eta_a}^{\eta_b} p_\theta^0(\eta)\, K\big(p_\theta^1(\eta)/p_\theta^0(\eta)\big)\, d\eta,$$

where $K(u) := u - \log u - 1$. Since the information divergence extends to non-probability densities, one may also define a distortion measure as

$$\kappa(p_\theta^0; p_\theta^1) := \kappa(q^* \| p_\theta^0/p_\theta^1) = \int_{\eta_a}^{\eta_b} K\big(p_\theta^0(\eta)/p_\theta^1(\eta)\big)\, d\eta,$$

where $q^*(\eta) := 1 + \delta(\eta - \eta_a) + \delta(\eta - \eta_b)$ is the density of "uniform" distribution.

III. Robustness

Suppose the signal is contaminated by a narrow-band noise so that X_t has a spectral density of the form $f_1(\omega) = (1-\epsilon) f_0(\omega) + \epsilon g(\omega)$, where f_0 is the noise-free spectrum and g is the noise spectrum with $g(\omega) = \frac{1}{2}\Delta^{-1}$ for $|\omega \pm \omega_0| < \frac{1}{2}\Delta$ and $g(\omega) = 0$ otherwise ($\Delta \ll 1$). To quantify the robustness of a distortion measure against the contamination, we use the second derivative of the distortion measure at $\epsilon = 0$, known as the *local curvature* of the distortion measure.

For the widely used Kullback-Leibler (KL) spectral divergence [3], $D_{\mathrm{KL}}(f_1, f_0) := \int K(f_1(\omega)/f_0(\omega))\, d\omega$, it is easy to show [4] that $(d^2/d\epsilon^2)\, D_{\mathrm{KL}}(f_1, f_0)|_{\epsilon=0} = \mathcal{O}(\Delta^{-1})$. This confirms again that the KL divergence is *not* robust to narrow-band contaminations [3].

Compared to the KL spectral divergence, the PF-based distortion measures exhibit more robustness to narrow-band contaminations. In fact, it is not difficult to show that with $1 - \max\{|\eta_a|, |\eta_b|\} \gg \Delta$ the local curvatures of $\kappa(p_\theta^0 \| p_\theta^1)$ and $\kappa(p_\theta^0; p_\theta^1)$ take the form of $\mathcal{O}(1)$ as $\Delta \to 0$. Similar results can be obtained for distortions due to spectral notch filtering [4].

References

[1] T. H. Li, "Discrimination of time series by parametric filtering," *J. Amer. Statist. Assoc.*, to appear.

[2] T. H. Li and J. D. Gibson, "Discriminant analysis of speech by parametric filtering," *Proc. Conf. Inform. Sci. Syst.* (Princeton, NJ, 1994), pp. 575–580.

[3] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[4] T. H. Li, "Poisson integral, time series characterization and robust distortion measures," submitted.

*T. H. Li is with the Department of Statistics.

†J. D. Gibson is with the Dept. of Electrical Engineering. He was partly supported by NSF grant NCR-9303805.

A Two-Step Markov Point Process*

Majeed M. Hayat and John A. Gubner
Dept. Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706

Abstract — Existence and uniqueness are established for a translation-invariant Gibbs measure corresponding to a spatial point process that has, in addition to inhibition and clustering, the new feature of penalizing isolated points. This point process has the so-called two-step Markov property, and the associated density function is characterized in terms of 2-interaction functions. The asymptotic normality of certain statistics of the point process is established when the size of the observation window tends to $\mathbb{R}^2$.

I. A Two-Step Markov Point Process on Bounded Sets

Let Ω_F denote the set of all finite lists of points from $\mathbb{R}^2$. A typical point $x \in \Omega_F$ has the form $x = (x_1, \ldots, x_n)$ for some n, where each $x_i \in \mathbb{R}^2$. For $x \in \Omega_F$, the number of *isolated points* in x is given by $I(x) := \big|\{i : \|x_i - x_j\| > d_2, \forall j \neq i\}\big|$, where $d_2 > 0$ is a specified threshold, $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^2$, and $|\cdot|$ denotes the cardinality of the indicated set. Fix $0 < d_1 < d_2$, and let $\psi : [0,\infty) \to [0,\infty)$ be a bounded function such that $\psi(r) = 1$ whenever $r \geq d_2$, and $\psi(r) = 0$ if $r < d_1$. Fix $0 < \gamma \leq 1$. For $x \in \Omega_F$, consider the density function $f(x) = a\gamma^{I(x)} \prod_{i<j} \psi(\|x_i - x_j\|)$, where a is a normalizing constant [3, Section 2]. The function ψ is responsible for pairwise interaction and may give rise to clustering and inhibition [1]. What is new here is the constant γ which is responsible for penalizing realizations with isolated points. We call two points neighbors if the distance between them is no more than d_2. It can be shown [3] that the ratio $f(x \cup \{\xi\})/f(x)$ depends on ξ, on the points of x that are neighbors of ξ, and on the neighbors' neighbors. If we consider the probability measure $f d\nu_\Lambda$, where Λ is a bounded set and ν_Λ is the measure corresponding to a Poisson process in Λ with constant intensity λ, then the conditional probability of an event in $\Lambda' \subset \Lambda$, given what is in $\Lambda \setminus \Lambda'$, depends on the points in Λ', on the points in $\Lambda \setminus \Lambda'$ that are neighbors of Λ', and the neighbors' neighbors. This fact motivates the term "two-step Markov." In [3], we extended Ripley and Kelly's [7] characterization theorem on Markov density functions to m-step Markov densities. As a result, we obtain the representation $f(x) = a \prod_{y \subset x} \Phi(y)$, where Φ is a so-called 2-interaction function, i.e., $\Phi(y) \neq 1$ implies that every two points in y are either neighbors, or else, there is a third point of y which is a neighbor to both points. Furthermore, $\Phi(y) = 1$ whenever $\max_{i,j} \|y_i - y_j\| > 2d_2$.

II. Existence and Uniqueness of a Gibbs Measure on $\mathbb{R}^2$

The goal of this section is to define a point process (having the features of the point process in the previous section) on the set Ω of all lists of points from $\mathbb{R}^2$ whose intersections with bounded sets are finite. Note that some of the elements of Ω are infinite lists. Following Preston [6, Chapter 6], we define the translation-invariant potential function V on Ω_F by $V(x) := -\log(f(x)/a) = -\sum_{y \subset x} \log \Phi(y)$. For any subset A in $\mathbb{R}^2$ and $s \in \Omega$, let s_A denote the restriction of s to A. Let Λ be any bounded subset of $\mathbb{R}^2$. If x is a list of points from Λ, and y is a list of points from Λ^c, define the conditional potential [6, p. 98] $V_\Lambda(x|y) := -\lim_{A \uparrow \mathbb{R}^2} \big\{\sum_{z \subset x \cup y_A, z \cap x \neq \emptyset} \log \Phi(z)\big\}$. For each temperature and each Λ, we can define a conditional probability measure corresponding to the above conditional potential. This conditional measure partially inherits the property of penalizing isolated points. The following result is a consequence of [5, Theorem 2.2 & Remark 2.3], [2] and [3, Lemma 3.2].

Theorem 1: For every sufficiently large temperature, there exists a unique translation-invariant Gibbs measure defined on events on Ω, and corresponding to the above conditional measures.

III. Asymptotic Normality

Assume that $\psi(r) = \sum_{i=1}^{M+1} \psi_i \mathbf{1}_{[r_{i-1}, r_i)}(r)$, where $0 = r_0 < d_1 = r_1 < \cdots < r_M = d_2, r_{M+1} = \infty, \psi_1 = 0, \psi_{M+1} = 1$, and $M \geq 1$. Observe that the density f now takes the form of $f(x) = a\gamma^{I(x)} \prod_{i=1}^{M} \psi_i^{S_i(x)}$, where $S_i(x)$ is the number of pairs of points that are r_{i-1} to r_i units apart. Let Λ_n, $n = 1, 2, \ldots$, be a sequence of bounded subsets of $\mathbb{R}^2$ such that $\Lambda_n \uparrow \mathbb{R}^2$ and $n/\text{area}(\Lambda_n)$ converges to a finite constant as $n \to \infty$. Let n be fixed, and define $X := (N, S_1, \ldots, S_M, I)$, where N is the total number of points in a realization of the Gibbs process of the previous section, I is the total number of isolated points, and S_i is the number of pairs of points that are within a distance r_{i-1} to r_i, all in Λ_n. For each $j = (j_1, j_2) \in \mathbb{Z}^2$, let $U_j := \{u = (u_1, u_2) \in \mathbb{R}^2 : d_2 j_i \leq u_i \leq d_2(j_i + 1), i = 1, 2\}$. Let $\mathcal{J}_n$ be the set of indices j for which $\Lambda_n \cap U_j$ is not empty. The asymptotic normality of X relies on [4, Theorem 2.2] and on [3, Lemma 3.2].

Theorem 2: If the temperature is sufficiently large, then as $n \to \infty$, $(X - \mathrm{E}[X])/|\mathcal{J}_n|^{1/2}$ converges in distribution to a zero-mean normally distributed $\mathbb{R}^{M+2}$-valued random variable with a covariance matrix specified in [3, Section 4]

References

[1] A. Baddeley and J. Møller, "Nearest-neighbour Markov point processes and random sets," *Int. Statist. Rev.*, vol. 57, pp. 89–121, 1989.
[2] H. Föllmer, "A covariance estimate for Gibbs measures," *J. Functl. Anal.*, no. 46, pp. 387–395, 1982.
[3] M. M. Hayat and J. A. Gubner, "A two-step Markov point process," submitted for publication, Mar. 1995.
[4] J. L. Jensen, "Asymptotic normality of estimates in spatial point processes," *Scand. J. Statist.*, vol. 20, pp. 97–109, 1993.
[5] D. Klein, "Convergence of grand canonical Gibbs measures," *Commun. Math. Phys.*, vol. 92, pp. 295–308, 1984.
[6] C. J. Preston, *Random Fields*. Heidelberg: Springer Lecture Notes in Mathematics 534, 1976.
[7] B. D. Ripley and F. P. Kelly, "Markov point processes," *J. London Math. Soc.*, vol. 2, pp. 188–192, 1977.

*Supported by ONR under Grant N00014-94-1-0366.

Noisy Attractors of Markov Maps.

G. Salazar - Anaya[1]
Department of Mathematics and Statistics, Carleton University. Ottawa, Canada
and
Jesús Urías[2]
Instituto de Investigación en Comunicación Optica, Universidad Autónoma de San Luis Potosí, 78000 SLP México.

Abstract — **It is shown that Markov maps when subjected to weakly continuous random perturbations have an attractive invariant measure that incorporates the dispersive effects of perturbations as well as the ordering effects of the mapping.**

I. Markov Maps.

Are defined on the basis of a finite set of functions $\{f_i \mid i = 1,2...N\}$ on a compact metric space (X,d). Associating a probability p_i to every function f_i, normalized as $\sum p_i = 1$, a probabilistic dynamics on X is defined by the map $x \mapsto f_i(x)$, with probability p_i. This probabilistic dynamics on X defines a deterministic dynamics on the set of probability measures on X, $\mathcal{P}(X)$, by the Markov operator, M. For a measure $\nu \in \mathcal{P}(X)$ and each measurable set $A \subset X$, the action of M is defined by

$$M\nu(A) = \int d\nu\, P_0(A|\cdot) = \sum_{i=1}^{N} p_i \nu \circ f_i^{-1}(A),$$

where $P_0(A|x)$ is the usual Markov transition probability and $\mathcal{P}(X)$ is endowed with Hutchinson's metric [1]. When the functions f_i have contractivity factors $s_i < 1$, the Markov operator M has contractivity factor $s = \max\{s_i\} < 1$. The dynamics of a Markov map is very simple: for all initial μ, $M^n\mu$ converges weakly to the invariant measure. Techniques to encode images as fractals are based on this fact.

II. Random Perturbations.

An operator $S: \mathcal{P}(X) \to \mathcal{P}(X)$ describes the stationary random perturbations on X. We restrict to random perturbations that are specified by their action on atomic measures (for examples, see [3]), by giving a function $N: B(X) \times X \to R$ such that $N(A|x) = S\delta_x(A)$ for every point $x \in X$ and any measurable subset $A \subset X$. The function $N(A|\cdot)$ is measurable for each A. The action of S on any measure $\nu \in \mathcal{P}(X)$ is then given by $S\nu(A) = \int d\nu\, N(A|\cdot)$. The perturbed Markov map is defined by the combined operation $R\nu = S \circ M$ and it follows that $R\nu \in \mathcal{P}(X)$ whenever $\nu \in \mathcal{P}(X)$. Written in terms of N, the perturbed map is $R\nu(A) = \int P(A|\cdot)d\nu$, where we introduced $P(A|x) = \sum p_i N(A|x) \circ f_x(x)$, the perturbed transition probability. In the unperturbed limit, $N(A|x) = \delta_x(A)$ and $P(A|x)$ reduces to $P_0(A|x)$ as it should. Notice that it is enough to define the function $N(A|\cdot)$ at points $x \in W(X) \equiv \cup_i f_i(X) \subset X$.

III. Stability under Random Perturbations.

The effects of a random perturbation are the opposite to the effects of M. The stability under random perturbations is not evident [2]. A Markov map is stable under a given perturbation if the corresponding randomly perturbed Markov operator has a unique attractive invariant measure. Under a severe random perturbation not every Markov map would be stable. However, we have found a class of random perturbations that do not change the contractivity of Markov maps. Perturbations in this class we call weakly continuous perturbations. A perturbation $N(A|x)$ is weakly continuous if

$$\left| \int f\, dN(\cdot|x) - \int f\, dN(\cdot|y) \right| \le d(x,y),$$

for every pair of points x, y in X. Notice that this condition is not a restriction on the amplitude of the perturbation, it simply is a weak form of continuity on N.

THEOREM(weak continuity implies stability) *Let N be a weakly continuous random perturbation. Then R has the same contractivity factor s as M.*

The theorem says that under the class of weakly continuos perturbations, the perturbed Markov operator R has a unique attractive fixed point ν_∞. In other words, Markov maps are stable under weakly continuous perturbations, in the sense that there exists an attractive invariant measure satisfying the equation $R\nu_\infty = \nu_\infty$, for any choice of the set of probabilities, $\{p\}$.

Weakly continuous perturbations conform a class big enough as to include *the full class* of homogeneous perturbations, i.e., those that are introduced with the help of independent identically distributed random variables [3]. Hence, under any translational invariant perturbation a Markov map always has an attractive invariant measure. Interesting features of the noisy invariant measure [4] are that it shows details much finer than the length scale settled by the noise amplitude and that the self-similar property of the unperturbed invariant measure is lost. At small noise amplitudes a degraded self-similarity is retained.

References.

[1] J. Hutchinson. Indiana Univ. J. Math. **30**, 713 (1981).
[2] R. García-Pelayo and W.C. Schieve. J. Math. Phys. **33**, 570 (1992)
[3] F. Moss and P.V.E. McClintock (editors). *Noise in nonlinear dynamical systems* Vols. 1-3. Cambridge U. Press (1988).
[4] G. Salazar-Anaya and J. Urías. J. Math. Phys. Submitted (1995).

[1] CONACYT (México) fellow.
[2] This work was partially supported by CONACYT (México) under contract 2109-E.

Revisiting the Huber-Strassen Minimax Theorem for Capacities

Heinrich Schwarte[1] and John S. Sadowsky[2]

[1]Department of Mathematics, University of Essen, 45117 Essen, Germany

[2]Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA

I. Introduction

Let $\mathcal{M}$ be the collection of probability measures on a measurable space $(E, \mathcal{B})$. Consider the binary decision problem H_0 vs. H_1 where the statistical hypotheses are represented by nonparametric families of probability distributions $\mathcal{P}_i \subset \mathcal{M}, i = 0, 1$. Two important special cases are *ϵ-contamination* and *total variation* families. These families are defined by

$$\mathcal{P}_i = \{Q | Q = (1 - \epsilon_i)P_i + \epsilon_i H, H \in \mathcal{M}\}$$

and

$$\mathcal{P}_i = \{Q \in \mathcal{M} | \sup_{A \in \mathcal{B}} |Q(A) - P_i(A)| \leq \epsilon_i\},$$

respectively, for some $P_i \in \mathcal{M}$ and $0 \leq \epsilon_i \leq 1$. In these cases the $\mathcal{P}_i$ formalize the possibility of deviations from the nominal models P_i. We seek Neyman-Pearson and Bayes minimax tests between $\mathcal{P}_1$ and $\mathcal{P}_2$.

A pair of distributions $(Q_0, Q_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ is called a *least favourable pair* if $Q_0'(q_1/q_0 > t) \leq Q_0(q_1/q_0 > t)$ and $Q_1'(q_1/q_0 > t) \geq Q_1(q_1/q_0 > t)$ for all $t \in \mathbf{R}$ and $(Q_0', Q_1') \in \mathcal{P}_0 \times \mathcal{P}_1$. Here q_0 and q_1 denote the Radon-Nikodym derivative of Q_0 and Q_1 with respect to a dominating measure μ. As is well known, a solution to above minimax problems is provided as a threshold test on the least favourable pair likelihood ratio [2], [3]. Thus, identification of the least favourable pair is the key to solving these nonparametric decision problems.

In his 1965 paper [2], P. J. Huber gave the construction of the least favourable pair for both *ϵ-contamination* and *total variation* families. This construction is quite general, in particular, it works in any measurable space.

Later, Huber and Strassen proved their celebrated abstract minimax theorem in [3] and [4]. Here the authors assume E to be a *Polish space* (i.e., a separable, complete metrizable topological space) with associated Borel σ-field $\mathcal{B}$. They consider families of the type $\mathcal{P}_i = \{P \in \mathcal{M} | P \leq \nu_i\}$ where ν_i are set functions defined on $\mathcal{B}$ satisfying

$$\nu_i(\emptyset) = 0, \nu_i(E) = 1, \tag{1}$$

$$A \subset B \text{ implies } \nu_i(A) \leq \nu_i(B), \tag{2}$$

$$A_n \uparrow A \text{ implies } \nu(A_n) \uparrow \nu(A), \tag{3}$$

$$F_n \downarrow F, F_n \text{ closed, implies } \nu_i(F_n) \downarrow \nu_i(F), \tag{4}$$

$$\nu_i(A \cup B) + \nu_i(A \cap B) \leq \nu_i(A) + \nu_i(B). \tag{5}$$

A set function satisfying (1)–(4) is called a *capacity* and a set function satisfying (5) called *2-alternating*. Their theorem establishes the existence of a least favourable pair, but does not give constructions [3, Theorem 4.1]. Moreover, the conditions (1)–(4) imply the weak compactness of $\mathcal{P}_i$ [3, Lemma 2.2].

Define ν_i by either $\nu_i(A) = (1 - \epsilon_i)P_i + \epsilon_i$ for $A \neq \emptyset$, (called ϵ-contamination capacity,) or $\nu_i(A) = \min(P_i(A) + \epsilon_i, 1)$ for $A \neq \emptyset$ (called total variation capacity). If E is compact, the ν_i satisfy (1)–(5) and $\mathcal{P}_i = \{P \in \mathcal{M} | P \leq \nu_i\}$ are either ϵ-contamination or total variation families [3, Example 3, Example 4]. However, if E is not compact, the ν_i do not satisfy (4).

A related discussion can be found in [5]. The author introduces a class of capacities, denoted by special capacities, containing both ϵ-contamination and total variation. For this class, an explicit construction of the least favourable pair is given.

II. Summary

In this paper we revisit the abstract minimax theorem of Huber & Strassen with the goal of removing the weak compactness condition. To do so, we require different topological conditions: We take E to be a locally compact space for which every open set is a K_σ (i.e., a countable union of compacts). This setting includes $\mathbf{R}^N, N < \infty$, and "well-behaved" subsets of $\mathbf{R}^N$ with their relative topology. We allow set functions satisfying (1)–(3), (5) and

$$K_n \downarrow K, K_n \text{ compact, implies } \nu_i(K_n) \downarrow \nu_i(K) \tag{6}$$

instead of (4). Note that both ϵ-contamination and total variation capacities satisfy (6). A set function satisfying (1)–(3) and (6) will be called a *regular Choquet capacity*.

We point out that a regular, 2-alternating Choquet capacity can be extended to the one-point compactification E' of E with the point at infinity in such a manner that the Huber-Strassen construction of a least favourable pair applies to the compactified space. Thus, our work is to construct the appropriate capacity extensions ν_i'. This is done within the setup of the theory of capacities as developed by G. Choquet [1]. Then the $\mathcal{P}_0' = \{P \leq \nu_0'\}$ vs. $\mathcal{P}_1' = \{P \leq \nu_1'\}$ least favourable pair on E' must be related to the original problem $\mathcal{P}_0$ vs. $\mathcal{P}_1$ on E. In particular, there is the issue that the $\mathcal{P}_0'$ vs. $\mathcal{P}_1'$ least favourable pair may put mass at infinity.

The contributions of this paper are as follows. First, we present the extension via one point compactifications as discussed above and argue that the Huber-Strassen construction of the least favourable pair applies to the compactified space. Second, if the ν_i satisfy $\nu_i(A) = \inf\{v_i(O) | A \subset O, E \backslash O \text{ compact in } E\}$, the $\mathcal{P}_0'$ vs. $\mathcal{P}_1'$ least favorable pair will not have mass at infinity, and hence, we obtain the desired $\mathcal{P}_0$ vs. $\mathcal{P}_1$ least favorable pair. Both ϵ-contamination and total variation do indeed satisfy this condition.

References

[1] G. Choquet, "Theory of capacities," *Ann. Inst. Fourier*, **5**, pp. 131–292, 1953/1954.

[2] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, **36**, pp. 1753–1758, 1965.

[3] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, **1**, pp. 251–263, 1973.

[4] P. J. Huber and V. Strassen, "Correction to minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, **2**, pp. 223–224, 1974.

[5] H. Rieder, "Least favorable pairs for special capacities," *Ann. Statist.*, **5**, pp. 909–921, 1977.

Hypothesis Testing for Arbitrarily Varying Source with Exponential-Type Constraint

Fang-Wei Fu and Shi-Yi Shen
Department of Mathematics, Nankai University
Tianjin 300071, P.R.China

The problem of hypothesis testing, which is to decide between two alternative explanations for the observed data, is one of the standard problem in statistics. A discrete memoryless source (DMS) is a sequence of i.i.d random variables. The distribution of the DMS is either P_1 or P_2. When a sample is emitted from the source, the observer attempts to decide which hypothesis of $H_1 : P_1$ or $H_2 : P_2$ is correct. The main concern of this problem is to determine the best asymptotic exponent of the second kind of the error probability when the first kind of the error probability is (1) fixed (2) less than 2^{-nr}. These are specified by (1) the well-known lemma of Stein (2) the theorem of Hoeffding ([1]), Blahut ([2]), Csiszár and Longo ([3]) for hypothesis testing problem with exponential-type constraint.

DMS is an ideal model, A more robust model is arbitrarily varying source (AVS), where the source distribution may vary within a certain set of distribution from one time instant to the next. The varying behavior of the distribution of AVS is not known exactly to us, and there are only two alternatives. We consider the problem of hypothesis testing for AVS in the same way for DMS, and determine the best asymptotic exponent of the second kind of the error probability when the first kind of the error probability is (1) fixed (2) less than 2^{-nr}. These results generalize the well-known lemma of Stein and the theorem of Hoeffding, Blahut, Csiszár and Longo in statistics. As a corollary in information theory, The best asymptotic error exponent and Strassen's theorem for AVS coding are obtained. furthermore, we determine the best asymptotic error exponent and r-optimal rate (the minimum compression rate when the error probability is less than 2^{-nr}, $r \geq 0$) of AVS coding with a fidelity criterion.

Let $\mathcal{W} = \{ W(\bullet \mid s) \mid s \in \mathcal{S} \}$ be a set of probability distributions on $\mathcal{X}$. An AVS defined by $\mathcal{W}$ is a sequence of random variables $\{X_i\}_{i=1}^{\infty}$ such that the distribution of $\underline{X} = (X_1, \ldots, X_n)$ is an unknown element of $\mathcal{W}^n$.

In the problem of hypothesis testing, $\mathcal{W}$ is not exactly known to statistician. There are only two alternative hypotheses for $\mathcal{W}$. Let $\mathcal{W}_i = \{W_i(\bullet \mid s) | s \in \mathcal{S}\}$, $i = 1, 2$ be two sets of probability distributions on $\mathcal{X}$. $H_1 : \mathcal{W} = \mathcal{W}_1$, $H_2 : \mathcal{W} = \mathcal{W}_2$. When a sequence $\underline{x} = (x_1, \ldots, x_n)$ is emitted from the source, the statistician attempts to decide, by observing the data $\underline{x}$, which hypothesis of H_1 or H_2 is correct. The decision rule is characterized by a set $A \subseteq \mathcal{X}^n$. The statistician declares that H_1 is true if $\underline{x} \in A$, and that H_2 is true if $\underline{x} \in A^c$. The first kind of error probability is

$$\alpha = \max_{\underline{s} \in \mathcal{S}^n} W_1^n(A^c \mid \underline{s})$$

The second kind of error probability is

$$\beta = \max_{\underline{t} \in \mathcal{S}^n} W_2^n(A \mid \underline{t})$$

Given $r > 0$, we denote

$$\beta_n(r) = \inf_{\alpha \leq 2^{-nr}} \beta$$

When $|\mathcal{S}| = 1$, it is the problem of hypothesis testing for DMS. for $i = 1, 2$, denote

$$\widehat{\mathcal{W}_i} = \{ \sum_{s \in \mathcal{S}} \lambda_s W_i(\cdot \mid s) \mid 0 \leq \lambda_s \leq 1, \sum_{s \in \mathcal{S}} \lambda_s = 1 \}$$

Theorem 1

$$\lim_{n \to \infty} [-\frac{1}{n} \log \beta_n(r)] = \min_{P \in \widehat{\mathcal{W}_2}} \min_{Q \in \widehat{\mathcal{W}_1}} \min_{\overline{P}: D(\overline{P} \| Q) \leq r} D(\overline{P} \parallel P)$$

here the right term is positive.

References

[1] W. Hoeffding, " Asymptotically optimal tests for multinomial distributions," *AMS,* Vol.36, pp.369–400, 1965

[2] R.E. Blahut, " Hypothesis testing and information theory," *IEEE Trans. Inform. Theory,* Vol.IT–20, pp.405–417, July, 1974

[3] I. Csiszár and G. Longo, " On the error exponent for source coding and for testing simple statistical hypothesis," *Studia Sc. Math. Hungarica,* Vol.6, pp.181–191, 1971

A Simple Formula for the Rate of Maxima in the Envelope of Normal Processes Having Unsymmetrical Spectra

Ali Abdi and Said Nader-Esfahani

Dept. of Elec. & Comp. Eng., University of Tehran
P. O. Box 14395-515, Tehran, Iran
E-Mail: abdi@ece.ut.ac.ir & nader_sa@irearn.bitnet

Abstract -- **In this paper, we present a new formula for rate of maxima in the envelope of a normal process. In contrast to the Rice formula, our result is simple; and also is not limited to a process with even symmetry in its one-sided spectrum.**

I. INTRODUCTION

In the early days of statistical communication theory, pioneering works of Rice, Middleton, Lawson, Uhlenbeck, etc. developed certain fundamental statistical properties of the Normal Process Envelope (NPE). But, still there is a large number of unresolved problems about the NPE. One of these unresolved problems is the rate of maxima in a NPE having unsymmetrical spectrum.

II. RICE FORMULA FOR THE RATE OF MAXIMA

In his classic paper [1], Rice has derived the following formula for a NPE with even symmetry in its one-sided spectrum:

$$N=\frac{(a^2-1)^2}{(2a)^{5/2}}\left(\frac{b_2}{\pi\, b_0}\right)^{1/2}\sum_{n=0}^{\infty}\frac{\Gamma(n/2+5/4)}{\Gamma(n/2+7/4)}\frac{A_n}{a^n} \qquad (1)$$

where:

$$a^2=\frac{b_0\, b_4}{{b_2}^2},\ b=\frac{3-a^2}{2},\ A_n=\sum_{m=0}^{n}\frac{1}{2}\cdots\frac{2m-1}{2}(n-m+1)\frac{b^m}{m!} \qquad (2)$$

In the above formulas, b_n is the n'th spectral moment:

$$b_n=(2\pi)^n\int_0^{\infty}w(f)(f-f_c)^n\,df \qquad (3)$$

where $w(f)$ is the one-sided spectrum of the normal process; and f_c is its center frequency.

The formula (1) is very complicated, and holds only under the assumption of even symmetry for $w(f)$ around f_c.

III. NEW FORMULA FOR THE RATE OF MAXIMA

One way for computing N is to derive an analytic expression for the bivariate joint pdf of the two samples of $R'(t)$, the time derivative of the NPE. But unfortunately, it is very difficult to derive this bivariate pdf assuming an unsymmetric $w(f)$. Thus, we have approximated it by the first few terms of its 2D Hermite polynomials expansion [2].

Using a level-crossing formula developed in [2], and after admittedly very cumbersome calculations, the above approximation for the bivariate pdf yields the following result:

$$N_{aprx}=\frac{1}{2\pi}\left(\frac{b_0\, b_4+3{b_2}^2-4b_1\, b_3}{b_0\, b_2-{b_1}^2}\right)^{1/2} \qquad (4)$$

To examine the degree of accuracy of (4), we considered several spectra with various mathematical forms. For the cases of symmetrical spectra, N was computed using (1). For the cases of unsymmetrical spectra, however, there is no closed form formula in [1] and in the related literature; so N was computed numerically using a very complicated triple integral (This triple integral can be derived from [1]). In all cases, the relative error of (4) was found to be below 5%!, which is really a great success for this formula.

Surprisingly, (4) is exactly like the same result that we have reported in [3]; which is obtained by a completely different approach.

It should be noted that (4) is really an approximate formula and it is not reasonable to determine its accuracy just by several numerical examples. However, our recent results, which will be reported later, show that (4) can be corrected just by multiplying a correction coefficient:

$$N=KN_{aprx} \qquad (5)$$

where K depends on the spectral moments. We observed that for the above examples, K is very close to one.

REFERENCES

[1] S. O. Rice, "Mathematical analysis of random noise," reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed., Dover, 1954.

[2] V. I. Tikhonov and A. A. Tolkachev, "The effect of non-normal fluctuations on linear systems," in *Non-Linear Transformations of Stochastic Processes*, P. I. Kuznetsov, R. L. Stratonovich, and V. I. Tikhonov, Eds, Pergamon, 1965.

[3] S. Nader-Esfahani and A. Abdi, "A new formula for the mean number of the peaks of a normal process envelope," in *Proc. Iranian Conf. Elec. Eng.*, Tarbiat-Modarres University, Tehran, Iran, 1994, vol. 5, pp. 450-456 (in Persian).

Measuring Time-Frequency Information and Complexity Using the Rényi Entropies

Richard G. Baraniuk,* Patrick Flandrin,◇ and Olivier Michel◇ [1]

*Department of Electrical and Computer Engineering
Rice University
Houston, Texas, USA

◇Laboratoire de Physique (URA 1325 CNRS)
Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France

Abstract — **In search of a nonparametric indicator of deterministic signal complexity, we link the Rényi entropies to time-frequency representations. The resulting measures show promise in several situations where concepts like the time-bandwidth product fail.**

I. Introduction

The term *component* is ubiquitous in the signal processing literature. Intuitively, a component is a concentration of energy in some domain, but this notion is difficult to translate into a quantitative concept. In fact, the concept of a signal component has never been — and may never be — clearly defined. In this paper, rather than address the question "what is a component?" directly, we investigate a class of quantitative measures of deterministic signal *complexity* and *information content*. While they do not yield direct answers regarding the locations and shapes of components, these measures are intimately related to the concept of a signal component, the connection being the intuitively reasonable supposition that signals of high complexity (and therefore high information content) must be constructed from large numbers of elementary components.

Our approach to complexity is based on entropy functionals and exploits the powerful analogy between deterministic signal energy densities and probability densities. For example, the Wigner time-frequency representation (TFR), $W_s(t,f) = \int s\left(u+\frac{\tau}{2}\right) s^*\left(u-\frac{\tau}{2}\right) e^{-j2\pi\tau f} d\tau$, which indicates the joint time-frequency content in a signal s, marginalizes to the time and frequency energy densities $\int W_s(t,f)\,df = |s(t)|^2$ and $\int W_s(t,f)\,dt = |S(f)|^2$. The TFRs $C_s(t,f)$ of Cohen's class form an infinite set of generalizations of the Wigner TFR.

The probabilistic analogy evoked by the marginals suggests the Shannon entropy $H(C_s) = -\iint C_s(t,f)\log_2 C_s(t,f)\,dt\,df$ as a natural candidate for estimating the complexity of a signal through its TFR: The peaky TFRs of signals comprised of small numbers of elementary components would yield small entropy values, while the diffuse TFRs of more complicated signals would yield large entropy values. Unfortunately, however, the negative values taken on by the Wigner distribution and most other Cohen's class TFRs prohibit the application of the Shannon entropy due to the logarithm.

II. The Rényi Entropies

We propose to sidestep this negativity issue by employing the Rényi entropies [1,2] $H_\alpha(C_s) = \frac{1}{1-\alpha}\log_2 \iint C_s^\alpha(t,f)\,dt\,df$, which generalize the Shannon entropy to a family parameterized by $\alpha > 0$. The resulting time-frequency information measure has a number of attractive properties. In addition to immunity to the negative TFR values that invalidate the Shannon approach [2], the third-order Rényi entropy measures signal complexity [1,2]: The information $H_3(C_s)$ in the TFR of the sum $s(t) = g(t) + g(t-T)$ of two separated signal components saturates (as $T \to \infty$) exactly one bit above the value $H_3(C_g)$ for a single component.

Our goal has been a detailed study of the properties and applications of these promising complexity measures, with emphasis on establishing a firm mathematical foundation. Interesting properties include the following [2]:

1. For integer orders $\alpha > 1$, $H_\alpha(C_s)$ is defined for essentially all key TFRs, including even those distributions taking locally negative values.
2. For odd orders $\alpha > 1$, $H_\alpha(C_s)$ is asymptotically invariant to TFR "cross-components" and therefore does not count them.
3. $H_\alpha(W_s)$ exhibits extreme sensitivity to phase differences between closely spaced components (ameliorated by time-frequency smoothing).
4. The range of $H_\alpha(W_s)$ values is bounded above and below. A single Gaussian pulse attains the lower bound, while "deterministic white noise" nears the upper bound.
5. The value of $H_\alpha(W_s)$ is invariant to arbitrary time and frequency shifts, scale changes, and shears and rotations in the time-frequency plane.

In recent work, we have applied the Rényi measures to random signals, introduced the notion of a Rényi dimension, and suggested how these measures can be employed to improve TFR performance through adaptivity.

Finally, we have introduced a new "Jensen-like" divergence measure [3]. While this quantity promises to be a useful indicator of the distance between two time-frequency distributions, it is currently limited to the analysis of positive definite TFRs. In spite of this rather severe limitation, this measure could prove useful for time-frequency based detection and recognition.

[1] Supported by NSF Grant MIP–9457438, ONR Grant N00014–95–1–0849, Texas ATP Grant 003604–002, and URA 1325 CNRS.

References

[1] W. Williams, M. Brown and, A. Hero, "Uncertainty, Information, and Time-Frequency Distributions," *SPIE 1566*, 1991.

[2] P. Flandrin, R. Baraniuk, and O. Michel, "Time-Frequency Complexity and Information," *Proc. IEEE ICASSP '94*, 1994.

[3] O. Michel, R. Baraniuk, and P. Flandrin, "Time-Frequency Based Distance and Divergence Measures," *Proc. IEEE Symp. Time-Frequency Analysis*, 1994.

Random Fields and their Wavelet Transforms and Representation: Covariance and Spectral Properties

E. Masry[1]
Electrical & Computer Engineering
University of California at San Diego
La Jolla, CA 92093, U.S. A.

Abstract — **The covariance and spectral properties of the wavelet transform and of the discrete wavelet coefficients, in the orthonormal series representation, of second-order random fields on R^n are determined. Both weakly homogeneous random fields as well as random fields with weakly homogeneous increments are considered. Weakly isotropic fields and fields with weakly isotropic increments are also considered. Applications to fractional Brownian fields on R^n are given.**

I. Introduction

Let $X = \{X(\underline{t},\omega), \underline{t} \in R^n\}$ be a possibly complex-valued random field which is jointly measurable in t and ω. We consider second-order random fields with zero mean and covariance function $C_X(\underline{t},\underline{s}) = E[X(\underline{t})X^*(\underline{s})]$ where $*$ denotes complex conjugate. Let $\psi(\underline{t}), \underline{t} \in R^n$, be an analyzing wavelet. The continuous wavelet transform of the random field X at scale $a > 0$ is defined by

$$W_a(\underline{t},\omega) = a^{-n/2} \int_{R^n} X(\underline{u},\omega)\psi((\underline{u}-\underline{t})/a))d\underline{u} \qquad (1)$$

so that $\{W_a(\underline{t},\omega), \underline{t} \in R^n\}$ is a random field for each scale $a > 0$.

Let $\{V_j, j \in Z\}$ be a multiresolution approximation of $L_2(R^n)$ and W_j the orthogonal complement of V_j in V_{j+1}. Let $\{\phi_{l,\underline{k}}(\underline{t}), \underline{k} \in Z^n\}$ be an orthonormal basis for V_l and let $\{\psi_{p,j,\underline{k}}(\underline{t}), p = 1,...,2^n-1, \underline{k} \in Z^n\}$ be an orthonormal basis for W_j [2]. Define the approximation coefficients $\{a_{l,\underline{k}}, \underline{k} \in Z^n\}$ at resolution 2^{-l} by

$$a_{l,\underline{k}}(\omega) = \int_{R^n} X(\underline{t},\omega)\phi_{l,\underline{k}}(\underline{t})d\underline{t} \qquad (2)$$

and the detail coefficients $\{b_{p,j,\underline{k}}, \underline{k} \in Z^n\}$ at detail level 2^{-j} by

$$b_{p,j,\underline{k}}(\omega) = \int_{R^n} X(\underline{t},\omega)\psi_{p,j,\underline{k}}(\underline{t})d\underline{t}. \qquad (3)$$

Under certain integrability conditions (see [1] for details), $\{a_{l,\underline{k}}, \underline{k} \in Z^n\}$ and $\{b_{p,j,\underline{k}}, \underline{k} \in Z^n\}$ are discrete-time second-order random fields on Z^n.

Our goal is determine the covariance and spectral properties of the random fields $\{W_a(\underline{t}), \underline{t} \in R^n\}$, $\{a_{l,\underline{k}}, \underline{k} \in Z^n\}$ and $\{b_{p,j,\underline{k}}, \underline{k} \in Z^n\}$ and to see whether they inherit the features of the input process X (weakly homogeneous, weakly homogeneous increments, weakly isotropic). A representative result is given in Section II.

II. Representative Result

We suppress the ω-argument in the sequel. Consider a possibly complex-valued measurable random field $\{X(\underline{t}), \underline{t} \in R^n\}$ with weakly homogeneous increments [3].

We are concerned with the covariance and spectral properties of the wavelet transform and approximation and detail coefficients, as defined in (1), (2), and (3), respectively, of the random field $\{X(\underline{t}), \underline{t} \in R^n\}$ itself (*not* of its increments).

Theorem 1. Assume that $\int_{R^n} \psi(\underline{t})d\underline{t} = 0$. Then the wavelet transforms $\{W_a(\underline{t}), \underline{t} \in R^n\}$, $a > 0$, are ***jointly weakly homogeneous*** random fields with zero means and covariance/cross-covariance function

$$C_{W_{a_1},W_{a_2}}(\underline{t}) \equiv E[W_{a_1}(\underline{t}+\underline{u})W^*_{a_2}(\underline{u})]$$

having the spectral representation

$$C_{W_{a_1},W_{a_2}}(\underline{t}) = (a_1a_2)^{n/2} \int_{R^n\setminus\{\underline{0}\}} e^{i\underline{t}\cdot\underline{\lambda}}\, \tilde{\psi}^*(a_1\underline{\lambda})\, \tilde{\psi}(a_2\underline{\lambda})\, F(d\underline{\lambda})$$
$$+(a_1a_2)^{1+n/2} \int_{R^n}\int_{R^n} (A\underline{u})\cdot\underline{v}\, \psi(\underline{u})\psi(\underline{v})d\underline{u}d\underline{v} \qquad (4)$$

where $F(d\underline{\lambda})$ is a measure on $R^n\setminus\{\underline{0}\}$ satisfying

$$\int_{R^n\setminus\{\underline{0}\}} \frac{\|\underline{\lambda}\|^2}{1+\|\underline{\lambda}\|^2} F(d\underline{\lambda}) < \infty, \qquad (5)$$

$R^n\setminus\{\underline{0}\}$ denotes the Euclidean space R^n minus the vector $\underline{0}$, $A = [\alpha_{ij}]$ is a nonnegative definite Hermitian matrix, and $\tilde{\psi}(\underline{\lambda})$ is the Fourier transform of $\psi(\underline{u})$.

Remark 1. Note that while the input field X is *not* weakly homogeneous, the wavelet transforms at distinct scales are jointly weakly homogeneous. Their spectral and cross-spectral distributions can be obtained from (4). When the the first-order moments of ψ vanish, the second term on the right side of (4) is equal to zero.

Results analogous to Theorem 1 are given for the discrete-time second-order random fields $\{a_{l,\underline{k}}, \underline{k} \in Z^n\}$ and $\{b_{p,j,\underline{k}}, \underline{k} \in Z^n\}$. Applications to fractional Brownian fields on R^n are also given. Full details can be found in [1].

References

[1] E. Masry, "Covariance and spectral properties of the wavelet transform and discrete wavelet coefficients of second-order random fields," submitted for publication in *IEEE Trans. Inform. Theory* , September 1994.

[2] Y. Meyer. *Wavelets and Operators.*. Cambridge: Cambridge Univ. Press, 1992.

[3] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions. I: Basic Results.* New York: Springer 1987.

[1] This work was supported by ONR Grant N00014-90-J-1175.

Finite Field Wavelet Transforms and Multilevel Error Protection

Sandip Sarkar and H. Vincent Poor[1]
Dept. of Electrical Engineering, Princeton University, Princeton, NJ 08544

Abstract — **In this paper, a technique for providing unequal error protection is investigated. It relies on a transform approach to coding and makes use of the wavelet transform over finite fields.**

I. Introduction

This paper deals with the application of finite field wavelets to build error correcting codes that provide unequal error protection to the codewords. Traditional coding theory usually constructs codes providing uniform error protection to all codewords. But, many image and speech processing applications require some codewords to be more protected than others. Examples of this kind include Differential Pulse Code Modulation (DPCM), where the effect of an error on the most significant bit (MSB) is much more than on the least significant bit (LSB). Similarly, in Linear Predictive Coding (LPC), a technique often used for speech transmission, the filter coefficients are much more important than the raw data. One way to provide additional error protection to some of the codewords is to give all codewords the highest protection required for any data, but this is not bandwidth efficient. Additional error protection calls for more redundancy, leading to a lower rate.

II. Finite Field Wavelet Transforms

A general theory of multiresolution analysis can be developed (cf. [1]) over $L^2(\mathbb{R})$. In this paper, we will only use finite length cyclic wavelet transforms as described in [2], [3] and [4]. We will refer to the mother wavelet in such a formulation by g and the 2-circulant [5] matrix generated by it to be G. Similarly, we have the complementary matrix H. (See [3] for details.)

III. Design of the Code

Transform domain study of codewords have been of great interest [6]. To use wavelet transforms to design codes, we make use of the fact that by choosing a mother wavelet properly, for a wide class of codes, codewords that have a zero bandpass coefficient have non-zero lowpass coefficients. In the successive transform levels, only a few codewords still yield non-zero coefficients and hence can be protected more. An example of this kind of a code is the extended Hamming [8,4] code { i.e. a Hamming (7,4) code with a parity bit } with the mother wavelet $(1\ -1\ 0\ 0\ 0\ 0\ 0\ 0)$. We will call this the Haar wavelet transform and call the generated matrix G_H.

IV. Reed-Muller Codes

Reed-Muller codes can in general be represented as Boolean functions completely specified by specifying a set of basis vectors. A first order Reed-Muller code uses only the first order terms, while an n^{th} order code uses product terms up to order n. Of course, if the length of the codewords are 2^n, only codes of up to order n exist. The matrix G_H of appropriate order works well for these codes, as do some other wavelets derived from a set of codewords.

[1] This research was supported by the US Office of Naval Research under Grant N00014-94-1-0115

V. Concatenated Codes

Concatenated codes can be dealt with in general. As an example, let $\mathcal{C}$ be an (n,k) code. Then, we form a $(2n, 2k)$ code of the form (C, C) by concatenating two codewords, where $C \in \mathcal{C}$. It is easy to see that if G is the 2-circulant matrix formed from the mother wavelet that works for the code $\mathcal{C}$, then $G \otimes I_2$ will work for the concatenated code.

As the next level of complexity, let us consider a code $\mathcal{C}'$ generated from a linear code $\mathcal{C}$ in the following manner : Let A be the generator matrix for the code $\mathcal{C}$. Then the generator matrix for the code $\mathcal{C}'$ is given by

$$A' = \begin{pmatrix} A & A & 0 \\ 0 & A & A \end{pmatrix}$$

Thus, any codeword in $\mathcal{C}'$ is of the form $(c_1, c_1 + c_2, c_2)$, where $c_1, c_2 \in \mathcal{C}$. With the assumption that the code $\mathcal{C}$ is linear, we get $c_1 + c_2 \in \mathcal{C}$. Hence, the matrix $G \otimes I_3$ works for this case. It is now obvious how to deal with any concatenated code of this form. In fact, depending on applications, we can choose a proper transform to achieve the required bit-rate.

VI. The Dual Code

Direct sum of two different codewords can be handled by using the fact that if $\mathcal{C}$ and $\mathcal{D}$ are two codes, then $(\mathcal{C}+\mathcal{D})^{\perp} = \mathcal{C}^{\perp} \cap \mathcal{D}^{\perp}$ (see [7]). The idea is to use codes that have as a subcode a self-dual code. Then, using the above property, if we can find a mother wavelet in the intersection, a description can be obtained. Even-weight repetition codes are an example of such kind of codes.

VII. Future Directions

It is thus possible to characterize a large class of codes. The next step of complexity is in finding descriptions of codes that are direct sums of two other codes. This is useful in finding a description for the Golay code.

Acknowledgements

We are grateful to Dr. A. R. Calderbank of AT&T for his comments and valuable suggestions.

References

[1] C. K. Chui, editor, *An Introduction to Wavelets*, Academic Press, 1991.

[2] S. G. Mallat, "A theory for multiresolution signal decomposition : The wavelet representation", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[3] R. L. Grossman G. Caire and H. V. Poor, "Wavelet transforms associated with finite cyclic groups", *IEEE Trans. Inform. Theory*, vol. 39, pp. 1157–1166, July 1993.

[4] S. Sarkar and H. V. Poor, "Certain generalizations of the cyclic wavelet transform", *Proc.. 1995 Conf. Inform. Sci. Syst.*, The Johns Hopkins University, Mar. 22 - 24, 1995.

[5] P. J. Davis, *Circulant Matrices*, John Wiley, New York, 1979.

[6] R. E. Blahut, *Theory and Practice of Error Correcting Codes*, Addison-Wesley, 1984.

[7] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, New York, 1977.

Galois Theory and Wavelet Transforms

Andreas Klappenecker[1] and Thomas Beth
Universität Karlsruhe, Institut für Algorithmen und Kognitive Systeme,
D–76 128 Karlsruhe, Germany, e-mail: klappi@ira.uka.de

Abstract — **Computing the Fast Wavelet Transform of rational input sequences using *algebraic* scaling coefficients affords only a *finite extension field* K over $\mathbb{Q}$ rather than the field of complex numbers. We use Galois theoretic methods to study this extension field.**

I. Introduction

Orthonormal wavelet bases are usually constructed by the tools of multiresolution analysis, cf. [2]. At the heart of a multiresolution analysis stands a so-called *scaling function* φ. This scaling function satisfies a dilation equation, which can be written in Fourier space as $\hat{\varphi}(\omega) = m_0(\omega/2)\,\hat{\varphi}(\omega/2)$, where $m_0(\omega) = \sum h_n\, e^{-in\omega}$. In what follows, we assume compactly supported scaling functions with algebraic coefficients h_n, i. e., every coefficient h_n is element of an algebraic number field. From the multiresolution analysis axioms one derives the simple relation $|\,m_0(\omega)\,|^2 + |\,m_0(\omega+\pi)\,|^2 = 1$. Therefore, it is convenient to construct the transfer function $m_0(\omega)$ from its squared modulus $|\,m_0(\omega)\,|^2$ with the help of the following:

Theorem 1 (Fejér-Riesz) *Let $A(\omega)$ be a real nonnegative even trigonometric polynomial*

$$A(\omega) = \sum_{m=0}^{M} a_m \cos m\,\omega, \quad \textit{with} \quad a_m \in \mathbb{R}.$$

Then it is possible to construct a real trigonometric polynomial $B(\omega) = \sum_{m=0}^{M} b_m\, e^{im\omega}$, with $b_m \in \mathbb{R}$, of the same order M, such that $A(\omega) = |\,B(\omega)\,|^2$.

II. Algebraic Scaling Coefficients

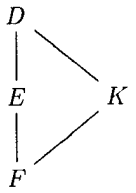
In the case of trigonometric polynomials $|\,m_0(\omega)\,|^2$ with algebraic coefficients, the following theorem ensures that $m_0(\omega)$ has algebraic coefficients, too.

Theorem 2 ([1]) *The coefficients a_m of the trigonometric polynomial $A(\omega)$ are algebraic if and only if the coefficients b_m of $B(\omega)$ are also algebraic.*

Theorem 2 can be proved by extending Daubechies' proof of Theorem 1 [2], but using minimal splitting fields instead of the algebraically closed field $\mathbb{C}$. The main steps in the proof can be sketched as follows:

1. Rewrite the trigonometric polynomial $A(\omega)$ as a polynomial p_A in $\cos\omega$. The polynomial p_A can be factorized over a minimal splitting field E as $\mathrm{lc}(p_A)\prod_{n=1}^{M}(c - c_j)$. Here, $\mathrm{lc}(\,\cdot\,)$ denotes the leading coefficient.
2. Build a self-reciprocal polynomial P_A by substituting $c := (z + z^{-1})/2$ in $p_A(c)$ and multiplying with z^M. Therefore, the resulting polynomial is of the following form $P_A(\omega) = \mathrm{lc}(p_A)\prod_{n=1}^{M}\left(1/2 - c_j z + 1/2z^2\right)$. Factorize $P_A(z)$ in a minimal splitting field D.
3. Choose a zero z_j from every factor $\left(1/2 - c_j z + 1/2z^2\right)$, $1 \le j \le M$, and build a new trigonometric polynomial $P_B(z) = \nu\prod_{j=1}^{M}(z - z_j)$, where $\nu \in K$ is just a normalization factor. The trigonometric polynomial $B(\omega)$ is obtained from P_B by $B(\omega) = P_B(e^{-i\omega})$. Thus, the field K is generated by elementary symmetric functions of the zeros z_j.

Hence, from a field theoretic point of view the situation can be summarized by the following diagram:



III. Galois Theoretic Analysis

From the very construction, we see that the fields E and D are Galois extensions over F. We discuss some of their properties through a sequence of lemmas and corollaries.

Lemma 1 *The Galois group $\mathrm{Gal}(D/E)$ is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^m$, with $m \le M$.*

From this observation we easily derive the following result about the structure of the Galois group.

Lemma 2 *The Galois group $\mathrm{Gal}(D/F)$ is the extension of the elementary abelian normal 2-subgroup $\mathrm{Gal}(D/E)$ by the group $\mathrm{Gal}(E/F)$.*

As a consequence, we get an upper bound for the order of the Galois group $\mathrm{Gal}(D/F)$, which is helpful in the estimation of this group.

Corollary 3 *We have the following upper bound for the field degree of D/F:*

$$[D : F] \le 2^M \cdot |\,\mathrm{Gal}(E/F)\,| \le M! \cdot 2^M.$$

By carefully studying the structure of K, we obtain

Lemma 4 *The field D is generated by the composition field EK.*

Corollary 5 *The field degree $[K:F]$ is at least $|\,\mathrm{Gal}(D/E)|$.*

The close connection between the fields D and K can be exemplified by the following

Lemma 6 *If the field degree D/E is maximal, i. e., $[D : E] = 2^M$, then the Galois closure of K is the field D.*

[1] This work was supported by DFG under project Be 877/6-2.

References

[1] T. Beth, A. Klappenecker, and A. Nückel. Construction of algebraic wavelet coefficients. *Proc. ISITA '94, Sydney*, 1994.

[2] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Reg. Conf. Series Appl. Math. SIAM, 1992.

The Discrete-Time Biorthogonal Wavelet Transform

Manuel Á. Sola and Sebastià Sallent
Applied Mathematics and Telematics Dept., Gran Capità s/n UPC C3, 08034 Barcelona, Spain

Abstract — **In this paper we present a methodology to analyze functions in ℓ^2 in terms of self-similar discrete-time biorthogonal functions at different resolution levels. We call these functions *discrete-time biorthogonal wavelets*, and they verify in ℓ^2 the same properties that biorthogonal wavelets do in L^2, including self-similarity.**

I. Introduction

One of the most well-known cases of Multiresolution Analysis (MA) [1] for functions $f \in L^2(\mathrm{R})$ is characterized by solutions of the equation

$$\phi(x) = \sum_k a_k \phi(2x - k) \tag{1}$$

with $\phi(x) \in L^2$, $a_k \in \mathrm{R}$, $k \in \mathrm{Z}$.

The family $\phi_{mk}(x) = 2^{-m/2}\phi(2^{-m}x - k)$ with $k, m \in \mathrm{Z}$, is a powerful tool to analyze the behaviour of functions at different locations and resolutions. As $\phi(x)$ is defined on L^2, it is not possible to apply this theory directly on a discrete-time signal $g \in \ell^2$. For discrete functions, it is necessary first to build $f \in L^2$ from g, and then apply a MA on f to study its properties and, from them, extrapolate the properties of g. In this paper we overcome these drawbacks developing the theory of wavelets directly on ℓ^2 and giving conditions to obtain families of discrete-time biorthogonal wavelets.

II. Discrete Multiresolution Analysis

We define a Discrete Multiresolution Analysis (DMA) V by a set of closed subspaces $V_m \subset \ell^2(\mathrm{R})$, $m \in \mathrm{N}$, $\ldots \subset V_2 \subset V_1 \subset V_0 = \ell^2$, where $\bigcap_m V_m = \{0\}$, and where each subspace verifies that $\phi_m \in V_m$ exists so that the set of functions $\{\phi_{mk}\}_{k \in Z}$ is a base for V_m, with $\phi_{mk}[n] = \phi_m[n - 2^m k]$, $n \in Z$. A direct consequence of this definition is the relationship between basis functions from adjacent subspaces given by $\phi_{m+1} = \sum_k a_k^m \phi_{mk}$, $a_k^m \in \mathrm{R}$.

We introduce the self-similarity criterion among functions at different resolutions levels stating that a DMA is self-similar (SSDMA) if and only if

$$\forall m \in N, f[n] \in V_{m+1} \Leftrightarrow f[2n] \in V_m. \tag{2}$$

In [2] we prove that all SSDMA's must be homogeneous ($a^m = a^0, m > 0$) and can be obtained from $\phi^0, a^0 \in \ell^2$ solutions of

$$\phi_0[n] = \sqrt{2} \sum_k a_k^0 \phi_0[2n - k].$$

We will call that equation *discrete two-scale equation* due to its analogy with (1). Under certain conditions of convergence, an SSDMA leads to an MA [2]. Techniques appearing in [3] for increasing regularity of $\phi(x)$ can be applied to study the regularity of the MA generated by an SSDMA.

III. Biorthogonality

Let V and $\tilde{V}$ be two DMA's. We say that V is biorthogonal to $\tilde{V}$ if and only if V_m is biorthogonal to $\tilde{V}_m$ for $m > 0$, that is, if and only if ϕ_m is biorthogonal to $\tilde{\phi}_m$, with the scalar product as the projection operator. A necessary and sufficient condition for V_m being biorthogonal to $\tilde{V}_m$ is that ϕ_1 be biorthogonal to $\tilde{\phi}_1$, and that a^m be biorthogonal to $\tilde{a}^m$. If V and $\tilde{V}$ have to be self-similar, they must verify (2). Let $f \in l^2$, and suppose that f is projected on V. As V and $\tilde{V}$ are biorthogonal, f can be reconstructed from $\tilde{V}$ and from the projections of f on V. If f must be decomposed in terms of self-similar functions, at least $\tilde{V}$ have to be self-similar. If V is not forced to be self-similar, there will be more degrees of freedom to design the families of biorthogonal discrete-time wavelets.

A special case of discrete-time biorthogonal wavelet can be obtained when a^0 is an interpolation function. From the relationship of this case with filter bank theory, one can obtain simple filter bank structures verifying the perfect reconstruction property, solving the drawbacks that interpolation filters present in this context [4], and pointing interesting applications in areas such as multiresolution image and video coding.

IV. Generalization of the Self-Similarity Criterion

The self-similarity condition given in (2) can be extended to

$$f[2^{\beta+\delta}n] = 2^{-\delta/2} g[2^\beta n], \beta \in N, \delta \in N^+, f \in V_{m+\delta}, g \in V_m.$$

The most restrictive case, that is, the case that would imply more constrains on discrete-time wavelets due to self-similarity, corresponds to $\beta = 0$ and leads to (2). Functions in SSDMA's with different β's will have different grade of self-similarity (GSS). An expression that measures this property for a given β results to be $GSS = 2^{-\beta}$. When relaxing the self-similarity criterion, the design of families of functions is also made more flexible. Constrain (2) can be generalized for a integer scaling factors greater than 2. Then, one can obtain SSDMA's defined by discrete-time multiwavelets on which biorthogonality conditions similar to those given in the former section can be imposed.

References

[1] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commu. Pure and Appl. Math.*, vol.XLI, pp.909–996, 1988.

[2] Manuel Á. Sola and Sebastià Sallent, "A theory for discrete self-similar functions in a discrete multiresolution analysis", *Technical Report*, Dept. of Applied Mathematics and Telematics, UPC, Barcelona, Spain, May 1995.

[3] I. Daubechies, J. C. Lagarias, "Two-scale difference equations II. Local regularity, infinite products of matrices and fractals," *SIAM J. Math. Anal.*, vol.23, no.4 , pp.1031–1079, 1992.

[4] M. Unser, "Efficient dyadic wavelet transformation of images using interpolation filters," in Proc. *ICASSP-93*, vol.V, pp.149–152, 1993.

Compression of Square Integrable Functions: Fourier vs Wavelets

R.E.Krichevskii, *Member, IEEE*, V.N.Potapov[1]
Math. Institute and State University, 630090 Novosibirsk, Russia

Abstract — **P_α is the class of functions with α-th derivative bounded in L_2-norm, $\alpha > 0$. Kolmogorov and Tichomirov have ε-specified any $f \in P_\alpha$ by a $O(\varepsilon^{-1/\alpha})$ bits length code obtained from the Fourier (trigonometric) spectrum of f. We prove that the code can be derived from f in linear time. We show that wavelets are equivalent to the trigonometric basis with respect to both the length of the code and the time to get it from the spectrum (to within multiplicative constants). On the other hand, some bases of wavelets outperform Fourier's, if we want to find the value of f at some point given the code of f.**

I. Introduction

A.Kolmogorov and V.Tichomirov in collaboration with V.Arnold introduced in [1] a compact class P_α of square integrable functions, $\alpha > 0$. A 2π-periodic function f, $f \in P_\alpha$, belongs to P_α, if

$$\int_0^{2\pi} |f(t)|dt \le 1, \int_0^{2\pi} |f^{(\alpha)}(t)|^2 dt \le 1 \ ,$$

where $f^{(\alpha)}$ is the α-th derivative of f, $\alpha > 0$. Every f was given a binary code through which one can recover f with ε-accuracy, $\varepsilon \to +0$, in L_2-norm. The length of the code was minimal (to within a multiplicative constant) and equal to the ε-entropy of P_α, which is $O(\varepsilon^{-1/\alpha})$. A function f was first expanded in trigonometric (Fourier) series. A partial sum of the series is a polynomial differing from f by ε in L_2-norm. The set of the coefficients of the polynomial is called the harmonic ε-spectrum of f. Kolmogorov-Tichomirov's code of f is a compressed form of that spectrum.

With the minimal code known, the next question arises: how difficult is it to go from f to its code and back?

An orthonormal basis is chosen in L_2. A function f is specified by it's ε-spectrum over that basis. There are two variants of the above question. The first: we want to know the running time of computer's transforming the ε-spectrum of f to a code of length $O(\varepsilon^{-1/\alpha})$-bits, ε-specifying f. We want to know also the running time of computer's transforming the code back to the ε-spectrum. The second: we want to know the running time of computer's transforming a code of length $O(\varepsilon^{-1/\alpha})$-bits of f and a number $x, 0 \le x \le 2\pi$ to $f(x)$. I.e., what is the time required to compute a value of f via a code of f? Our purpose is to find out which basis is best suited for solving that question. We will compare the wavelet bases with Fourier's.

II. Main Results

We give the following answers to those two variants of the question. The first variant: we develop a simple algorithm that takes an independent on ε number of operations with bits per an input bit to transform the ε-spectrum of a function into its code of length $O(\varepsilon^{-1/\alpha})$-bits. We call the algorithm simplex, not to be confused with the known Dantzig's simplex method. It is optimal to within the constants in O and in the number of the operations with bits. The same is true for the inverse algorithm. There is a wavelet basis which is as good in solving the first variant of the question as the Fourier's, although the constant in O is greater for wavelets. So, as regards the spectrum-code transformation wavelets are equivalent to the trigonometric basis in the sense mentioned.

As regards the calculation of functions via codes of their spectra (second variant of the above question), wavelets outperform the trigonometric basis. Namely, it takes either $O(\varepsilon^{-1/\alpha}(\log 1/\varepsilon)^c)$, $c > 0$, or $O((\log 1/\varepsilon)^3)$ operations with bits to compute $f(x)$ given a $O(\varepsilon^{-1/\alpha})$ code of f, depending on which spectrum the code is based on: Fourier's or wavelet's.

The simplex code plays an important part in our construction. First of all it is used to enumerate vectors with integer coordinates belonging to a multidimensional simplex. Then the code is applied to a ball and to an ellipsoid. Both the length and the running time of the simplex code are minimal. Moreover, one can recover a sole coordinate $x_i, i = 1, ..., p$ of a vector $(x_1, ..., x_p), p \ge 0$ rather quickly. This property of the simplex code is combined with the fact that there are not so many wavelets not vanishing at a point. As a result, we calculate $f(x)$ rapidly if we use the simplex code of a wavelet expansion of f. The wavelet basis is selected for the class P_α. On the contrary, in the trigonometric case we should use all the members of the trigonometric polynomial.

One of the open questions: what is the tradeoff between the length of codes of functions in P_α and the time required to compute either the code of f or $f(x)$, $0 \le x \le 2\pi$ given the spectrum of f?

References

[1] A.N.Kolmogorov, V.M.Tichomirov, "ε-entropy and ε-capacity of Sets in Metric Spaces," *Uspechi Math. Nauk* , vol. 14(2), pp. 3-86, 1959. (Rus).

[2] S.Mallat, "Multiresolution approximation and wavelet orthonormal bases of L^2," *Trans. Amer. Math. Soc.*, vol.315, pp. 69-87, 1989.

[3] I.Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol.41, pp. 909-996, 1988.

[4] I.Daubechies, *Ten Lectures on wavelet,* CBMS Lecture Notes nr.61, SIAM, Philadelphia, 1990.

[1]This work was supported by Grant NQ8000 from the ISF

Generalized Vector Quantization: Jointly Optimal Quantization and Estimation [1]

Ajit Rao, David Miller, Kenneth Rose, and Allen Gersho
Center for Information Processing Research
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

Given a pair of random vectors $\mathbf{X}, \mathbf{Y}$, we study the problem of finding an efficient or optimal estimator of $\mathbf{Y}$ given $\mathbf{X}$ when the range of the estimator is constrained to be a finite set of values. A *generalized vector quantizer* (GVQ), with input dimension k, output dimension m, and size N maps input $\mathbf{X} \in \mathcal{R}^k$, to output $V(\mathbf{X}) \in \mathcal{R}^m$. The output $V(\mathbf{X})$ is constrained to be one of the *estimation codevectors* in the *codebook*, $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$. The performance of the GVQ is measured by the average distortion, $D = E[d(\mathbf{Y}, V(\mathbf{X}))]$ for a suitable output-space distortion measure $d(\cdot, \cdot)$. A GVQ reduces to a conventional vector quantizer in the special case where $\mathbf{X} = \mathbf{Y}$. The GVQ problem has been approached in the information theory literature from many different standpoints. In particular, it appears in the context of noisy source coding, which is the special case where we quantize $\mathbf{X}$, the observable, noisy version of a source, $\mathbf{Y}$.

A GVQ partitions the input space $\mathcal{R}^k$ into N decision regions or *cells*. Each cell is mapped by the GVQ to a particular codevector. In principle, a GVQ is fully characterized by specifying (a) the input space partition and (b) the codebook. Correspondingly, one can view the GVQ operation as the composition of two operations, an *encoder*, $\mathcal{E}$, which assigns an index i to each input vector $\mathbf{X}$, and a *decoder*, $\mathcal{D}$, which is a table-lookup operation that generates $\mathbf{y}_i$, given i. Thus, $\mathcal{E}$ is a classifier whose performance measure is the distortion in $\mathbf{Y}$ induced by the classification, and $\mathcal{D}$ is the conditional estimator of $\mathbf{Y}$, given the classification index assigned by $\mathcal{E}$. We summarize the necessary conditions and properties of the optimal GVQ. However, the optimal encoder has, in general, unmanageable complexity since its partition regions may be neither convex nor connected. We propose therefore, to constrain the complexity of the encoder, $\mathcal{E}$ by restricting its structure. Finding the optimal GVQ subject to the structural constraint is a hard optimization problem and to address it, we apply ideas from statistical physics. Although the approach we propose is extendible to a variety of structures, we restrict our derivation to the specific structure of the *multiple prototype classifier* and we refer to such a GVQ system as the multiple-prototype generalized vector quantizer (MP-GVQ). In MP-GVQ, a codevector, $\mathbf{y}_j$ owns M_j prototypes, $\{\mathbf{x}_{j1}, \mathbf{x}_{j2} \ldots \mathbf{x}_{jM_j}\}$. The encoding rule finds the nearest prototype to the input $\mathbf{X}$ and maps it to the estimation vector associated with that prototype. Thus, the encoder partition region R_j is the union of M_j nearest neighbor Voronoi cells.

The MP-GVQ design problem is to *jointly* optimize the prototypes $\{\mathbf{x}_{jk}\}$ and codevectors $\{\mathbf{y}_j\}$ to minimize the distortion, D. The problem cannot be directly solved with a variant of Lloyd's algorithm nor by a gradient descent approach, due to the discrete nature of the classifier partition. We tackle the problem by introducing a probabilistic framework for the encoding rule where, for a given input, a probability distribution is assigned to the set of prototypes and the estimation vector assigned to the input is determined by the class index of the randomly chosen prototype. The degree of randomness is measured by the Shannon entropy. Randomization of the nearest-neighbor partition subject to a constraint on the encoder entropy results in the Gibbs distribution for the encoding rule. The Lagrange parameter, γ controls the degree of randomness , and as $\gamma \rightarrow \infty$, the encoding rule approaches the (non-random) nearest-neighbor rule and the entropy goes to zero. Furthermore, this Lagrangian framework is extended to re-formulate the entire MP-GVQ problem as a minimization of the expected distortion, D subject to an entropy constraint. The corresponding Lagrange multiplier, β is inversely related to the temperature in the physical analogy, as explained below.

The method consists of starting with a highly random encoder (large value of the entropy constraint) and gradually reducing the entropy while solving the optimization at each level. At the limit of zero entropy, we obtain a deterministic solution *satisfying the structural constraint* and *minimizing the output distortion.*

This is an annealing process corresponding to the physical analogy where a system whose energy is the output distortion and whose temperature is inversely related to the Lagrange multiplier , β, is gradually cooled down to zero temperature. This analogy also explains the ability of the method to avoid many local minima that riddle the distortion surface. The physical analogy is taken a step further by observing that the system undergoes phase transitions in the sequence of solutions obtained for decreasing values of entropy. These transitions correspond to an increase in the effective size of the model (the number of distinct codevectors found in the solution for each entropy value). We provide a result yielding the critical temperature (at which a set of codevectors "split" into a larger set) as a function of the covariances and cross-covariances of $\mathbf{X}$ and $\mathbf{Y}$ in the respective clusters. The result extends the original results for phase transitions of deterministic annealing process previously studied for conventional vector quantizer design.

We demonstrate the usefulness of our MP-GVQ design procedure for a variety of examples from the source coding literature.

[1] This work was supported in part by the National Science Foundation under grant no. NCR-9314335, the University of California MICRO program, Rockwell International Corporation, Hughes Aircraft Company, Echo Speech Corporation, Signal Technology Inc., Lockheed Missile and Space Company and Qualcomm, Inc.

Universal Trellis Coded Quantization

James H. Kasner and Michael W. Marcellin[1]
Department of Electrical and Computer Engineering
The University of Arizona

Abstract — **A new form of trellis coded quantization is presented based on uniform quantization thresholds and "on-the-fly" codeword training. The universal trellis coded quantization (UTCQ) technique requires no stored codebooks. UTCQ performance is comparable with fully optimized ECTCQ for most rates. Performance for the memoryless Gaussian source is presented.**

TCQ has been shown to be an effective quantizer for memoryless sources with low to moderate complexity [1]. ECTCQ was developed in [2, 3] and achieves MSE performance near (within about 0.5 dB) the rate-distortion bound of the memoryless Gaussian source, at all non-negative encoding rates. In [4], the TCQ subset labelling of Figure 1 was introduced. This index shift makes the quantizer symmetric with respect to codebook supersets ($S_0 = D_0 \cup D_2$ & $S_1 = D_1 \cup D_3$). With the modified labelling, both supersets have access to a zero codeword.

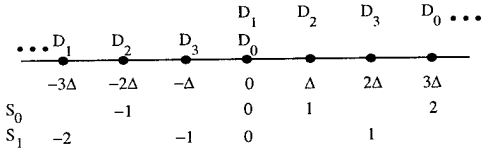


Fig. 1: Modified Subset Labels

The following relationships are evident (assuming a symmetric pdf),

$$CW_{i \in S_0} = -CW_{-i \in S_1} \quad (1)$$

$$p_{S_0}[CW_i] = p_{S_1}[CW_{-i}]. \quad (2)$$

These relationships allow the use of a single variable-rate code for *both* supersets [5]. The encoder returns the S_0 indices and the negative of the S_1 indices. The decoder may uniquely recover the index stream by tracking the trellis state.

UTCQ uses uniform thresholds and codewords for quantization. The encoder is completely characterized by Δ (see Figure 1). For CW_i, ($|i| \geq 2$), the decoder uses uniform codewords. The remaining codewords are trained on the actual sequence being encoded,except $CW_0 \equiv 0$[2]. The trained codewords, are determined by taking the mean of all samples mapping to $i \in S_0$ and the negative of those mapping to $-i \in S_1$. These codewords are quantized within their cells using 256 levels and passed to the decoder. This quantization requires a four byte overhead and guarantees that the quantized codewords are within 0.4% of their trained values.

[1]This work was supported in part by SAIC and by the National Science Foundation under Grant No. 9258374.

[2]Although suboptimum in an MSE sense, we have found this sometimes results in perceptual improvements when used in image coding applications.

In [4] a system similar to UTCQ was presented. There *all* codewords were trained using a training sequence and the codebooks stored. Figure (2) gives the relative distortion between UTCQ when training four codewords versus training all of the codewords. By simply training four codewords, UTCQ achieves virtually the same performance and stores no codebooks. Futhermore, by training on the sequence data itself, UTCQ may perform better when the source statistics are not precisely known.

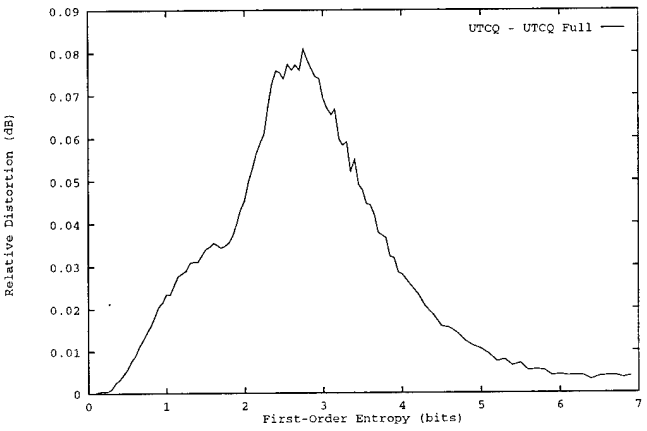


Fig. 2: UTCQ Memoryless Gaussian Performance

References

[1] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Comm.*, vol. COM-38, pp.82-93, Jan., 1990.

[2] T. R. Fischer and M. Wang, "Entropy-constrained trellis-coded quantization," *IEEE Trans. Info. Th.*, vol. IT-38, pp.415-426, Mar., 1992.

[3] M. W. Marcellin, "On entropy-constrained trellis coded quantization," *IEEE Trans. Comm.*, pp.14-16, Jan., 1994.

[4] R. L. Joshi, T. R. Fischer, J. H. Kasner, and M. W. Marcellin "Arithmetic and trellis coded quantization," *Proc. IEEE Int. Symp. on Info. Th.*, Trondheim, Norway, Jun., 1994.

[5] R. Laroia and N. Farvardin "Trellis-based scalar-vector quantizer for memoryless sources," *IEEE Trans. Info. Th.*, vol. IT-40, pp. 860-870, May, 1994.

Maximum Mutual Information Vector Quantization

Lynn D. Wilcox and Les T. Niles
Xerox PARC, Palo Alto, CA 94304 USA

Abstract — **A method is proposed for designing a maximum mutual information (MMI) vector quantizer, for applications in which quantization is used to extract a set of discrete features for use in classification.**

I. Introduction

Vector quantization is commonly used as a feature extraction technique for classification. Typically, the vector quantizer for feature extraction is designed identically to a vector quantizer for coding, that is, to achieve a minimum distortion representation of the original data [3]. While this type of quantizer has proven successful as a feature extraction technique for recognition systems, it seems reasonable to question whether such minimum distortion quantizers are actually optimal for feature extraction.

II. MMI Vector Quantization

We propose a technique for designing a maximum mutual information (MMI) quantizer which maximizes the mutual information $I((\mathbf{X}, C); Q)$ between data $\mathbf{X}$ labeled with class C, and the quantization rule Q. We consider the case when the quantization rule $Q(\mathbf{X}, C) \in \{1, ..., K\}$ is a function of the data and class label, as well as the case when $Q(\mathbf{X})$ is a function of only the data. The quantization rule $Q(\mathbf{X}, C)$ or $Q(\mathbf{X})$ is based on centroids $\mathbf{Y}_1, ..., \mathbf{Y}_K$ associated with each quantization index. The mutual information $I((\mathbf{X}, C); Q)$ between the data and the quantizer is given by [1]

$$I((\mathbf{X}, C); Q) = H(\mathbf{X}, C) - H(\mathbf{X}, C|Q) \tag{1}$$

where $H(\mathbf{X}, C|Q)$ is the conditional entropy of $\mathbf{X}$ and C given the quantization Q. Since $H(\mathbf{X}, C)$ does not depend on the quantization, finding the quantizer to maximize the mutual information between $(\mathbf{X}, C)$ and Q is equivalent to finding the quantizer to minimize $H(\mathbf{X}, C|Q)$.

Now $P(\mathbf{X}, C|Q) = P(\mathbf{X}|Q)P(C|\mathbf{X}, Q)$. We make the simplifying assumption that $P(C|\mathbf{X}, Q) = P(C|Q)$. Thus

$$H(\mathbf{X}, C|Q) = H(\mathbf{X}|Q) + H(C|Q). \tag{2}$$

Let $P(\mathbf{X}|Q)$ be Gaussian, with mean $\mathbf{Y}_Q$ and identity covariance. Then the quantizer which maximizes the mutual information between $(\mathbf{X}, C)$ and Q can be found by minimizing

$$H(\mathbf{X}, C|Q) = \frac{1}{2}E\{(\mathbf{X} - \mathbf{Y}_Q)^2\} - E\{\log(P(C|Q))\}. \tag{3}$$

A class-dependent quantization rule $Q(\mathbf{X}, C)$ can be designed with an MMI criterion by using the standard k-means algorithm [2] to find the centroids $\mathbf{Y}_1, ..., \mathbf{Y}_K$ that minimize the MMI distortion

$$d_{\mathrm{MMI}}(\mathbf{X}, C; Q) = \frac{1}{2}(\mathbf{X} - \mathbf{Y}_Q)^2 - \log(P(C|Q)), \tag{4}$$

averaged over the labeled training data $(\mathbf{X}_1, C_1), ..., (\mathbf{X}_N, C_N)$. The second term in $d_{\mathrm{MMI}}(\mathbf{X}, C; Q)$ requires an estimate of $P(C|Q)$, obtained empirically based on the class labels of the training data.

In practice, the class labels of the data are unknown before quantization. Thus the quantization rule $Q(\mathbf{X})$ must be a function of only the data $\mathbf{X}$. We assume that the form of the quantization rule for $\mathbf{X}$ is to choose the quantization index of the centroid $\mathbf{Y}_k$ that has minimum Euclidean distance

$$d_{\mathrm{E}}(\mathbf{X}; Q) = \frac{1}{2}(\mathbf{X} - \mathbf{Y}_Q)^2. \tag{5}$$

The quantizer design involves finding the centroids $\{\mathbf{Y}_k\}$ to minimize (3). More precisely, since the expectation is over the empirical distribution observed in some labeled training data $(\mathbf{X}_1, C_1), ..., (\mathbf{X}_N, C_N)$, we in fact seek to minimize

$$J = \sum_{i=1}^{N} \left[\frac{1}{2}(\mathbf{X}_i - \mathbf{Y}_{Q(\mathbf{X}_i)})^2 - \log P(C_i|Q(\mathbf{X}_i))\right]. \tag{6}$$

Since the criterion for estimating the centroids (Eq. 6) is now different from the distortion measure used to assign vectors to centroids (Eq. 5), the simple k-means algorithm can't be used for optimizing the centroids. Instead, we will use a gradient descent procedure. Estimating $P(C|Q)$ using simply a count of the samples with quantization index Q and class C, as in the previous section, yields a function which is piecewise-constant with respect to the centroids $\{\mathbf{Y}_k\}$, and thus is not amenable to gradient descent. Instead, we use the estimate

$$\hat{P}(C = m|Q = k) = \frac{\sum_{i=1}^{N} \delta_{C_i,m} P(Q = k|\mathbf{X}_i)}{\sum_{i=1}^{N} P(Q = k|\mathbf{X}_i)}, \tag{7}$$

where $\delta_{C_i,m}$ is the Kronecker delta: $\delta_{C_i,m} = 1$ if $C_i = m$, 0 if $C_i \neq m$. Now

$$P(Q = k|\mathbf{X}_i) = \frac{P(\mathbf{X}_i|Q = k)P(Q = k)}{P(\mathbf{X}_i)}. \tag{8}$$

As before, $P(\mathbf{X}|Q = k)$ is Gaussian, with mean $\mathbf{Y}_k$ and identity covariance, and $P(\mathbf{X}_i) = \sum_{k=1}^{K} P(\mathbf{X}_i|Q = k)P(Q = k)$. We assume $P(Q = k) = 1/K$.

Using these forms for $P(Q = k|\mathbf{X}_i)$ and $\hat{P}(C = m|Q = k)$ in (6), the gradient of J with respect to the cluster centroids $\{\mathbf{Y}_k\}$ can be computed. A standard gradient descent procedure is then applied to minimize J and hence design the codebook.

References

[1] Cover, T.M. and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[2] Linde, Y., A. Buzo, and R.M. Gray. "An Algorithm for Vector Quantizer Design". IEEE Trans. Comm., Vol. COM-28, pp. 84-95, Jan. 1980.

[3] Makhoul, J., S. Roucos, and H. Gish. "Vector Quatnization and Speech Coding". Proc. IEEE, Vol. 73, No. 11, pp. 1551-1588, Nov. 1985.

Robust Quantization for Transmission Over Noisy Channels

Qing Chen and Thomas R. Fischer[1]

School of Elect. Engin. & Comp. Science, Washington State Univ., Pullman, WA 99164-2752, USA

Abstract — **A robust quantizer is proposed for transmission over a binary symmetric channel (BSC). The quantization scheme combines the Gaussian channel-optimized scalar quantizer (COSQ) with an all-pass filtering before/after quantizing. For a broad class of sources the resulting performance is approximately that of the Gaussian COSQ for the memoryless Gaussian source.**

I. Introduction

There are several approaches to designing scalar quantizers (SQs) and vector quantizers (VQs) for use over a binary symmetric channel [1]-[8]. A comparison of the performance of these methods leads to the following conclusion: For the encoding of memoryless (generalized Gaussian) sources, if the channel bit error rate is significant (larger than about 10^{-3}) very little improvement over channel optimized scalar quantization has been achieved.

Figure 1 compares the performance of COSQ to the distortion-rate function evaluated at the channel capacity (termed the optimal performance theoretically attainable (OPTA)) for Gaussian, Laplacian, and generalized Gaussian (with shape parameter $\nu = 0.5$) sources. Two features are evident. The first is that there is large potential performance gain possible whenever the BSC bit error rate is significant. The second is that the general ordering of the COSQ performance curves for Gaussian, Laplacian, and generalized Gaussian ($\nu = 0.5$) sources is exactly opposite to the ordering of their respective optimal performance theoretical attainable.

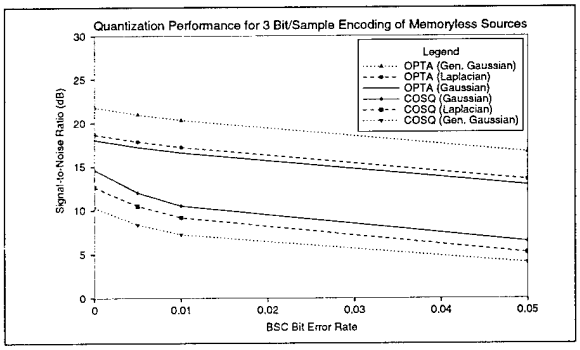


Figure 1: OPTA and COSQ performance.

II. Robust Quantization

All-pass filtering can be used to change the marginal distribution of a source sequence into one that is approximately Gaussian [9][10]. Since the transformation is unitary, the proper concatenation of all-pass filtering, quantization using the Gaussian COSQ, and inverse filtering (at the receiver) provides consistent quantization performance, at the level of the Gaussian COSQ, for a wide variety of source distributions.

Table I compares the robust quantization performance to the COSQ performance [1] for several sources. The all-pass filtering was done using the binary phase scrambling method [10].

	$\epsilon = 0.001$	$\epsilon = 0.010$	$\epsilon = 0.100$
GG	9.18	7.23	2.32
Lap	12.09	9.18	3.79
Gaus	13.99	10.57	4.69
P-GG	13.96	10.57	4.68
P-Lap	13.98	10.56	4.67

Table 1: SNR (in dB) for memoryless Gaussian, Laplacian, and generalized Gaussian (with shape parameter $\nu = 0.5$) sources for COSQ and the robust quantization method (P-GG, P-Lap).

References

[1] N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design for Noisy Channels: An Approach to Combined Source-Channel Coding," *IEEE Trans. on Information Theory*, vol. IT-33, no. 6, pp. 827-838, Nov. 1987.

[2] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. on Information Theory*, vol. IT-37, no. 1, pp. 155-160, Jan. 1991.

[3] E. Ayanoğlu and R. M. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. on Information Theory*, vol. IT-33, pp. 855-865, Nov. 1987.

[4] M. Wang and T. R. Fischer, "Trellis coded quantization designed for noisy channels," *IEEE Trans. on Information Theory*, to appear.

[5] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. on Information Theory*, vol. IT-36, no. 4, pp. 799-809, July 1990.

[6] K. Zeger and A. Gersho, "Pseudo-Gray Coding," *IEEE Trans. on Communications*, vol. COM-38, no. 12, pp. 2147-2158, Dec. 1990.

[7] R. Hagen and P. Hedelin, "Robust vector quantization by a linear mapping of a block code," submitted to *IEEE Trans. on Information Theory*.

[8] P. Knagenhjelm, "Competitive learning in robust communication," Ph.D dissertation, Chalmers University (Sweden), 1993

[9] A. C. Popat and K. Zeger, "Robust quantization of memoryless sources using dispersive FIR filters," *IEEE Trans. on Communications*, vol. 40, pp. 1670-1674, Nov. 1992.

[10] C. J. Kuo and C. S. Huang, "Robust coding technique-transform encryption coding for noisy communications," *Optical Engineering*, Vol. 32, No. 1, pp. 150-156, Jan. 1993.

[1]This work was supported by NSF Grant NCR-9303868.

SOFT DECODING FOR VECTOR QUANTIZATION IN COMBINATION WITH BLOCK CHANNEL CODING

Mikael Skoglund and Per Hedelin
Department of Information Theory, Chalmers University of Technology, S - 412 96 Göteborg, Sweden

I. INTRODUCTION

According to the two-step source/channel coding procedure introduced by Shannon, the source and the channel codes are designed and used separately. Recent research has striven to find efficient combined approaches for source/channel coding. Much of this research has considered vector quantization (VQ) for noisy channels. In this paper we present a method for *joint decoding* of the combination of a vector quantizer and a channel code. Our decoder is *soft* in the sense that no decisions are involved in the decoding, and the unquantized channel outputs are utilized (c.f. [1, 2] and [3]). We depart from the traditional way of decoding, in that we make the decoding into a one-step procedure, without any intermediate channel decoding. A similar approach for scalar quantization and a discrete channel was presented in [4]. We will also demonstrate that estimates of the transmitted binary data can be efficiently obtained in our framework.

II. BLOCK SOURCE AND CHANNEL ENCODING

Consider a VQ encoder in tandem with a block channel encoder. Assume that the *VQ encoder*, α, maps $\mathbf{R}^d$ onto $\Im_N = \{0,1,...,N-1\}$, where $N = 2^k$, and that the binary representation, $\mathbf{b}(i) \in \{\pm 1\}^k$, of the chosen index, $i = \alpha(\mathbf{x})$, for the source vector , $\mathbf{x} \in \mathbf{R}^d$, is encoded into a channel codeword. Let $P_i = \Pr(\alpha(\mathbf{X}) = i)$. The *channel encoder* is described by the mapping $\beta : \Im_N \to \Im_M$, where $\Im_M = \{0,1,...,M-1\}$, $M = 2^n$ and $n \geq k$, such that $i' = \beta(i)$. We take the channel codeword, $\mathbf{c}(i') \in \{\pm 1\}^n$, to be the binary representation of the index i'. The two mappings of the VQ encoder and the channel encoder can be joined into one mapping, $\varepsilon : \mathbf{R}^d \to \Im_M$, where $\varepsilon = \beta \circ \alpha$. With this mapping we associate the probabilities $P'_{i'} = \Pr(I' = i')$, such that $P'_{i'} = P_{\beta^{-1}(i')}$, if $i' \in \beta(\Im_N)$, and $P'_{i'} = 0$ if $i' \in \Im_M \setminus \beta(\Im_N)$. Consequently, the tandem of the original VQ encoder and the channel encoder is equivalent to *a new VQ encoder*, having members of a subset of the index probabilities equal to zero.

III. OPTIMAL SOFT DECODING

Assume that the channel corrupts the transmitted codeword with AWGN. The received vector, $\mathbf{R} = (R_1, R_2, ..., R_n)^T$, can then be expressed as $\mathbf{R} = a \cdot \mathbf{c}(I') + \mathbf{W}$, where a is a known amplitude and $\mathbf{W}$ is Gaussian with covariance matrix $\sigma^2\mathbf{I}$. This model is valid for binary modulation in AWGN, then $\mathbf{R}$ corresponds to samples of the matched filter output at the receiver. The main result of this paper is a Hadamard-based expression for the MMSE soft decoder decoding the source/channel encoder. We use the word *soft* to emphasize that the decoder utilizes the unquantized channel output, the vector $\mathbf{R}$.

The decoder function, $\hat{\mathbf{X}}$, that minimizes $E\|\mathbf{X} - \hat{\mathbf{X}}\|^2$, can easily be shown to be $\hat{\mathbf{X}}(\mathbf{r}) = E[\mathbf{y}_{I'} | \mathbf{R} = \mathbf{r}]$, where $\mathbf{y}_i = E[\mathbf{X} | I' = i]$. This expression can be rewritten using a Hadamard-transform approach. For this purpose we express the vector $\mathbf{y}_i$ as $\mathbf{y}_i = \mathbf{T} \cdot \mathbf{h}_i$, where $\mathbf{h}_i$ is the ith column of an M by M Hadamard matrix $\mathbf{H}$. The matrix $\mathbf{T}$ is fully specified by the vectors $\mathbf{y}_i$ (c.f. [3]). Thus the MMSE-decoder can be written $\hat{\mathbf{X}}(\mathbf{r}) = \mathbf{T} \cdot \hat{\mathbf{h}}(\mathbf{r})$ where $\hat{\mathbf{h}}(\mathbf{r}) = E[\mathbf{h}_{I'} | \mathbf{R} = \mathbf{r}]$. Using this expression it can be shown that optimal soft decoding can be based on estimates, $\hat{b}(r_m) = \tanh(ar_m / \sigma^2)$, of the individual bits of the codeword $\mathbf{c}(I')$. The bit-estimates are used to build a vector $\hat{\mathbf{p}}(\mathbf{r})$, according to $\hat{\mathbf{p}}(\mathbf{r}) = (1, \hat{b}(r_n))^T \otimes \cdots \otimes (1, \hat{b}(r_1))^T$, where $\otimes$ denotes the Kronecker matrix product. It can then be shown that the expression for the vector $\hat{\mathbf{h}}(\mathbf{r})$ becomes $\hat{\mathbf{h}}(\mathbf{r}) = f(\mathbf{r}) \cdot \mathbf{R}_{\mathbf{hh}} \cdot \hat{\mathbf{p}}(\mathbf{r})$, where $\mathbf{R}_{\mathbf{hh}} = E[\mathbf{h}_{I'} \mathbf{h}_{I'}^T]$, and the scalar function $f(\mathbf{r})$ is defined as $f(\mathbf{r}) = \{\mathbf{m}_{\mathbf{h}}^T \cdot \hat{\mathbf{p}}(\mathbf{r})\}^{-1}$, where $\mathbf{m}_{\mathbf{h}} = E[\mathbf{h}_{I'}]$. By modifying an algorithm given in [5] to the framework of the present study, the calculation of $\hat{\mathbf{h}}(\mathbf{r})$, based on the received vector $\mathbf{r}$, can be carried out using an order of $n \cdot 2^n$ operations.

Traditionally, decoding is based on hard bit-estimates that are calculated from the received signal. Our approach performs decoding in a single-step procedure, with no hard decisions involved. However, for applications where hard bit-values are desired, symbol-by-symbol MAP-estimates of the bits can easily be obtained from $\hat{\mathbf{h}}(\mathbf{r})$. Since the vector $\hat{\mathbf{h}}(\mathbf{r})$ will have MMSE-estimates of the information bits in positions 2^m, $m = 0,...,k-1$ (assuming that the channel code is given in systematic form), we obtain the hard bit-estimates as $\hat{b}_{\mathrm{MAP}}(m) = \mathrm{sign}(\hat{h}_{2^m})$, where $\hat{h}_n$ denotes the nth component of $\hat{\mathbf{h}}(\mathbf{r})$. In this paper we investigate the VQ performance in terms of SNR, but we emphasize that transmission of binary data is also easily treated in the soft Hadamard-based framework.

IV. RESULTS

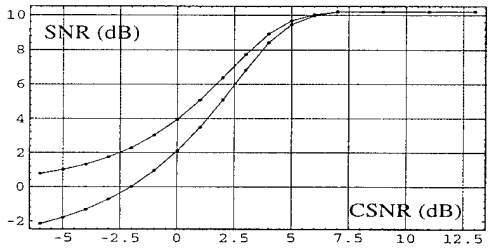


Figure 1. *SNR in dB, as a function of the channel SNR (CSNR).*

A simple example is illustrated in fig. 1. In this example, a 4 bit 4-dimensional VQ trained for a first order Gauss-Markov source having the correlation 0.9 between samples is used in tandem with the Hamming (7,4) block channel code. The simulation shows the soft decoder (upper curve), and decoding based on a two-stage procedure which first decodes the channel code with soft decision ML-decoding, and then uses the decision to perform table-look-up VQ decoding (lower curve). As we can see our soft decoder performs better than the two-stage procedure, and the difference becomes larger for bad channels. This difference is mainly due to the fact that knowledge of the CSNR and the source statistics is utilized in the soft decoder. Furthermore, soft decoding is favorable as a principle since ML detection-based decoding destroys information when taking hard decisions. This information can be utilized by the soft decoder to enhance the performance of the decoding of the VQs.

REFERENCES

[1] V. A. Vaishampayan and N. Farvardin, "Joint Design of Block Source Codes and Modulation Signal Sets," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 1230-1248, July 1992.

[2] F.-H. Liu, P. Ho, and V. Cuperman, "Joint Source and Channel Coding Using a Non-Linear Receiver," in *Proc. ICC '93*, pp. 1502-1507, Geneva, Switzerland, May 1993.

[3] M. Skoglund and P. Hedelin, "A Soft Decoder Vector Quantizer for a Noisy Channel," in *Proc. ISIT '94*, p. 401, Trondheim, Norway, June 1994.

[4] G. A. Wolf and R. Redinbo, "The Optimum Mean-Square Estimate for Decoding Binary Block Codes," *IEEE Transactions on Information Theory*, vol. IT-20, no. May, pp. 344-351, 1974.

[5] M. Skoglund, "A Soft Decoder Vector Quantizer for a Rayleigh Fading Channel - Application to Image Transmission," in *Proc. ICASSP 95*, pp. 2507-2510, Detroit, May 1995.

A Fixed-Rate Trellis Source Code for Memoryless Sources

Liuyang Yang and Thomas R. Fischer[1]
School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA

Abstract — **The trellis-based scalar-vector quantizer (TB-SVQ) for memoryless sources was introduced by Laroia and Farvardin and outperforms all other reasonable complexity fixed-rate quantizers. Unfortunately, the resulting code is catastrophic – a single bit error within a block can propagate indefinitely into other blocks. This paper presents a new algorithm, termed a fixed-rate trellis source code (FRTSC), that achieves essentially the same, or in some cases better, performance as the TB-SVQ for error-free channels, but limits the propagation of channel errors.**

I. Introduction

Vector Quantizers can achieve various gains over uniform scalar quantizers. These gains are classified into boundary (entropy) gain, granular gain and non-uniform density gain [1],[5]. The scalar-vector quantizer (SVQ) [4], introduced by Laroia and Farvardin, is a structured vector quantizer which can achieve both boundary gain and non-uniform density gain without infinite error propagation due to transmission bit errors. The trellis coded quantizers introduced by Marcellin and Fischer [2] are effective structured multidimensional quantizers that realize a significant portion of the ultimate granular gain. Laroia and Farvardin combined the SVQ with TCQ to realize these three gains. The resulting quantizer is called the trellis- based scalar-vector quantizer (TB-SVQ) [1].

II. The Trellis-Based Scalar-Vector Quantizer (TB-SVQ)

Laroia and Farvardin impose two constraints on the TB-SVQ design so that the TB-SVQ enumeration encoding is state-independent and the SVQ enumeration algorithm can be applied directly in the TB-SVQ. This elegant formulation avoids the difficulty of state-dependent enumeration, but unfortunately yields a catastrophic code.

Lemma 1. Given the same binary SVQ codeword, different initial states at the beginning of a block can cause the TB-SVQ decoder to produce different TB-SVQ code-sequences with different ending states at the end of the block.

Theorem 1. The TB-SVQ is a catastrophic code, whether or not a feedback-free encoder is used.

III. A Fixed-Rate Trellis Source Code

A new algorithm, termed a fixed-rate trellis source code (FRTSC), follows the basic idea of the TB-SVQ for combining the SVQ with TCQ, but differs in at least two ways. The first is that no constraints are imposed on the SVQ alphabet as in TB-SVQ. This more general setting allows a zero level to be included easily as a quantization level, and potentially provides performance improvement over the TB-SVQ. The second difference is that a state-dependent enumeration algorithm is used, which is a generalization of the enumeration developed for pyramid trellis codes in [3]. This new enumeration explicitly specifies the ending state for each block.

Let m be the dimension per block, r the bit rate per dimension, and μ the constraint length of the convolutional encoder. Following the notation in [1], let $L(s,t)$ denote the length threshold for the m-vectors with initial state s and final state t. $L(s,t)$ are selected so that no more than rm bits are used to encode each m-vector. For each block, given initial state s, out of the rm available bits, μ bits are used to specify the ending state, and the remaining $rm-\mu$ bits are used to code the trellis sequences with initial state s and final state t.

Infinite error propagation due to channel transmission errors is avoided in the FRTSC because the ending state is explicitly coded and transmitted. Simulation shows that the FRTSC achieves similar performance as TB-SVQ for Gaussian and Laplacian sources at encoding rates of $r = 1, 2, 3$ bits per sample. For sharp-peaked, broad-tailed sources, like the generalized Gaussian with shape parameter $\alpha = 0.5$, some performance improvement is achieved. The improvement is as large as 0.8 dB for a 4-state trellis and an encoding rate of $r = 3$.

References

[1] R. Laroia and N. Farvardin, "Trellis-based scalar-vector quantizer for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 40, pp. 860-870, May 1994.

[2] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, pp. 82-93, Jan. 1990.

[3] T. R. Fischer and J. Pan, "Enumeration encoding and decoding algorithms for pyramid cubic lattice and trellis coded," submitted to *IEEE Trans. Inform. Theory.*

[4] R. Laroia and N. Farvardin, "A structured fixed-rate vector quantizer derived from a variable-length scalar quantizer – Part I: Memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 851-867, May 1993.

[5] M. V. Eyuboglu and G. D. Forney, Jr., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks – High-rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 46-59, Jan. 1993.

[1]This work was supported by NSF Grant NCR-9303868

Why Vector Quantizers Outperform Scalar Quantizers on Stationary Memoryless Sources

David L. Neuhoff

Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109

Abstract

Why do vector quantizers outperform scalar quantizers when the source is stationary and memoryless? This question is frequently asked by newcomers to VQ, who recognize that, in this case, its ability to exploit correlation is of no use. An interesting approach (c.f. [1]) is to a compare k-dimensional VQ with rate R to the k-dimensional product quantizer (PQ) induced by applying a scalar quantizer (SQ) with rate R to k successive source samples. It is then evident that one advantage of VQ is that its cells are more spherical than those of the PQ, which are rectangular. Another is that the points of the VQ are better distributed. Indeed, it is often thought that the PQ distributes points in a "cubic" fashion, whereas the VQ matches its point distribution to the source; e.g. spherical for a Gaussian density.

Using asymptotic quantization theory, we show that aside from the rectangularity of the induced PQ's cells, the shortcoming of SQ's is not that they are incapable of inducing a PQ with an optimal point density. Rather, the structure of the PQ links the point density and cell shapes in a way that causes the best SQ to be a compromise between that which induces the best point density and that which induces the best cell shapes. Consequently, the optimum SQ suffers a point density loss and a cell shape loss. For large rates, we find formulas for these and evaluate them in the Gaussian and Laplacian cases. For example, in the Gaussian case, relative to high-dimensional VQ, an SQ has a 1.88 dB "point density" loss, a 1.53 dB "cubic" loss and a .94 dB "oblongitis" loss.

Summary of Results

Applying an SQ with N_1 points and point density $\lambda_1(x_1)$ to k successive source samples induces a k-dimensional PQ with N_1^k points. Its point density can be shown to be

$$\lambda^{pr}(x) = \lambda_1(x_1)\ldots\lambda_1(x_k), \quad \text{where } x = (x_1,\ldots,x_k).$$

Its cells are rectangular (cubic on the diagonals), and the effect of their shapes on the mean squared error (MSE) is contained in the *inertial profile* m^{pr}, which equals the normalized moment of inertia (nmi) of the cells in the vicinity x. It can be shown that

$$m^{pr}(x) = \frac{1}{12}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{1}{\lambda_1(x_i)^2}\right)\left(\prod_{i=1}^{k}\frac{1}{\lambda_1(x_i)^2}\right)^{-1/k}.$$

Notice that λ_1 affects both λ^{pr} and m^{pr}. In comparison, an optimal k-dimensional VQ for a stationary memoryless source with first-order density $p_1(x_1)$ and kth-order density $p(x)$ has point density [2]

$$\lambda_k^*(x) = c\,p(x)^{k/(k+2)} = c\,p_1(x_1)^{k/(k+2)}\ldots p_1(x_k)^{k/(k+2)}$$

where c is a constant. Its inertial profile is $m_k^*(x) = M_k^*$, where M_k^* is the least nmi of k-dimensional polytopes that tesselate.

To quantitively assess the suboptimality of the point density and inertial profile of a PQ, consider the ratio of its MSE, D^{pr}, to the MSE, $D_{k,N}^*$, of an optimal k-dimensional VQ, which we call the *loss* L. Using the vector version of Bennett's integral [2] and assuming N_1 is large, we find

$$L \triangleq \frac{D^{pr}}{D_{k,N}^*} \cong \int \frac{m^{pr}(x)}{\lambda^{pr}(x)^{2/k}}\,p(x)\,dx \Big/ \int \frac{M_k^*}{\lambda_k^*(x)^{2/k}}\,p(x)\,dx.$$

It is useful to factor this loss into three terms

$$L = L_{pt} \times L_{cu} \times L_{ob}.$$

The *point density loss*, L_{pt}, is the ratio of the MSE of a VQ with point density λ^{pr} and a constant (e.g. optimal) inertial profile to that of a VQ with optimum point density and the same inertial profile. The *cubic loss*, L_{cu}, is the ratio of the MSE of a VQ with cubic cells to one whose cells have nmi equal to M_k^* and the same point density. The *oblongitis loss*, L_{ob}, is the ratio of the MSE of the PQ to that of a VQ with the same point density, but cubic cells; i.e. it is due to rectangularity. The product of cubic and oblongitis losses is the *cell shape loss*.

To optimize the PQ, the scalar point density λ_1 must be chosen to minimize $L_{pt}L_{ob}$. On the one hand, choosing λ_1 to be uniform minimizes L_{ob}. On the other hand, choosing $\lambda_1(x_1) = c'\,p_1(x_1)^{k/(k+2)}$ minimizes L_{pt}. In this case, $\lambda^{pr} = \lambda_k^*$, but there is so much "oblongitis" that $L_{ob} = \infty$. The best scalar point density, $\lambda_1^*(x_1) = c\,p_1(x_1)^{1/3}$, is a compromise. It is more uniform than the point density that minimizes L_{pt}, which reduces "oblongitis". The fact that a PQ can have the optimal point density is often overlooked, probably due to the "squarish" arrangement of its points.

For the optimal scalar point density, formulas for the point density and oblongitis losses can be straightforwardly derived. For a Gaussian density these reduce to

$$L_{pt} = 3\left(\frac{3k}{3k-2}\right)^{k/2}\left(\frac{k}{k+2}\right)^{(k+2)/2} \to 3e^{-2/3} \text{ as } k\to\infty$$

$$L_{ob} = \sqrt{3}\left(\frac{3k-2}{3k}\right)^{k/2} \to \sqrt{3}\,e^{-1/3} \text{ as } k\to\infty$$

which are listed in Table 1, along with L_{cu}, for various k. For a Laplacian density, the point density and cubic losses are the squares of those for the Gaussian density. They are larger (by a factor of 2 in dB) because the sharper peak and heavier tails cause an optimal SQ to be more nonuniform.

A related analysis shows that for a Gaussian source with memory, an optimal transform VQ suffers precisely the same losses as in Table 1.

References

[1] T. Lookabough and R.M. Gray, "High resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inform. Thy.*, IT-35, pp. 1020-1033, Sept. 1989.

[2] S. Na and D.L. Neuhoff, "Bennett's Integral for Vector Quantizers," to appear in *IEEE Trans. Inform. Thy.*

[3] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Thy.*, IT-25, pp. 373-380, July 1979.

[4] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, New York: Springer-Verlag, 1988.

k	cubic L_{cu}	oblong's L_{ob}	pt dens. L_{pt}	shape $L_{ob}L_{pt}$	total $L_{cu}L_{ob}L_{pt}$
2	0.1671	0.6247	0.5115	1.1362	1.3033
4	0.3949'	0.8020	1.0721	1.8741	2.2690'
8	0.6572'	0.8741	1.4373	2.3113	2.9686'
12	0.8084'	0.8962	1.5744	2.4705	3.2789'
24	1.0385'	0.9175	1.7203	2.6377	3.6762'
∞	1.5329	0.9380	1.8759	2.8139	4.3468

Table 1: Losses (in dB) for optimal PQ's for a stationary, memoryless Gaussian source. The "primed" losses are based on a conjectured lower bound to M_k^* [4, pp. 61,62].

Vector Quantization of Spherically Invariant Random Processes

Frank Müller[1] and Franz Geischläger

Institut für Elektrische Nachrichtentechnik, Aachen University of Technology (RWTH), 52056 Aachen, Germany

Abstract — **Vector quantization of spherically invariant random processes (SIRP) is considered. Especially, trellis coded quantization (TCQ) and lattice vector quantization (LVQ) are investigated. For performance evaluations a random number generator has been developed producing sequences which can be regarded as SIRP realizations. It turns out that in most cases the TCQ outperforms all other investigated quantization methods, even those LVQ schemes which are matched to the properties of SIRP sources. Comparisons with bounds from rate distortion theory are given as well.**

I. Introduction

Vector quantization (VQ) plays a key role in lossy data compression. The rate distortion bounds of any source can be reached in principle by VQ when the vector dimension tends to infinity. Unfortunately, with increasing dimension storage and computational complexity tend to infinity as well. To cope with this problem, it makes sense to consider methods which reduce complexity by employment of strongly *structured* codebooks as there are lattice vector quantization (LVQ) and trellis coded quantization (TCQ).

II. SIRP Model Source

A SIRP (in the strict sense) is a random process defined by the property that every n-variate pdf of random variables taken from the process can be written as $f(\mathbf{x}) = \pi^{-n/2} g_n(\mathbf{x}^T\mathbf{x})$. The pdf is constant on hyper-spheres centered around the origin.

A representation theorem due to Yao [1] states that every SIRP can be regarded as a variance mixture of Gaussian processes. For the density function of a SIRP then holds

$$f^{\mathrm{SIRP}}(\mathbf{x}) = \int_0^\infty f^{\mathrm{Gauss}}(\mathbf{x}, r) f_\sigma(r) dr. \qquad (1)$$

Here $f^{\mathrm{Gauss}}(\mathbf{x}, r)$ denotes the multivariate density function of a Gaussian process with standard deviation r. $f_\sigma(r)$ is an univariate density function called sigma density which controls the distribution of the variance. The resulting source itself is non-ergodic, as most natural processes (e.g. image and speech processes) are.

In this contribution the sigma density is modeled in discrete fashion. Particularly, f_σ is modeled with two Dirac impulses with a weight of 0.5 at the locations σ_1 and σ_2. We constructed a random generator, where the sigma density was controlled by a finite state machine with two states. The state transition probability had been fixed to a value of 0.2. The overall variance of the model source was normalized to one which is equal to the condition $\sigma_2 = \sqrt{2 - \sigma_1^2}$.

The univariate pdf of this particular SIRP is then given by

$$f(x) = \frac{0.5}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} + \frac{0.5}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{x^2}{2\sigma_2^2}}. \qquad (2)$$

Note that in the special case $\sigma_1 = \sigma_2 = 1$ the SIRP is Gaussian and only in this case the samples of the process are independent.

III. Simulation Results

The parameter σ_1 of the SIRP random generator has been varied in the range from 0.3 to 1.0. All data samples obtained in this way were encoded by TCQ at a rate R of 1, 2 and 3 bit/sample using a codebook with 2^{R+1} codewords. All sequences were also encoded with a Lloyd-Max scalar quantizer and with lattice vector quantizers.

The SNR for a SIRP source with $\sigma_1 = 0.3$ has been plotted in figure 1 for different quantization methods. For comparison, the Shannon lower bound for SIRP sources "SLB" (according to [2]) and the first order rate-distortion function "RDF" are plotted as well. "D24" denotes direct quantization using the 24 dimensional D-lattice, and "D24Tr" a LVQ scheme due to Herbert [3] which is matched to SIRP sources employing a companding approach. Lastly, "Lloyd" denotes scalar quantization using a Lloyd-Max quantizer.

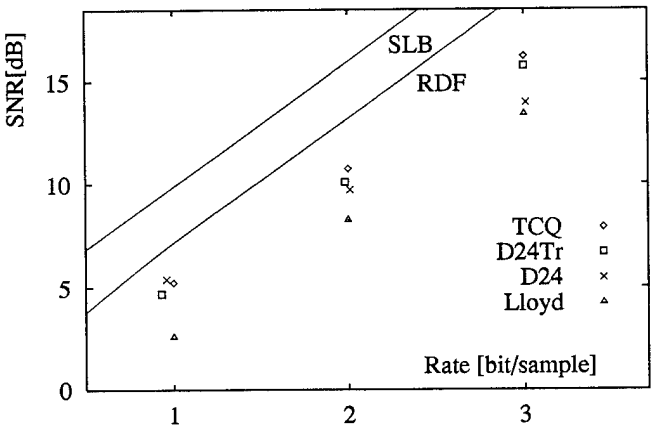


Fig. 1: Comparison of VQ algorithms for SIRP sources with $\sigma_1 = 0.3$

At rates of 2 and 3 bit/sample the TCQ with optimized codebooks outperforms all other investigated quantizers. Only at 1 bit/sample the direct Lattice quantization yields a slight improvement. A comparison with the Shannon lower bounds again demonstrated the good performance of TCQ in the SIRP case.

References

[1] K. Yao, A Representation Theorem and Its Applications to Spherically-Invariant Random Processes *IEEE Trans. Inform. Theory*, IT-19, pp. 600-608, 1973

[2] H. M. Leung und S. Cambanis, "On the rate distortion functions of spherically invariant vectors and sequences", *IEEE Trans. Inform. Theory*, IT-24, S. 367–373 (1978)

[3] M. Herbert, "Lattice quantization of spherically invariant speech-model signals", *Archiv für Elektronik und Übertragungstechnik*, AEÜ–45, S. 235–244 (1991)

[1] E-mail: mueller@ient.rwth-aachen.de

Optimal Quantization for Distributed Estimation via a Multiple Access Channel[1]

Tolga M. Duman Masoud Salehi
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115

Abstract — **Quantizer design algorithms for decentralized estimation are presented. Scalar quantizer design for the problem of multiple descriptions over a multiple-access channel is also studied. The importance of the initial index assignment is explained and an algorithm to choose a good initial index assignment is derived.**

I. Summary

In a typical decentralized detection and estimation system, the objective is to estimate a certain random variable at a fusion center by using the observations of a set of sensors. In general, the observations have very large entropy rates, and therefore, information reduction at the sensors before transmission is necessary. We assume that this information reduction is accomplished by scalar quantization and the quantized values are transmitted to the fusion center via a multiple-access channel (MAC). We derive quantizer design algorithms to minimize the mean squared error (MSE) for both noiseless and noisy observations. We will also present a set of numerical results.

In the rest of this summary, we study the related problem of multiple descriptions over a MAC.

Consider the two channel diversity system where the objective is to transmit a certain source output to a receiver. Assume that one of the links may break down during the transmission. The problem is to send descriptions of the source output over both links in such a way that the overall distortion is minimized when both links are available and at the same time a minimum fidelity is guaranteed when one of the links is broken. This setting is called the multiple descriptions problem.

In particular, when the distortion measure is the mean squared error, the problem is to minimize

$$D_{12} = E[(X - \hat{X})^2 | \text{both links available}]$$

subject to the constraints,

$$D_l = E[(X - \hat{X})^2 | \text{only } l^{th} \text{ link available}] \leq D_{l,max}$$

where $l = 1, 2$. We assume that the transmitter does not know whether there is a broken link or not, on the other hand, the receiver does.

El Gamal and Cover [2] studied this problem from an information theoretical point of view, and derived an achievable rate-distortion region for a memoryless source, independent channels and a single letter fidelity criterion. In [3], Ozarow proved that the region found in [2] is actually the rate distortion region for a Gaussian source with mean squared distortion measure. Vaishampayan has considered the multiple description scalar quantizer design problem for independent noiseless channels [4]. Our work here is the generalization of the work in [4] for the case of a noisy and possibly dependent transmission medium (i.e. a multiple-access channel).

We assume that the source statistics and the channel characteristics are known, and derive a quantizer design algorithm to minimize the Lagrangian by employing some type of joint source and channel coding.

It turns out that, even for the case of independent noiseless links the initial index assignment is very important. In [4], some good index assignment strategies for independent noiseless links are presented. In our setting, the index assignment problem is twofold — one due to the noisy (asymmetric) nature of the links and the other due to the multiple descriptions. Since the transmission medium is not fixed, it is not plausible to obtain a fixed index assignment strategy. We present an algorithm based on simulated annealing to choose a good initial index assignment.

In order to complete the solution of the problem, one has to vary the Lagrange multipliers, apply the design algorithm, and consider time-sharing of the resulting strategies. Therefore, the computational requirements of the solution of the problem is high, on the other hand the computations can be made off-line, and no on-line computation is necessary. It can also be shown that time-sharing of three strategies is always sufficient to obtain the optimal performance. A set of numerical examples that illustrates the use of the algorithm and the general performance improvement by a good initial index assignment will also be presented.

[1] This work was partially supported by the NSF Grant NCR-9101560

References

[1] T. M. Duman, "Multiterminal quantization for distributed detection and estimation," M.S. Thesis, Northeastern University, Boston, MA, May 1995.

[2] A. A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. on Inform. Theory.*, vol. IT-28, pp. 851-857, Nov. 1982.

[3] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909-1921, Dec. 1980.

[4] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Inform. Theory.*, vol. IT-39, pp. 821-834, May 1993.

Vector Quantization of Images with the ECMPNN Algorithm

Dorin Comaniciu[1]

Dept. Electronics & Telecommunications
Polytechnic University of Bucharest, Bucharest, Romania

Abstract - **This paper presents an entropy-constrained version of the Modified Pairwise Nearest Neighbor (MPNN) algorithm [1] for the design of efficient vector quantizers. We called this new algorithm *Entropy-Constrained Modified Pairwise Nearest Neighbor* (ECMPNN).**

I. Introduction

The proposed ECMPNN technique follows the idea of the MPNN algorithm to design an entropy-constrained codebook. MPNN is derived from the well known PNN algorithm [2]. The MPNN clustering starts with an initial codebook of the desired size, containing vectors from the training set (TS). Each vector in the initial codebook is considered a separate cluster. At each step, one new cluster is formed by taking a new TS vector. The number of clusters is maintained to the desired size by merging the two closest clusters. MPNN maintains a superior quality of the generated codebooks, and it requires as many multiplications as the LBG algorithm [3] needs for two iterations [1].

The problem of entropy-constrained vector quantization is to choose the clustering and the codebook in such a way as to minimize the overall distortion subject to an entropy constraint. The solution proposed by the CLG-ECVQ [4] technique uses the Lagrangian formulation. The algorithm minimizes the functional $J=D+\lambda R$, where the parameter λ is the slope of the distortion-rate curve. By varying λ, all the distortion/rate pairs on the convex hull of the operational distortion-rate curve can be found.

II. ECMPNN Algorithm

The ECMPNN clustering begins with an initial codebook C_0 of size N, filled-up with N randomly chosen TS vectors $\{\mathbf{X}_0, \mathbf{X}_1,..., \mathbf{X}_{N-1}\}$. Each vector in the initial codebook corresponds to one cluster. The initial codebook can be written as

$$C_0 = \{\mathbf{Y}_0, \mathbf{Y}_1,..., \mathbf{Y}_{N-1}\} \equiv \{\mathbf{X}_0, \mathbf{X}_1,..., \mathbf{X}_{N-1}\} \quad (1)$$

At each i-th step, one new cluster is formed by a new vector from the TS. The N+1 clusters are converted into N clusters by merging two clusters. The merge is chosen so that it is optimum in the distortion-rate sense. The strategy of finding the best merge is described as follows. Let us denote by (D_{i-1}, R_{i-1}) the distortion/rate pair corresponding to the (i-1)th step of ECMPNN. At the i-th step, we can consider each possible merge of two clusters and compute the slope to any other distortion/rate pair. To find the best merge, it is sufficient to find the merge which yields the smallest magnitude slope. If (D_i, R_i) is the distortion/rate pair that results from the best merge, then, any other merge which yields a slope of larger magnitude will necessarily lie above the line connecting (D_{i-1}, R_{i-1}) and (D_i, R_i). Thus, the merge of two clusters must be taken so that the ratio of distortion increment to entropy decrement induced by the merge to be minimum, that is

$$\lambda_i = \Delta D_i / \Delta R_i = \min \quad (2)$$

We are now in a position to describe the proposed algorithm. At the first step the (N+1) clusters are given by

$$\{\{\mathbf{X}_0\}, \{\mathbf{X}_1\},..., \{\mathbf{X}_{N-1}\}, \{\mathbf{X}_N\}\} \quad (3)$$

where $\mathbf{X}_N$ is the new vector from the TS. Let us suppose that the vectors $\mathbf{X}_0$ and $\mathbf{X}_1$ give the minimum ratio (2). Then, the algorithm classifies together these two vectors in the same cluster, and the resulted N clusters are described by

$$\{\{\mathbf{X}_0, \mathbf{X}_1\}, \{\mathbf{X}_N\}, \{\mathbf{X}_2\},..., \{\mathbf{X}_{N-1}\}\} \quad (4)$$

The codebook is then modified by replacing the first codeword in the codebook with the centroid of vectors $\mathbf{X}_0$ and $\mathbf{X}_1$, and the second codeword with the vector $\mathbf{X}_N$. Note that the remaining codewords are unchanged. The resulted codebook is

$$C_1 = \{\mathbf{X}_{01}, \mathbf{X}_N, \mathbf{X}_2,..., \mathbf{X}_{N-1}\} \quad (5)$$

where $\mathbf{X}_{01}$ signifies the mean of vectors $\mathbf{X}_0$ and $\mathbf{X}_1$. At the second step, by taking a new vector $\mathbf{X}_{N+1}$ from the TS, the (N+1) clusters are given by

$$\{\{\mathbf{X}_0, \mathbf{X}_1\}, \{\mathbf{X}_N\}, \{\mathbf{X}_2\},..., \{\mathbf{X}_{N-1}\}, \{\mathbf{X}_{N+1}\}\} \quad (6)$$

If we further suppose that $\mathbf{X}_2$ and $\mathbf{X}_{N+1}$ are the closest two vectors in the set (6), then, the second step gives N clusters

$$\{\{\mathbf{X}_0, \mathbf{X}_1\}, \{\mathbf{X}_N\}, \{\mathbf{X}_2, \mathbf{X}_{N+1}\},..., \{\mathbf{X}_{N-1}\}\} \quad (7)$$

and the resulted codebook is

$$C_2 = \{\mathbf{X}_{01}, \mathbf{X}_N, \mathbf{X}_{2,N+1},..., \mathbf{X}_{N-1}\} \quad (8)$$

where $\mathbf{X}_{2,N+1}$ is the mean of vectors $\mathbf{X}_2$ and $\mathbf{X}_{N+1}$. This process is continued until all the training vectors have been considered. Since the entropy decrement in (2) depends only of the number of vectors in the two considered clusters, its values can be computed and stored off-line. Therefore, ECMPNN requires only one additional division per step compared to MPNN (see [1] for a discussion on MPNN's computational complexity).

III. Computer Simulation Results

Simulations on a variety of test images showed that the ECMPNN algorithm runs significantly faster than CLG-ECVQ, without sacrificing performance. A block size of 4 x 4 pixels (or vector size 16) was used, and the mean value of each training vector was removed before the codebook design. With the codebook so obtained, each test image was entropy coded. In particular, image Lenna (from outside the TS) of 512 x 512 pixels, 256 gray levels, was coded at a bit rate of 0.359 bpp with a PSNR of 30.46 dB.

Acknowledgments

The author is grateful to Professor Victor Neagoe of Polytechnic University of Bucharest for originally proposing this direction of research.

References

[1] D.Comaniciu, "An Efficient Clustering Algorithm for Vector Quantization", Proc. 9th Scandinavian Conf. on Image Analysis, Uppsala, Sweden, June 1995

[2] W.H.Equitz, "A New Vector Quantization Clustering Algorithm", IEEE Trans. on ASSP, Vol. 37, No. 10, pp. 1568-1575, Oct. 1989

[3] Y.Linde, A.Buzo, and R.M.Gray, "An Algorithm for Vector Quantizer Design", IEEE Trans. on Commun., vol. COM-28, pp. 84- 95, Jan. 1980

[4] P.A.Chou, T.Lookabaugh, and R.M.Gray, "Entropy-Constrained Vector Quantization", IEEE Trans. on ASSP, Vol. 37, pp.31-42, Jan.1989

[1] Correspondence to:
Dorin Comaniciu, C.P. 16-105, Bucharest 16, Romania 77500

A New Universal Random Coding Bound for the Multiple-Access Channel

Yu-Sun Liu and Brian Hughes[1]
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, Maryland 21218, USA

Abstract — **The minimum average error probability achievable by block codes on the two-user multiple-access channel is investigated. A new exponential upper bound is found which can be achieved universally for all discrete memoryless multiple-access channels with given input and output alphabets. It is shown that the exponent of this bound is greater than or equal to those of previously known bounds. Moreover, examples are given where the new exponent is strictly larger.**

Summary

One of the central problems in multiuser information theory is to determine the minimum average error probability which can be achieved on a two-user discrete memoryless multiple-access channel using a block code with rate pair (R_X, R_Y) and blocklength n. The most fundamental result of this theory is the coding theorem of Ahlswede [1] and Liao [4] which asserts that, for any (R_X, R_Y) in the interior of a certain set $\mathcal{C}$ and all sufficiently large n, there exists a multiuser code with an error probability arbitrarily close to zero. Conversely, for any (R_X, R_Y) outside of $\mathcal{C}$, the error probability is bounded away from zero. The set $\mathcal{C}$, which is called the *capacity region*, is the convex closure of the set of rate pairs (R_X, R_Y) satisfying

$$\begin{aligned} 0 &\le R_X \le I(X \wedge Z|Y), \\ 0 &\le R_Y \le I(Y \wedge Z|X), \\ R_X + R_Y &\le I(XY \wedge Z) \end{aligned} \tag{1}$$

for some choice of independent input random variables X and Y, where Z is the corresponding channel output.

Over the past twenty years, stronger versions of this coding theorem, which give exponential upper bounds on the error probability, have been derived by Slepian and Wolf [7], Dyachkov [2], Gallager [3], and Pokorny and Wallmeier [6]. Pokorny and Wallmeier's coding theorem is particularly strong because it asserts the existence of *universal* multiuser codes. By this we mean that a fixed choice of codewords and decoding sets achieves the upper bound for all multiple-access channels with given input and output alphabets.

In this work, we derive a new upper bound for the minimum error probability which can be achieved on the multiple-access channel using a block code with rate pair (R_X, R_Y) and blocklength n. Like Pokorny and Wallmeier's result, our bound is universally achievable for all multiple-access channels with given input and output alphabets. The proof involves a new multiuser packing lemma and a new universal decoding rule which seeks to minimize the empirical equivocation of the users' codewords given the channel output. We show that the exponent of this bound is always greater than or equal to those given in [2, 3, 6, 7]. Moreover, we give examples in which the new exponent is strictly larger. Hence, the corresponding bound on the minimum error probability is tighter for large n.

[1]This work was supported by the National Science Foundation under grant NCR-9217457.

References

[1] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory*, Tsahkadsor, Armenian S.S.R., 1971, Hungarian Acad. Sc., pp. 23-52, 1973.

[2] A. G. Dyachkov, "Random constant composition codes for multiple access channels," *Probl. Control and Inform. Theory*, vol. 13, no. 6, pp. 357-369, 1984.

[3] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 124-142, Mar. 1985.

[4] H. Liao, "A coding theorem for multiple access communications," in *Proc. Int. Symp. Information Theory*, Asilomar, CA, 1972; also "Multiple Access Channels," Ph.D. dissertation, Dept. of Elec. Eng., Univ. of Hawaii, 1972.

[5] Y. S. Liu and B. L. Hughes, "A new universal random coding bound for the multiple-access channel," Dept. of Elec. Comp. Eng., The Johns Hopkins University, Tech. Rep. 95-02, Jan. 1995.

[6] J. Pokorny and H. M. Wallmeier, "Random coding bound and codes produced by permutations for the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 741-750, Nov. 1985.

[7] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System Tech. J.*, vol. 52, pp. 1037-1076, Sept. 1973.

On the factor-of-two bound for Gaussian multiple access channels with feedback

Erik Ordentlich

Info. Systems Lab., Durand 141A, Stanford University, Stanford, CA 94305-4055, U.S.A. Email: eor@isl.stanford.edu

Abstract — **Pombra and Cover [1] and Thomas [2] showed that the maximum achievable throughput (sum of rates of all users) of a Gaussian multiple access channel with feedback is at most twice that achievable without feedback. We prove a stronger result which establishes the factor-of-two bound not only for the total throughput but for the entire capacity region as well. Specifically, we show that the capacity region of a Gaussian multiple access channel with feedback is contained within twice the capacity region without feedback.**

I. Introduction

A channel use at time j of a Gaussian multiple access channel (MAC) involves m independent users each transmitting a real number $X_{ij}, i \in \{1, \ldots, m\}$. Thus, X_{ij} denotes the transmission of the i^{th} user at time j. A single receiver observes $Y_j = \sum_{i=1}^{m} X_{ij} + Z_j$ where Z_j is a sample from an arbitrary Gaussian noise process with known n-block covariance $K_Z^{(n)}$. The channel is assumed to operate in one of two modes: with or without feedback. In the no-feedback mode, the users base their transmissions exclusively on the messages they wish to send to the receiver which are assumed random and independent of each other and the noise. With feedback, the users can adapt their transmissions based on previously received symbols (the Y_j's) available to each user over a noiseless and delayless feedback link. In both cases the users' transmissions must satisfy average power constraints, $n^{-1}\sum_{j=1}^{n} X_{ij}^2 \le P_i$. Since the same feedback signal is observed by all users, they can cooperate to some extent and achieve higher reliable transmission rates than in the absence of feedback. Memory in the noise, if it is non-white, can also be exploited for additional gains. Therefore, the capacity region of the Gaussian MAC with feedback strictly includes the capacity region without feedback. The gains with feedback are, however, limited to a factor of two, which we prove in the following theorem.

Theorem 1 (Factor-of-two bound) *If* $(R_1^{FB}, \ldots, R_m^{FB})$ *is an achievable rate vector for a Gaussian MAC with feedback under power constraints* $(P_1, \ldots, P_m)$, *then* $(R_1^{FB}/2, \ldots, R_m^{FB}/2)$ *is an achievable rate vector without feedback.*

Figure 1 illustrates the theorem for the case of two users ($m = 2$). The boundary of the capacity region of a two user Gaussian MAC with feedback (C_{FB}) lies in the shaded region between the boundary of the no-feedback capacity region (C) and twice this boundary ($2C$).

II. Proof outline

The proof of Theorem 1 relies on two theorems. The first, proved by Keilers [3], gives the capacity region of an m user Gaussian MAC without feedback.

Theorem 2 (No-feedback theorem) *Rates* $(R_1, \ldots, R_m)$ *are achievable for expected average power constraints* $(P_1, \ldots, P_m)$ *if and only if for all* $\epsilon > 0$ *and all* n *sufficiently large, there exist* $n \times n$ *covariance matrices* $K_{X_1}^{(n)}, \ldots, K_{X_m}^{(n)}$

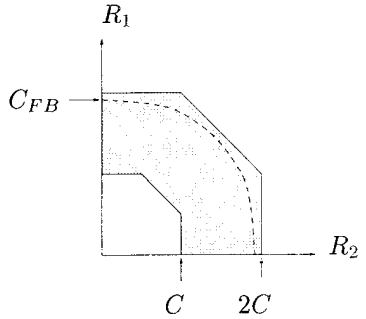


Fig. 1: The factor-of-two bound for two users.

satisfying

$$R(S) \le \frac{1}{2n} \log \frac{\det(K_Z^{(n)} + \sum_{i \in S} K_{X_i}^{(n)})}{\det K_Z^{(n)}} + \epsilon, \qquad (1)$$

for all $S \subseteq \{1, \ldots, m\}$, *with* $\frac{1}{n}\text{trace}(K_{X_i}^{(n)}) \le P_i$, *for all* i.

The second theorem is an extension of a result of Pombra and Cover [1] and Thomas [2] and provides an outer bound on the capacity region of the Gaussian MAC with feedback.

Theorem 3 (Feedback theorem) *If* $(R_1^{FB}, \ldots, R_m^{FB})$ *is an achievable rate vector with feedback for expected average power constraints* $(P_1, \ldots, P_m)$, *then for all* $\epsilon > 0$ *and all* n *sufficiently large, there exists a joint distribution on* $(X_1^n, \ldots, X_m^n, Z^n)$ *with the marginal on* Z^n *equal to the noise distribution, and covariance matrices satisfying*

$$R^{FB}(M) \le \frac{1}{2n} \log \frac{\det K_{[Z+X(M)-X(M^c \cap S)]}^{(n)}}{\det K_Z^{(n)}} + \epsilon, \qquad (2)$$

for all pairs of nested subsets $M \subseteq S \subseteq \{1, \ldots, m\}$, *with* $\frac{1}{n}\text{trace}(K_{X_i}^{(n)}) \le P_i$, *for all* i.

In the above theorems, $R(S) = \sum_{i \in S} R_i$ and $X(S) = \sum_{i \in S} X_i^n$ respectively denote rate sums and transmission sums over subsets of users.

We outline the proof of the factor-of-two bound (Theorem 1) as follows. A vector of rates $(R_1^{FB}, \ldots, R_m^{FB})$ achievable with feedback satisfies the inequalities of the feedback theorem (Theorem 3) for some covariance structure for all n sufficiently large. Using some combinatorial lemmas, we show that this covariance structure also satisfies the inequalities of the no-feedback theorem (Theorem 2) for $(R_1^{FB}/2, \ldots, R_m^{FB}/2)$. This establishes that the rate vector $(R_1^{FB}/2, \ldots, R_m^{FB}/2)$ is achievable without feedback.

References

[1] S. Pombra and T. M. Cover. Non white Gaussian multiple access channels with feedback. *IEEE Trans. Info. Theory*, IT-40:885–892, May 1994.

[2] J. A. Thomas. Feedback can at most double Gaussian multiple access channel capacity. *IEEE Trans. Info. Theory*, IT-33:711–716, Sept. 1987.

[3] C. W. Keilers. *The capacity of the spectral Gaussian multiple access channel.* PhD thesis, Stanford University, Stanford, CA, May 1976.

On the capacity for the T-user M-frequency noiseless multiple access channel without intensity information

A.J. Han Vinck, Jeroen Keuning* and Sang Wu Kim**
Institute for Experimental Mathematics, Ellernstr. 29, 45326 Essen, Germany
* Eindhoven University of Technology, Eindhoven, The Netherlands
** Korea Advanced Institute of Science and Technology, Taejon, Korea

Abstract — **We discuss the achievable ϵ-error throughput for the uncoordinated (asynchronous) T-user M-frequency multiple-access channel without intensity information. The problem is formulated in terms of frequencies, but the results are also applicable to Pulse Position Modulation (PPM) schemes. We show that the achievable sum rate for T users reduces from $(M-1)$ bits per channel use in the fully coordinated multi-access situation to $(M-1)\cdot\ln(2)$ bits per channel use if we assume no coordination between users. In particular, the result shows that for multi-tone M-ary frequency shift keying multiple access in asynchronous operation for instance, multiple user interference reduces the capacity only by a factor $\ln(2)=0.695$ relative to the ideal TDMA system.**

I. Summary

Cohen, Heller and Viterbi [1] presented a new approach to completely asynchronous multiple access digital communications. In asynchronous multiple access one assumes that T individual users can access the system independently of the other users. Each user transmits by means of on-off signaling without regard to, or knowledge of the remaining $T-1$ other users. In a synchronized Time-Division Multiple Access (TDMA) system, each user would be assigned $1/T$ of the available dimensions and with on-off signaling the transmission rate can be 1 bit/dimension, yielding a capacity per user of $1/T$ bits/dimension. In [1] it has been shown that in the asynchronous system, multiple user interference reduces the total capacity for T users only by a factor of $\ln(2)=0.695$ relative to the ideal TDMA system. This efficiency can be achieved by using low-duty-cycle signaling. A practical example of such a signaling is multi-tone (M-tone) frequency shift keying, where a specific user transmits *one* out of M orthogonal frequencies. For PPM, the signaling interval is partitioned into M sub-intervals or time slots. During a signaling interval, only one of the M sub-intervals is used to transmit a pulse.

Chang and Wolf [2] considered the synchronous T-user M-frequency noiseless multiple access channel where only *one* receiver decodes all users simultaneously. For a large number of users, the channel capacity approaches $(M-1)$ bits per signaling interval of M frequencies. The results were arrived at by a computer search.

We derive an achievable rate for the asynchronous T-user M-frequency noiseless multiple access channel and show that the achievable rate reduces from $(M-1)$ bits per channel use in the fully coordinated multi-access situation to $(M-1)\cdot\ln(2)$ bits per channel use if we assume no coordination between users, or one-to-one communication. We start with 2-tone signaling and show the nature of the detection problem. For each individual user, the 2-tone multiple access channel is, from a capacity point of view, equivalent to the binary input-binary output Z-channel. Although the 2-tone multiple access channel has a ternary output, capacity is the same as if we make a hard decision in case of an ambiguous reception of two frequencies. The asymptotic optimizing input distribution is highly asymmetric, indicating that each of the users must transmit a low duty-cycle signal.

We extend the system to $M\geq 3$ frequencies and give an input distribution from which it follows that the channel capacity is upper bounded by $M\cdot\ln(2)$ and lower bounded by $(M-1)\cdot\ln(2)$ bits per frequency interval. Since the channel transition probabilities are functions of the input distribution, we cannot use the Kuhn-Tucker conditions for a candidate (capacity achieving) input distribution as given above. Instead, we prove that the achievable rate $C(T,M)$ for the channel with M frequencies and T users, asymptotically approaches $C(T,M)\rightarrow(M-1)\cdot\ln(2)$, $M\geq 2$ fixed and $T\rightarrow\infty$.

II. Conclusions

We summarize the results as follows:

1. The capacity for the asynchronous T-user M-frequency noiseless multiple access channel approaches $\ln(2)$ bits *per frequency* (dimension);
2. The capacity achieving distribution puts all mass on one frequency and divides the remaining probability mass equally on the remaining $M-1$ frequencies;
3. Instead of using one out of 2^M-1 combinations of frequencies from a given frequency interval with M orthogonal frequencies, capacity can be achieved by using only a *single* frequency from an M-frequency interval, which is the advantage of the M-tone frequency shift keying systems.
4. The cut-off rate R_{comp} approaches $(M-1)\cdot 0.413$ bit per signaling interval for the same type of input probability distribution as given in 2.
5. A practical coding scheme achieving $(1-\epsilon)\cdot\ln(2)$ bits per dimension with vanishing decoding error probability is given. The coding method is equivalent to Frequency Hopping MFSK and extends and modifies the strategy as given in [1].

References

[1] A. Cohen, J. Heller and A. Viterbi, "A New Coding Technique for Asynchronous Multiple Access Communication", *IEEE Trans. Commun.*, vol. COM-19, pp. 849–855, October 1971.

[2] Shin-Chun Chang and J.K. Wolf, "On the T-User M-Frequency Noiseless Multiple-Access Channel with and without Intensity Information", *IEEE Trans. on Information Theory*, vol. IT-27, pp. 41-48, January 1981.

An Extension of the Achievable Rate Region of Schalkwijk's 1983 Coding Strategy for the Binary Multiplying Channel

H.B. Meeuwissen, J.P.M. Schalkwijk, and A.H.A. Bloemen

Eindhoven University of Technology, Dept. of Electrical Engineering, Group on Information and Communication Theory, P.O. Box 513, 5600 MB Eindhoven, the Netherlands.

Abstract — **A new region $\mathcal{R}$ of achievable rate pairs $(R_1, R_2) \in \mathcal{R}$ is established for the binary multiplying channel. The new region $\mathcal{R}$ has an equal rate point of $R_1 = R_2 = 0.63072$ bit per transmission.**

I. Definitions

This paper is concerned with the binary multiplying channel (BMC) [1]. The capacity region of the BMC is bounded by the Shannon inner bound region $\mathcal{G}_i$, and the Shannon outer bound region $\mathcal{G}_o$. These regions are plotted in Fig. 1.

Communication over the BMC by two distant terminals is modeled as follows. A message Θ_t at terminal t, $t = 1, 2$, is encoded into the channel input sequence $\boldsymbol{X}_t = (X_{t,1}, X_{t,2}, ..., X_{t,n})$. The common channel output sequence $\boldsymbol{Y}_1 = \boldsymbol{Y}_2 = \boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$ is formed such that $Y_j = X_{1,j} X_{2,j}$, $X_{t,j} \in \{0, 1\}$, $j = 1, 2, ..., n$. Note that the first channel input $X_{t,1}$ is based on the message Θ_t only, while the k-th channel input $X_{t,k}$, $k = 2, 3, ..., n$ is based on both the local message Θ_t, and the previous channel outputs $(Y_1, Y_2, ..., Y_{k-1})$. The decoder at terminal t estimates the other terminal's message Θ_{3-t} from both the channel output sequence $\boldsymbol{Y}$, and the local message Θ_t.

A coding strategy for the BMC is described as a progressive subdivision of the $[0, 1) \times [0, 1)$ square. Therefore, the probability of each resolution product that occurs in this progressive subdivision of the unit square is equal to its area.

II. Schalkwijk's 1983 Coding Strategy

The 1983 coding strategy is composed of alternating so-called inner and outer bound transmissions. Let $\Pr[i]$ and $\Pr[o]$ denote the average code word length of the inner and outer bound transmissions, respectively. Of course, $\Pr[i] = 1$. Let $I(\Theta_t; \boldsymbol{Y} | \Theta_{3-t}, i)$ and $I(\Theta_t; \boldsymbol{Y} | \Theta_{3-t}, o)$ denote the information rate of an inner and an outer bound transmission from encoder t to decoder $3 - t$, respectively. The achievable rate region of the 1983 coding strategy satisfies $\mathcal{R}' = \{(R_1, R_2) :$

$$0 \leq R_t \leq \frac{\Pr[i]\, I(\Theta_t; \boldsymbol{Y} | \Theta_{3-t}, i) + \Pr[o]\, I(\Theta_t; \boldsymbol{Y} | \Theta_{3-t}, o)}{\Pr[i] + \Pr[o]}\}.$$

The region $\mathcal{R}'$ has an equal rate point of $R_1 = R_2 = 0.63056$ bit per transmission and includes the region $\mathcal{G}_i$. In the unit square, a message pair (Θ_1, Θ_2) is always situated in a subrectangle after an inner bound transmission and a subsequent outer bound transmission. Thus, the inner and outer bound transmissions can be repeated ad infinitum in all these subrectangles.

III. The New Coding Strategy

The new coding strategy consists of a structure of inner bound transmissions of average code word length $3\Pr[i]$, such that (i) an efficient resolution product is generated, and (ii) an unlimited number of repetitions of this resolution product is generated. The subdivision of these efficient resolution products is completed by (i) outer bound transmissions of average code word length $3\Pr[o] - \mathrm{L}[loss]$, and (ii) three new transmissions of average code word length $\mathrm{L}[gain]$. In fact, the new coding strategy, see [3], is a modification of the 1983 coding strategy that results in both a loss and a gain with respect to its original. Let $\mathrm{I}_t[gain]$ denote the average mutual information of the three new transmissions from encoder t to decoder $3 - t$, then the achievable rate region of the new strategy satisfies $\mathcal{R} = \{(R_1, R_2) :$

$$0 \leq R_t \leq \frac{3\Pr[i]\, I(\Theta_t; \boldsymbol{Y} | \Theta_{3-t}, i)}{3\Pr[i] + 3\Pr[o] - \mathrm{L}[loss] + \mathrm{L}[gain]} + \frac{(3\Pr[o] - \mathrm{L}[loss])\, I(\Theta_t; \boldsymbol{Y} | \Theta_{3-t}, o) + \mathrm{I}_t[gain]}{3\Pr[i] + 3\Pr[o] - \mathrm{L}[loss] + \mathrm{L}[gain]}\}.$$

The new region $\mathcal{R}$ has an equal rate point of $R_1 = R_2 = 0.63072$ bit per transmission and includes the region $\mathcal{R}'$. The results of van Overveld [4] prove that all rate pairs $(R_1, R_2) \in \mathcal{R}$ are operationally achievable. The new region $\mathcal{R}$ is also plotted in Fig. 1.

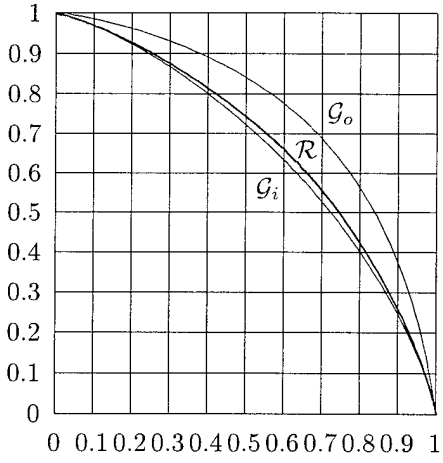


Fig. 1: The new region $\mathcal{R}$ of achievable rate pairs.

References

[1] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. on Math. Statist. and Prob.*, vol. 1, 1961, pp. 611-644.

[2] J. P. M. Schalkwijk, "On an extension of an achievable rate region for the binary multiplying channel," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 445-448, May 1983.

[3] J. P. M. Schalkwijk, H. B. Meeuwissen, and A. H. A. Bloemen, "Coding strategies and a new achievable rate region for the binary multiplying channel," preprint.

[4] W. M. C. J. van Overveld, *On the Capacity Region for Deterministic Two-Way Channels and Write-Unidirectional Memories*. Ph.D. dissertation, Dept. of Electrical Engineering, Eindhoven Univ. of Technol., Eindhoven, The Netherlands, 1991.

Interference Cancellation in Groups

S.V. Hanly[1], P.A. Whiting[2]

1. AT&T Bell Laboratories, 600 Mountain Ave. Murray Hill, NJ 07974, U.S.A.
2. Mobile Communications Research Centre, University of South Australia, 5095 S. Australia

I. Introduction

Consider a multi-user white noise channel of bandwidth W Hz, white noise spectral density $\frac{\eta}{2}$, with M users all received at power P, all requiring the same rate R bits/sec. It is well known that that there is a Shannon capacity for the channel and that it is achievable by FDMA. It is well known that the capacity can also be achieved by an interference cancellation procedure that involves M^2 cancellations, and Rimoldi and Urbanke [3] have recently shown that it is achievable with at most $2M$ cancellation steps. In the present paper we show that we can achieve rates arbitrarily close to capacity with $O(1)$ cancellation steps in the particular case of equal powers and equal rates. The results in the present paper first appeared in Hanly [1].

II. Interference cancellation

It is well known that the equal rate Shannon capacity can be achieved by a combination of time sharing and interference cancellation (Wyner [4]). It can equally well be achieved by a combination of frequency sharing and interference cancellation. In such a scheme there are M subchannels of bandwidth $\frac{W}{M}$ and all users send in all sub-bands. In sub-band 1 we might decode user 1 first, subtract its signal, then decode user 2, and so on. We choose the orderings of the users in the sub-bands in such a way that each user is decoded in the jth position precisely once out of all the sub-bands. This is really exactly the same as time sharing, except that we do not require time synchronization.

Both time-sharing and frequency sharing cancellations require M cancellation steps in each sub-band, for a total of M^2 cancellation steps. In the present paper we consider a scheme with $O(1)$ cancellation steps.

III. Interference cancellation of groups

Let us partition the bandwidth into J subchannels, each of bandwidth $\frac{W}{J}$ and partition the users into J groups. We order the *groups* among the subchannels, just as in the frequency sharing interference cancellation scheme.

Without loss of generality, assume that the groups in subchannel 1 are decoded in the order $\mathcal{G}_J, \mathcal{G}_{J-1}, \ldots, \mathcal{G}_1$. In the first cancellation step we decode all the users in $\mathcal{G}_J$ in parallel. The decoder of a $\mathcal{G}_J$ user treats the interference from all other users in $\mathcal{G}_J$, as well as $\mathcal{G}_{J-1}, \ldots, \mathcal{G}_1$ (ie *all* other users) as random noise. The decoded signals of the $\mathcal{G}_J$ users are passed to an adder, the sum $\mathcal{G}_J$ signal reconstituted, and this is then subtracted from the total received signal. Users in $\mathcal{G}_{J-1}$ are then decoded in parallel. A $\mathcal{G}_{J-1}$ decoder treats the interference from all other $\mathcal{G}_{J-1}$ users, and the users in $\mathcal{G}_1, \ldots, \mathcal{G}_{J-2}$ as random noise. Note that the interference from $\mathcal{G}_J$ users has been subtracted out. This process continues and requires a total of J cancellation steps. Finally, the decoder of a $\mathcal{G}_1$ user only has to contend with interference from other $\mathcal{G}_1$ users.

We are interested in the limiting regime in which M, the number of users, grows large. We scale the bandwidth linearly with M, $W \equiv W_0 M$, but the common received power P is fixed. The number of groups, J, is also fixed, so the number of users in each group is M/J. Let $R_j^{(M)}$ be the bit rate of a $\mathcal{G}_J$ user in subchannel 1.

Result 1 *Let $\alpha \equiv \frac{P}{\eta W_0}$ be fixed. Then*

$$R_j^{(M)} = \frac{W_0}{\ln 2} \frac{\alpha/J}{1+j\alpha/J} + O\left(\frac{1}{M}\right) \quad \textit{bits/sec.}$$

and the common bit rate $R^{(M)}$ is given by

$$R^{(M)} = \frac{W_0}{\ln 2} \sum_{j=1}^{J} \frac{\alpha/J}{1+j\alpha/J} + O(1/M) \quad \textit{bits/sec.}$$

The Shannon capacity of the channel is independent of M and is given by $C = W_0 \log_2\left(1+\frac{P}{\eta W_0}\right)$ bits/sec. Moreover,

Result 2 $C = \frac{W_0}{\ln 2} \sum_{j=1}^{J} \frac{\alpha/J}{1+j\alpha/J} + O(1/J)$ *bits/sec.*

Sketch Proof: We write $C = \frac{W_0}{\ln 2} \int_0^{1+\alpha} \frac{1}{1+u}\, du$ and then take a Riemann sum approximation.

IV. Conclusions

Our interference cancellation scheme involves J^2 cancellation steps. Suppose we wish to achieve a rate $(1-\epsilon)C$ for each user. We can first choose a J sufficiently large so that

$$C - \frac{W_0}{\ln 2} \sum_{j=1}^{J} \frac{\alpha/J}{1+j\alpha/J} < \epsilon$$

This J will then work for *all* sufficiently large M, in the sense that for sufficiently large M, $C - R^{(M)} < \epsilon$. We conclude that to be arbitrarily close to Shannon capacity, we do not need more than $O(1)$ cancellation steps, as the number of users increases.

The results of the present paper are extended to the multi-receiver radio network context, and to the case of multiple power levels, in Hanly [1] and Hanly and Whiting [2]. Rimoldi and Urbanke [3] give a scheme that can actually *achieve* Shannon capacity with at most $2M$ cancellation steps, and this is for any set of received powers, and any point in the feasible rate region. We suggest that the complexity of their scheme, at least for the equal rate and equal powers case, may be further reduced, at a small price in terms of bit rate, by incorporating our group cancellation approach in their procedure.

References

[1] Hanly S.V. (1993) Information capacity of radio networks *Ph. D. Thesis, Cambridge University*, August.

[2] Hanly S.V., Whiting P.A. (1995) Interference cancellation of groups of users *in preparation*

[3] Rimoldi B., Urbanke R. (1994) Onion peeling and the Gaussian multiple access channel *submitted to IEEE Trans. Inform. Theory*

[4] A.D. Wyner (1974) Recent results in the Shannon theory *IEEE Trans. on Information Theory*, Vol. 20, Jan.

Multiterminal Estimation Theory with Binary Symmetric Source

Hidetoshi Shimokawa[1] Shun-ichi Amari[1]

[1] Dept. of Math. Eng. and Info. Phys., Faculty of Eng., University of Tokyo, Bunkyou-ku, Tokyo 113, Japan

Abstract — **The multiterminal estimation theory discuss the maximum Fisher information under the Shannon information restriction. In the single-terminal case, it is trivial problem because the maximum Fisher information can be attained at asymptotically 0-rate. Han and Amari[1] discuss about this problem generally and give the lower bound of the maximum Fisher information under rate restriction. Its approach is based on Slepian-Wolf type rate region. In this paper, we give an example, binary symmetric case, which represents that sufficient statistics can be sent at the rate outside of SW-region using Körner and Morton's method[2], and show that it gives a better bound than the one of Han and Amari[1]. Finally we give the general form of such parametric family of which sufficient statistics can be sent at the rate in KM type region.**

I. Introduction

Let X and Y be discrete i.i.d. source which have a joint probability distribution $P_{XY}(\theta)$, where x^n and y^n are encoded at rate R independently. The encoded messages are denoted by $u_n = f_x^n(x^n)$ and $v_n = f_y^n(y^n)$. The estimator ϕ estimates θ by u_n, v_n. In this paper, we discuss about the minimum rate at which we can estimate θ by u_n, v_n as same estimation error as by x^n, v_n.

An encoder f^n and an estimator $\hat{\theta}_n$ must have following property.

- Rate restriction: $\frac{1}{n}\log\|f^n\| \le R$.
- Asymptotically efficient: $\lim_{n\to\infty} E_\theta[\hat{\theta_n}] = \theta$.

Here, we consider the variance of the estimator V_n, and it's inverse I_n.

$$V_n(\theta; f_x, f_y) = E_\theta[(\hat{\theta}_n - \theta)^2] = \frac{1}{I_n(\theta; f_x, f_y)}.$$

Our aim is to maximize the I_n under rate restriction. Let I^* be defined by

$$I^*(\theta; R) = \lim_{n\to\infty} \frac{1}{n} \max_{f_x^n, f_y^n, \hat{\theta}} I_n(\theta; f_x, f_y).$$

For simpleness, we consider binary symmetric case that P_{XY} is given by the following.

$$P_{XY}(\theta) = \begin{pmatrix} \theta/2 & (1-\theta)/2 \\ (1-\theta)/2 & \theta/2 \end{pmatrix} \qquad (1)$$

Han and Amari[1] gives the lower bound of the Fisher information $I_{Han}(\theta, R)$ at rate R as the following.

$$I_{Han}(\theta; R) = \frac{64a^4}{4 - 16(2\theta-1)a^2(1-2a^2)},$$

where

$$R = 1 - h(\alpha), \qquad \alpha = a + \frac{1}{2}.$$

II. Main Result

Theorem 1

We assume that the parametric family $P_{XY}(\theta)$ is defined in the region $0 < \theta < \theta'$ or $1-\theta' < \theta < 1$, where $0 < \theta' < \frac{1}{2}$. If $R \ge H(\theta')$, θ can be estimated without loss of information, that is, attain same variance as when x^n and y^n can be observed.

This Theorem can be proven by the following technic.

- Minimum entropy decoding for universal coding.
- The method to send binary addition (Körner and Marton[2]).

This Theorem implies that the sufficient statistics of θ can be sent at rate $H(\theta')$. In the other hand, Han and Amari[1] needs rate $(1 + H(\theta'))/2$ to attain same variance as when x^n and y^n can be observed.

Corollary 1

By simple time sharing method, $I^*(\theta; R)$ is bounded as the following.

$$I^*(\theta; R) \ge \begin{cases} \frac{1}{\theta(1-\theta)} \frac{R}{H(\theta')} & R < H(\theta') \\ \frac{1}{\theta(1-\theta)} & \text{otherwise} \end{cases}. \qquad (2)$$

This bound is tighter than I_{Han} especially when θ' is close to 0 or 1, and R is close to $H(\theta')$.

These results is obtained by considering alphabet on GF(2), and we can easily extend these results to GF(p^k), where p is a prime number. For example, on GF(2^2), we consider the following parametric family.

$$\begin{pmatrix} \theta_1/8 & (1-\theta_1)/8 & \theta_2/8 & (1-\theta_2)/8 \\ (1-\theta_1)/8 & \theta_1/8 & (1-\theta_2)/8 & \theta_2/8 \\ \theta_2/8 & (1-\theta_2)/8 & \theta_1/8 & (1-\theta_1)/8 \\ (1-\theta_1)/8 & \theta_1/8 & (1-\theta_2)/8 & \theta_2/8 \end{pmatrix}. \qquad (3)$$

In general, we consider such a parametric family on GF(p^k) that,

- X and Y has p^k alphabets respectively,
- $(p-1)p^{k-1}$ parameters,
- if $x+y = x'+y'$ (on GF(p^k)) then $\Pr\{X = x, Y = y\} = \Pr\{X = x', Y = y'\}$.

Theorem 2

About the parametric families above, It needs less rate to send sufficient statistics of parameters than to send x^n, y^n.

The detail and proof of this theorem will be shown in the full paper.

References

[1] Han, T. S. and Amari, S.: Multiterminal Estimation Theory, METR 93-19, Department of Mathematical Engineering and Information Physics, Faculty of Engineering, The University of Tokyo, 1993.

[2] Körner, J. and Marton, K.: How to Encode the Modulo-Two Sum of Binary Sources, *IEEE Tran. IT*, Vol. 25 (1979), 219–221.

Single User Coding for the Discrete Memoryless Multiple Access Channel

A. Grant[1], B. Rimoldi[2], R. Urbanke[2] and P. Whiting[1]

1. Mobile Communications Research Centre
University of South Australia
The Levels, 5095 Australia

2. Washington University, Dept. of Electrical Engineering
Electronic Systems and Signals Research Laboratory
St. Louis, MO 63130, USA

Abstract — **We show that the coding problem of any m-user asynchronous discrete multiple access channel can be reduced to at most $2m-1$ single user coding problems. This extends previous results for the Gaussian channel.**

I. Introduction

Consider an m-user discrete memoryless channel. This is defined in terms of m finite input alphabets $\mathcal{X}_i, i = 1, \cdots, m$, an output alphabet $\mathcal{Y}$ and a transition probability matrix $p(y|x_1, \cdots, x_m)$. It is well known [1] that the capacity region is the convex hull of the union of rate regions that are achievable for a fixed set of input distributions, $\prod_{i=1}^{i=m} p_i(x_i)$, such that $\sum_{i\in\mathcal{L}} R_i \leq I(X_{i\in\mathcal{L}}; Y|X_{i\in\mathcal{L}^c}), \forall \mathcal{L} \subseteq \{1, \cdots, m\}$. Hui and Humblet have shown [2] that without the convex hull operation, the remaining region describes the rate-tuples which can be achieved without time synchronization. The proposed scheme is for such asynchronous channels.

Although the theoretical limits of discrete memoryless multiple access channels are well understood, there are few examples of multiple access channels for which explicit and efficient codes are known. By contrast, significant progress has been made for single user channels of practical interest, most notably the Gaussian channel at low and high signal-to-noise ratios. It is to be expected that the single user problem will always be better understood and techniques to its solution will be more numerous and efficient than for the multiple access problem. The key contribution of this paper is to translate the problem of finding coding schemes for a given discrete memoryless multiple access channel into the one of finding such schemes for an appropriately defined single user channel.

Vertices of the capacity region can be achieved by successive cancellation [1] of m single user codes. We show that any point in an m-dimensional asynchronous capacity region can be viewed as a vertex in an appropriately defined $(2m-1)$-dimensional asynchronous capacity region. This extends the result in [3] for the Gaussian case.

II. The Result and Proof for Two User Case

Theorem 1 *Any rate tuple in the asynchronous capacity region of a discrete m-user multiple access channel can be achieved by means of single-user decoding of at most $2m-1$ users.*

Proof for $m = 2$. Without loss of generality [4] assume a rate-tuple (R_1, R_2) such that $R_1 + R_2 = I(X_1, X_2; W)$, $R_1 < I(X_1; Y|X_2)$, $R_2 < I(X_2; Y|X_1)$. Assume it is possible to write $X_1 = f(U, V)$ for some function f and random variables U and V which are mutually independent and independent of X_2. Then

$$\begin{aligned} R_1 + R_2 &= I(X_1, X_2; Y) = I(U, V, X_2; Y) \\ &= I(U; Y) + I(X_2; Y|U) + I(V; Y|U, X_2). \quad (1) \end{aligned}$$

If we can choose the distribution on U and V such that $R_2 = I(X_2; Y|U)$, then (1) shows that single user decoding can be employed, decoding first the input corresponding to U then the input corresponding to X_2, and finally the input corresponding to V, i.e., (1) describes a vertex.

Let U and V have the same alphabet as $X_1 \in \{1, \cdots, J\}$ and let $f(u, v) = \max\{u, v\}$. Let the distributions on U, V, and X_1 be p_U, p_V, and p_{X_1}, respectively. Define $p_U(\epsilon) = \epsilon p_{X_1} + (1-\epsilon)e$, $\epsilon \in [0, 1]$, where e is the distribution with all its weight on the first element. It can be verified that for any $\epsilon \in [0, 1]$ a well defined p_V exists such that $X_1 = f(U, V)$. Furthermore, if $\epsilon = 0$ then $I(X_2; Y|U) = I(X_2; Y)$ whereas if $\epsilon = 1$ then $I(X_2; Y|U) = I(X_2; Y|X_1)$. Since $I(X_2; Y) \leq R_2 \leq I(X_2; Y|X_1)$ the claim then follows by continuity. □

Note that $f(u, v) = \max\{u, v\}$ is not the only possible function, but this particular choice leads to a simple proof.

III. An Example

Consider the binary multiplier channel, where the channel inputs X_1 and X_2 as well as the channel output $W = X_1 X_2$ are elements of $\{0, 1\}$. The capacity region of the binary multiplier channel is well known [1, p. 390] and is characterized by the set of rate tuples (R_1, R_2) such that $R_1 + R_2 \leq 1$. To achieve the rate tuple $(R_1, R_2) = (0.5, 0.5)$ we may choose the input distributions to be $p_{X_1} = p_{X_2} = (1 - 1/\sqrt{2}, 1/\sqrt{2})$ and let $f(u, v) = \max\{u, v\}$. The appropriate input distributions for U and V are $p_U = (0.57, 0.43)$ and $p_V = (0.51, 0.49)$. It follows that $(R_V, R_{X_2}, R_U) = (0.41, 0.5, 0.09)$ is a single-user decodable rate triple for the new channel.

This work was supported in part by Telecom Australia under Contract No.7368 and by the Commonwealth of Australia under International S & T Grant No.56 as well as by National Science Foundation Grant NCR-9357689 and NCR-9304763.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[2] J. Y. Hui and P. A. Humblet, "The Capacity Region of the Totally Asynchronous Multiple-Access Channel " *IEEE Trans. Inf. Th. Vol. IT-31 No. 2 March 1986, pp 207 - 216*

[3] B. Rimoldi and R. Urbanke, "On Single-User Decodable Codes for the Gaussian Multiple Access Channel" in *IEEE Int. Symp. Info. Theory* , 1994.

[4] S. Hanly and P. Whiting, "Constraints on capacity in a multi-user channel, " in *IEEE Int. Symp. Info. Theory*, 1994.

Rate-Distortion Theory for a Triangular Communication System

Hirosuke Yamamoto
Dept. of Math. Eng. & Inform. Physics, Univ. of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Abstract — **The admissible rate-distortion region is determined for a triangular communication system shown in Fig. 1.**

I. Summary

Consider a triangular communication system (TCS) shown in Fig. 1. where the source outputs X and Y are i.i.d. but mutually correlated random variables, which take values in finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. The decoder's outputs $\hat{X} \in \hat{\mathcal{X}}$ and $\hat{Y} \in \hat{\mathcal{Y}}$ are allowed distortion, which is measured by distortion measures, $d_X(X,\hat{X}) < \infty$ and $d_Y(Y,\hat{Y}) < \infty$, respectively. We consider block coding. Hence, for $\mathrm{X}^K = (X_1, X_2, \cdots, X_K)$ and $\mathrm{Y}^K = (Y_1, Y_2, \cdots, Y_K)$, the encoder f and the decoders g_X and g_Y are defined as the following mappings.

$$\begin{aligned}(W_X, W_Y) &= f(\mathrm{X}^K, \mathrm{Y}^K),\\ \hat{\mathrm{X}}^K = (\hat{X}_1, \hat{X}_2, \cdots, \hat{X}_K) &= g_X(W_X, V),\\ \hat{\mathrm{Y}}^K = (\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_K) &= g_Y(W_Y, U),\end{aligned}$$

where W_X and W_Y are sent to decoders g_X and g_Y, respectively, and $U = (U_{[1]}, U_{[2]}, \cdots, U_{[L]})$ and $V = (V_{[1]}, V_{[2]}, \cdots, V_{[L]})$ are codewords to communicate between two decoders g_X and g_Y, and they are defined by

$$\begin{aligned}U_{[\ell]} &= g_U^{[\ell]}(W_X, V_{[1]}, V_{[2]}, \cdots, V_{[\ell-1]}), \quad \ell = 1, 2, \cdots, L,\\ V_{[\ell]} &= g_V^{[\ell]}(W_Y, U_{[1]}, U_{[2]}, \cdots, U_{[\ell-1]}), \quad \ell = 1, 2, \cdots, L.\end{aligned}$$

Letting $W_X \in \mathcal{I}(M_X)$, $W_Y \in \mathcal{I}(M_Y)$, $U_{[\ell]} \in \mathcal{I}(M_{U_{[\ell]}})$, and $V_{[\ell]} \in \mathcal{I}(M_{V_{[\ell]}})$, where $\mathcal{I}(M) \triangleq \{0, 1, 2, \cdots, M-1\}$, the rate of each channel is defined as

$$R_X \triangleq \frac{1}{K}\log M_X, \qquad R_Y \triangleq \frac{1}{K}\log M_Y,$$

$$R_U \triangleq \frac{1}{K}\sum_{\ell=1}^{L}\log M_{U_{[\ell]}}, \qquad R_V \triangleq \frac{1}{K}\sum_{\ell=1}^{L}\log M_{V_{[\ell]}}.$$

For $(\mathrm{X}^K, \hat{\mathrm{X}}^K)$ and $(\mathrm{Y}^K, \hat{\mathrm{Y}}^K)$, each distortion is measured by the averaged single letter distortion measure, i.e. $d_X^{(K)}(\mathrm{X}^K, \hat{\mathrm{X}}^K) = \frac{1}{K}\sum_{k=1}^{K} d_X(X_k, \hat{X}_k)$, $d_Y^{(K)}(\mathrm{Y}^K, \hat{\mathrm{Y}}^K) = \frac{1}{K}\sum_{k=1}^{K} d_Y(Y_k, \hat{Y}_k)$.

Rate-distortion tuple $(R_X, R_Y, R_U, R_V, D_X, D_Y)$ is called admissible if for any $\varepsilon > 0$, sufficiently large K, and some finite L, there exists a code $(f, g_X, g_Y, g_U^{[\ell]}, g_V^{[\ell]}; \ell = 1, 2, \cdots, L)$ that satisfies

$$\begin{aligned}E d_X^{(K)}(\mathrm{X}^K, \hat{\mathrm{X}}^K) &\le D_X + \varepsilon,\\ E d_Y^{(K)}(\mathrm{Y}^K, \hat{\mathrm{Y}}^K) &\le D_Y + \varepsilon.\end{aligned}$$

The admissible rate-distortion region $\mathcal{R}$ for the TCS is defined as

$$\begin{aligned}\mathcal{R} \triangleq \{&(R_X, R_Y, R_U, R_V, D_X, D_Y):\\ &(R_X, R_Y, R_U, R_V, D_X, D_Y) \text{ is admissible}\}.\end{aligned}$$

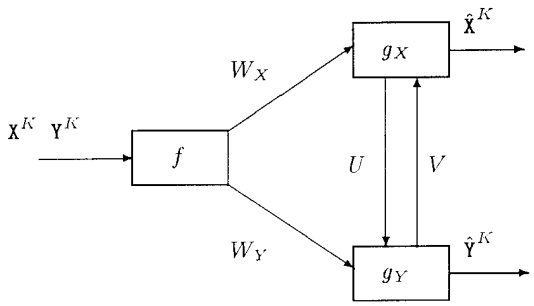


Figure 1: Triangular Communication System

This admissible region $\mathcal{R}$ is determined by the following theorem.

Theorem 1

$$\begin{aligned}\mathcal{R} = \{&(R_X, R_Y, R_U, R_V, D_X, D_Y):\\ &R_X \ge R_{X|S}(D_X) \quad R_Y \ge R_{Y|S}(D_Y)\\ &R_X + R_V \ge R_{X|S}(D_X) + I(XY;S)\\ &R_Y + R_U \ge R_{Y|S}(D_Y) + I(XY;S)\\ &R_U + R_V \ge I(XY;S)\\ &R_X + R_Y \ge R_{X|S}(D_X) + R_{Y|S}(D_Y) + I(XY;S),\\ &\textit{for some auxiliary random variable } S \in \mathcal{S}\\ &\textit{such that } |\mathcal{S}| \le |\mathcal{X}||\mathcal{Y}| + 2\},\end{aligned}$$

where $R_{X|S}(D_X)$ and $R_{Y|S}(D_Y)$ are the conditional rate-distortion functions.

Although the proof of the converse part is complicated, the direct part can easily be proved by using the code of the Gray-Wyner system [1]. From the proof of the direct part, the admissible region can be attained with $L = 1$. Hence, the iterative communication between the decoders is not necessary.

It is also worth noticing that when $R_X + R_Y = H(XY)$ in the distortionless case, the minimum of $R_U + R_V$ is equal to the Wyner's common information [2]. Hence, the Wyner's common information can be explained as the minimum rate necessary to communicate between the decoders in the TCS under the conditions that

- the total rate sent from the encoder to the decoders is minimum,
- X and Y must be reproduced with arbitrarily small error probability.

References

[1] R.M.Gray and A.D.Wyner, "Source coding for a simple network ", *Bell Sys. Tech. J.*, vol.58, pp.1681-1721, Nov. 1974

[2] A.D.Wyner, "The common information of two dependent random variables", *IEEE Tans. on Inform. Theory*, vol.IT-21, no.2, pp.163-179, March 1975

Extended Shannon's Inequality and the Asymptotic Capacity of T-User Binary Adder Channel

Wai Ho Mow[1]

Dept. of Elect. & Comp. Eng., Univ. of Waterloo, Waterloo, Ontario, CANADA N2L 3G1

Abstract — **Extension of Shannon's inequality for discrete probability distribution with an infinite number of elements is considered. As an application, the asymptotic capacity of T-user binary adder channel is exactly determined. Previously known *asymptotically good* (but not very) T-user code of Chang and Weldon is shown to be far from *asymptotically very good*. It is thus concluded that the achievability problem of the asymptotic capacity remains open.**

I. Extended Shannon's Inequality

Denote the set of N-D positive real-valued vectors by $\mathbf{R}_+^N$. Let $\mathcal{X}(N) = \{(x_1, x_2, \cdots, x_N) \in \mathbf{R}_+^N : \sum_{k=1}^N x_k = 1\}$.

Theorem 1 *(I) For any integer $N \in [2, \infty]$, and for all $p \in \mathcal{X}(N)$ and $q \in \mathbf{R}_+^N$, define the function $f_N(p,q) = \sum_{k=1}^N p_k \log(p_k/q_k)$, then*

$$f_N(p,q) \geq 0 \qquad \text{if } \sum_{k=1}^N q_k \leq 1, \tag{1}$$

$$f_N(p,q) \leq 0 \qquad \text{if } \sum_{k=1}^N p_k^2/q_k \leq 1. \tag{2}$$

(II) For $N \in [2, \infty)$, if $\sum_{k=1}^N q_k \leq 1$, the necessary and sufficient condition for $f_N(p,q) = 0$ is

$$\sum_{k=1}^N p_k^2/q_k = 1, \tag{3}$$

or equivalently,

$$p_k = q_k \text{ for } k = 1, 2, \cdots, N. \tag{4}$$

Remarks: (i) The base of the logarithmic function in the definition of f_N is arbitrary provided it is greater than 1. (ii) The condition (4) in part II is in fact a classical form of Shannon's inequality [1],[4]. The extension actually refers to part I. (iii) In part II, if the hypothesis is replaced by $\sum_{k=1}^N p_k^2/q_k \leq 1$ and (3) by $\sum_{k=1}^N q_k = 1$, the result is still valid. (iv) (3) is a sufficient condition for infinite N, and we conjecture that it is also a necessary condition, as it is for every finite N.

II. Asymptotic Capacity of T-User Binary Adder Channel

As an application of Part I of Theorem 1, we now consider the asymptotic capacity of T-user binary adder channel. We refer the reader to [2],[3] for the background of this topic. The (sum) capacity is defined by $C_{\text{sum}}(T) = -\sum_{i=0}^T c_i \log_2 c_i$, where $c_i = \binom{T}{i}/2^T$. Wolf [5] observed that the maximal achievable rate sum is about $\frac{1}{2}\log_2(\pi eT/2)$ for such channel. Chang and Weldon [3] proved that

$$\frac{1}{2}\log_2\frac{\pi T}{2} \leq C_{\text{sum}}(T) \leq \begin{cases} \frac{1}{2}\log_2(\pi eT/2) & \text{even } T, \\ \frac{1}{2}\log_2(\pi e(T+1)/2) & \text{odd } T. \end{cases} \tag{5}$$

They also conjectured that $\frac{1}{2}\log_2(\pi eT/2)$ is an upper bound for odd T. Recently, Blake [2] observed that $1 + \frac{1}{2}\log_2(T)$ is a much tighter lower bound, and that $\frac{1}{2}\log_2(\pi eT/2)$, as an upper bound, is very tight (c.f. [2, Table 1]). These observations motivated our work.

In [3], a T-user code is said to be *asymptotically good* if its rate sum satisfies

$$\lim_{T\to\infty} \frac{R_{\text{sum}}(T)}{\frac{1}{2}\log_2 T} = \lim_{T\to\infty} \frac{C_{\text{sum}}(T)}{\frac{1}{2}\log_2 T} = 1, \tag{6}$$

where the last equality follows from (5). Constructions of asymptotically good T-user codes were also given in [3].

Our next theorem, whose proof invokes Part I of Theorem 1, determines the exact asymptotic value of $C_{\text{sum}}(T)$.

Theorem 2

$$\lim_{T\to\infty} \left(C_{\text{sum}}(T) - \frac{1}{2}\log_2\frac{\pi eT}{2} \right) = 0. \tag{7}$$

In view of Theorem 2, a T-user code is said to be *asymptotically very good* if its rate sum satisfies

$$\lim_{T\to\infty} \left(R_{\text{sum}}(T) - \frac{1}{2}\log_2\frac{\pi eT}{2} \right) = 0. \tag{8}$$

The asymptotically good T-user code of Chang and Weldon [3] is not asymptotically very good. (In fact, the r.h.s. of (8) is ∞ instead of 0!) Hence, the achievability problem of the asymptotic capacity of T-user binary adder channel, or equivalently, the existence problem of the asymptotically very good T-user code remains open.

Acknowledgements

This author thanks Professor Ian F. Blake for bringing his attention to the coding problem for binary adder channels.

References

[1] J. Aczél, "On Shannon's inequality, optimal coding, and characterizations of Shannon's and Rényi's entropies," Research Report CS-73-05, Dept. of Applied Analysis and Computer Science, Univ. of Waterloo, Jan. 1973.

[2] I. F. Blake, "Coding for Adder Channels," in *Communications and Cryptography* (Eds. R. E. Blahut, D. J. Costello, U. Maurer and T. Mittelholzer), Kluwer Academic Publishers, pp. 179–185, 1994.

[3] S.-C. Chang and E. J. Weldon, Jr., "Coding for T-User Multiple-Access Channels," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 6, pp. 684–691, Nov. 1979.

[4] A. Feinstein, *Foundations of Information Theory*, McGraw-Hill, New York – Toronto – London, 1958, chapter 2.

[5] J. K. Wolf, "Multi-user communication networks," in *Communication Systems and Random Process Theory*, (NATO Advanced Study Institute Series), J. K. Skwirzynski, Ed. Leyden, The Netherlands: Noordfhoff International, 1978, pp. 37–53.

[1]This work was supported by the Croucher Foundation Fellowship 1994/95.

Coding for Computing

Alon Orlitsky[1] and James R. Roche[2]

[1] AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974
[2] Center for Communications Research, Thanet Road, Princeton, NJ 08540

Abstract

A sender communicates with a receiver who wishes to reliably evaluate a function of their combined data. We show that if only the sender can transmit, the number of bits required is a conditional entropy of a naturally defined graph. We also determine the number of bits needed when the communicators exchange two messages.

I Introduction

f is a function of two random variables X and Y. A sender $P_{\mathcal{X}}$ knows X, a receiver $P_{\mathcal{Y}}$ knows Y, and both want $P_{\mathcal{Y}}$ to reliably determine $f(X,Y)$. How many bits must $P_{\mathcal{X}}$ transmit?

Embedding this communication-complexity scenario (Yao [6]) in the standard information-theoretic setting (Shannon [4]), we assume that (1) $f(X,Y)$ must be determined for a *block* of many independent (X,Y)-instances, (2) $P_{\mathcal{X}}$ transmits after observing the whole block of X-instances, (3) a vanishing block error probability is allowed, and (4) the problem's *rate* $L_f(X|Y)$ is the number of bits transmitted for the block, normalized by the number of instances.

Two naive bounds are easily established. $L_f(X|Y) \geq H(f(X,Y)|Y)$, the number of bits required when $P_{\mathcal{X}}$ knows Y in advance, and by a simple application of the Slepian-Wolf Theorem, $L_f(X|Y) \leq \min\{H(g(X)|Y) : g(X) \text{ and } Y \text{ determine } f(X,Y)\}$. Both bounds are tight in special cases, but not in general.

Drawing on rate-distortion results, we show that for every X, Y, and f,

$$L_f(X|Y) = H_G(X|Y). \tag{1}$$

G is a simply-defined *characteristic graph* of X, Y, and f. $H_G(X|Y)$ is the *conditional G-entropy* of X given Y. It extends $H_G(X)$, the G-entropy of X, defined by Körner [3], also called the *graph entropy* of G and X. Graph entropy has recently been used to derive an alternative characterization of perfect graphs, lower bounds on perfect hashing, lower bounds for Boolean formula size, and algorithms for sorting.

The lower bound ($\geq$) in (1) is proven via an analogy between $H_G(X|Y)$ and rate-distortion results of Wyner and Ziv [5] and their extension in Csiszár and Körner [1]. The upper bound ($\leq$) strengthens these rate-distortion results, showing that in certain application the same rate suffices to achieve small block- and not just bit-error probability. The proof uses *robust* typicality, a more restrictive form of the asymptotic equi-partition property.

We also consider the more general scenario in which the communicators can exchange two messages. $P_{\mathcal{Y}}$ sends a message based on the block of Y instances, and $P_{\mathcal{X}}$ responds with a message based on $P_{\mathcal{Y}}$'s message and the block of X instances. Again, $P_{\mathcal{Y}}$ must accurately evaluate all $f(X,Y)$'s. $P_{\mathcal{X}}$'s transmission rate r_x, and $P_{\mathcal{Y}}$'s transmission rate r_y, are the number of bits they transmit, normalized by the block length. We determine the region $R_f(X|Y)$ of possible rate pairs for all X, Y, and f.

Two random variables U and V are *admissible* if (1) $U-Y-X$, (2) $V-UX-Y$, and (2) U, V and Y determine $f(X,Y)$. We show that for every (X,Y) and f,

$$R_f(X|Y) = \Big\{(r_x, r_y) \;:\; r_x \geq I(V;X|UY) \text{ and } r_y \geq I(U;Y|X) \text{ for some admissible } U \text{ and } V\Big\}.$$

The *inner* bound is derived by generalizing the one-way achievability results. To prove the (matching) *outer* bound, we extend results of Kaspi and Berger [2] to a larger class of distortion measures.

References

[1] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, 1978.

[2] A. H. Kaspi and T. Berger. Rate-distortion for correlated sources. *IEEE Transactions on Information Theory*, 28(6):828–840, 1982.

[3] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. *Proceedings of the 6th Prague Conference on Information Theory*, pages 411–425, 1973.

[4] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[5] A.D. Wyner and J. Ziv. The rate distortion function for source coding with side information at the receiver. *IEEE Transactions on Information Theory*, 22(1):1–11, January 1976.

[6] A.C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

Coding for Interactive Communication

Leonard J. Schulman[1]
Dept. of Applied Mathematics, Weizmann Inst. Science
College of Computing, Georgia Inst. Technology

Abstract — **Let the input to a computation problem be split between two processors connected by a communication link; and let an interactive protocol π be known, by which on any input, the processors can solve the problem using no more than T transmissions of bits between them, provided the channel is noiseless. We study the following question: If in fact there is some noise on the channel, what is the effect upon the number of transmissions needed in order to solve the communication problem reliably?**

I. Introduction

Shannon, in his seminal study of communication [3], studied the effect of noise upon "one-way" communication problems, i.e. data transmission. His fundamental observation was that coding schemes which did not treat each bit separately, but jointly encoded large blocks of data into long codewords, could achieve very small error probability (exponentially small in T), while slowing down by only a constant factor relative to the T transmissions required by the noiseless-channel protocol (which can simply send the bits one by one). The constant (ratio of noiseless to noisy communication time) is a property of the channel, known as its Shannon capacity.

The improvement in communication rate provided by Shannon's insight is dramatic: if the channel is memoryless, the naive protocol which repeates each bit several times can only achieve the same error probability by repeating each bit a number of times proportional to the length of the entire original protocol. (For a total of T^2 communications.) Moreover in order to achieve any communication on "adversarial" or "worst-case" channels in which any set of a given number of transmissions may be garbled, such error-correcting codes are necessary. A precise statement of Shannon's coding theorem (for the special case of binary symmetric channels, BSCs) follows. With some loss in the capacity, a similar statement can be made for "adversarial" channels.

Theorem 1 (Shannon) *Let a BSC of capacity C be given. For every T and every $\gamma > 0$ there exists a code $\chi : \{0,1\}^T \to \{0,1\}^{T\frac{1}{C}(1+\gamma)}$ and a decoding map $\chi' : \{0,1\}^{T\frac{1}{C}(1+\gamma)} \to \{0,1\}^T$ such that every codeword transmitted across the channel is decoded correctly with probability $1 - e^{-\Omega(T)}$.*

Recently, in computer science, communication has come to be critical to distributed computing, parallel computing, and the performance of VLSI chips. In these contexts interaction is an essential part of the communication process, and its role has been extensively studied through the "communication complexity" model initiated by of A. C. Yao [4] (see [1] for a survey). Noise afflicts interactive communications just as it does the one-way communications considered by Shannon, and for much the same reasons: physical devices are by nature noisy, and there is often a significant cost associated with making them so reliable that the noise can be ignored. (By providing very strong transmitters, cooled circuits, etc.) To mitigate such costs we can design our systems to operate reliably even in the presence of some noise. The ability to transmit data in the presence of noise, the subject of Shannon's and subsequent work, is a necessary but far from sufficient condition for sustained interaction and computation.

Observe that in the case of an interactive protocol, the processors generally do not know what they want to transmit more than one bit ahead, and therefore cannot use a block code as in the one-way case. Another difficulty that arises in our situation but not for data transmission, is that once an error has occurred, subsequent exchanges on the channel are affected. Such exchanges cannot be counted on to be of any use either to the simulation of the original protocol, or to the detection of the error condition. Yet the processors must be able to recover, and resume synchronized execution of the intended protocol, following any sequence of errors, although these may cause them to have very different records of the history of their interaction. In spite of these new difficulties we have:

Theorem 2 *In each direction between a pair of processors let a BSC of capacity C be given. There is a deterministic communication protocol which, given any noiseless channel protocol π of length (duration) T, simulates π on the noisy channel in time $\theta(T/C)$ and with error probability $e^{-\Omega(T)}$.*

In all but a constant factor in the rate, this is an exact analog, for the general case of interactive communication problems, of the Shannon coding theorem. A similar statement can be shown also for the case of "adversarial" channels.

As part of our work we introduce and show the existence of a new class of codes, "explicit" tree codes. (These are different from, though in part inspired by, the random tree codes of the sequential decoding literature.) Computationally effective (e.g. polynomial-time) construction of these codes is an open problem. We show that if these codes can be implemented with polynomial-time computation, then so can the encoding and decoding procedures of the protocol. To be precise: Given an oracle for a tree code, the expected computation time of each of the processors implementing our protocol, when the communication channels are BSCs, is polynomial in T.

Our results are described more fully in reference [2].

References

[1] L. Lovász. Communication complexity: A survey. In Korde et al, editor, *Algorithms and Combinatorics*. Springer-Verlag, 1990.

[2] L. J. Schulman. Coding for Interactive Communication. Manuscript.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, pp. 379–423, 1948; pt. II, pp. 623–656, 1948.

[4] A. C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Annual Symposium on Theory of Computing*, pages 209–213, 1979.

[1]Supported in part by an NSF Postdoctoral Fellowship.
schulman@cc.gatech.edu

Coding for Distributed Computation

Sridhar Rajagopalan[1]
Princeton University

Leonard J. Schulman[2]
Dept. of Applied Mathematics, Weizmann Inst. Science
College of Computing, Georgia Inst. Technology

Abstract — **We show that any distributed protocol which runs on a noiseless network in time T, can be simulated on an identical noisy network with a slowdown factor proportional to $\log(d+1)$, where d is the maximum degree in the network, and with exponentially small probability of error.**

I. Introduction: A Coding Theorem

Shannon's coding theorem [2] can be stated as follows: The number of transmissions sufficient to send a T bit message over a noisy channel with reliability $1 - e^{-\Omega(T)}$ is asymptotic to $\frac{1}{C}T$ where $0 < C < 1$ is the "channel capacity", a function only of the noise characteristics of the channel. In addition, Shannon proves the converse – that this many transmissions are required.

Can we extend the theory to networks, with a number of links available for simultaneous use? In his work, Schulman[1] shows an analog of the Shannon coding theorem for a pair of processors running an interactive protocol in a model introduced by Yao[3] in his work in communication complexity.

The main theorem in our work (stated most simply in the case of a noisy network in which each connection is made via a binary symmetric channel of capacity at least $C > 0$) is the following:

Theorem I.1 *Any protocol Π which runs in time T on a noiseless N-processor network of maximum degree d can be simulated on that network if it is noisy, in time $O(T\frac{\log(d+1)}{C})$. The probability that the simulation fails is at most $Ne^{-\Omega(T)}$.*

The simulation is said to fail if any processor terminates in a state other than that which it would have arrived at in the absence of noise.

Model: Consider a network $\mathcal{N}$ with maximum degree d. We make the following assumptions. First, all noise in our system occurs only in the communication links between processors. Second, we look only at the number of communication bits and ignore the computational cost of the protocol. Third, we require that our protocol be event driven and therefore implementable in the asynchronous setting but we analyze its correctness and efficiency in the synchronous setting. We do this so that the notion of the trajectory of the system and thus the notion of simulation is well defined.

II. Method

On every channel of our network we will implement communications using tree codes, introduced by Schulman in [1]. The noiseless protocol will be embedded within a simulation that uses locally initiated (hence asynchronous) "backups", followed by renewed transmissions, in response to perceived errors in the simulation.

[1]Received support from NSF PYI Award CCR 88-96202 and NSF grant IRI 91-20074 and a DIMACS postdoctoral fellowship.
sridhar@cs.princeton.edu
[2]Supported in part by an NSF Postdoctoral Fellowship.
schulman@cc.gatech.edu

III. Analysis

We will have in mind a "space-time" diagram, where space corresponds to the topology of the network. By a path in space-time we mean a sequence of nodes $\{p_\tau\}$ in the network such that for each τ, p_τ and $p_{\tau+1}$ are adjacent [1] in the network. For a protocol Π, denote by $\Pi(t)$ the state (i.e. the combined state of all the processors) after t time steps. We show that there is a c such that for each noiseless network protocol Π there is a protocol Σ simulating Π on the same network $\mathcal{N}$, so that in the presence of noise, $\Sigma(T)$ fails to reproduce $\Pi(t)$ only if there is a space-time path on which there are at least $T - ct$ corrupted bit transmissions. This follows from an (slightly involved) argument relating the delay of the simulation, which is only defined locally due to the asynchronous nature of the simulation, to a space-time path containing a corresponding number of errors. The argument uses the combinatorial properties of both the protocol and tree codes. The probability of having that many transmission errors is then bounded in standard fashion by using the channel model.

From these steps we obtain theorem I.1. Similar results can be derived for more general channel models which require only a replacement of the last segment of the argument.

IV. Discussion

Our results are non-constructive: though we can show the existence of a simulation Σ, we are unable to produce Σ explicitly. The impediment is exactly that explicit algorithmic constructions for tree codes are not known. Moreover, the problem of decoding tree codes is solvable given a "coding oracle." Resolving the computational complexity of coding and decoding tree codes is the most critical open issue at the conclusion of our work.

Second, there is a storage space overhead which is separate from that incurred due to the cost of coding and decoding. This comes from the need, in our simulation Σ, for processors to be able to roll back computation; so in our protocol processors keep a record of all their past states.

If one is willing to tolerate error probabilities of the order of $N/\text{poly}(T)$, then the above problems can be addressed: the storage overhead can be greatly reduced, and a sufficiently good tree code can be constructed.

We conjecture that the $\log(d+1)$ slowdown in our theorem is necessary.

References

[1] L. J. Schulman. Coding for interactive communication. Manuscript. Abstract in these proceedings.

[2] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423; 623–656, 1948.

[3] A. C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the 11th Annual Symposium on Theory of Computing*, pages 209–213, 1979.

[1]A processor is adjacent to itself for this purpose.

Random Access from Compressed Datasets with Perfect Value Hashing

John W. Miller
Microsoft Research, One Microsoft Way, Redmond, WA 98052

Abstract — **A representation technique is presented allowing for quick access of individual records from a static compressed dataset. Given a collection of key-record pairs, the representation allows the appropriate short record to be returned for any given key. The approach is a generalization of *Perfect Address Hashing*. The new approach, called *Perfect Value Hashing*, uses a carefully chosen pseudo-random number generator to directly produce the correct record for any key in the dataset. This contrasts with Address Hashing where the random number provides an address which is then used to recover the record from a separate table. Value Hashing doesn't have the theoretical limitations of Address Hashing, and in practice is more space efficient for records of size less than 36 bits. Value Hashing has the added benefit (important when the records are encoded for compression) that variable length records can be represented without an increase in the size of the encoded records. This new technique was used to provide random access from a highly compressed spelling dictionary.**

I. Background of the Problem

Given a dataset of key-record pairs, the general problem is to represent the dataset so that a record can be recovered without a slow search. A well known solution is to sort the keys and store them with each fixed length record. This, for example, is the method used to organize a conventional phone book. Lookup requires a search logarithmic in the number of records. A faster lookup can be performed by storing with the dictionary a pseudo-random number generator called an Address Hash. This function takes any key and returns a number which is used to tell where the associated record is stored. The equivalent example with a phone book is where an Address Hash would convert a name (the key), into a page (the address) where the phone number (the record) is stored. This allows for faster access because the search is now limited to one page of the phone book, and so the speed of access is independent of the total size of the book. It is possible to represent this information much more space efficiently by not storing the key at all. A *Perfect Address Hash* is a specially created Address Hash for a particular dataset which produces a different address for every record (i.e. every page in the example phone book has exactly one phone number). It provides for time efficient access because no search is needed among the records at a given address. An important result of Perfect Address Hashing by Melhorn [1] is that an overhead of approximately $(1/n)\ \ln(n^n/n!)/\ln(2) \approx 1.44$bits is required to map n keys to n unique addresses (additional overhead is required if the records do not have a fixed length). Practical algorithms for finding Perfect Address Hash functions for large number of records (10^6) have been reported with a cost of 3.6 bits per record [2]. For small variable length compressed records, the size of the Perfect Address Hash function may be unacceptably large compared to the size of the compressed records.

II. Solution

Value Hashing is the method of using a pseudo-random number generator to calculate information about the record itself. For the phone book example a pseudo-random number generator would be created so that the number it returns for a given name is that person's phone number (or the bits of a prefix encoded representation of the phone number). This approach overcomes Melhorn's theoretical bound on overhead because each key does not map to a unique address. The achievability of the Slepian-Wolf [3] bound for broadcast channels [4] shows that the size of the Value Hash function (at least in theory) can be made independent of n (and so the overhead goes to zero for large n). This is obvious if you consider that a random sequence of bits will duplicate the records of a database of size n with probability $(1/2)^n$. In principle you could create Perfect Value Hashes by evaluating approximately 2^n hash parameters to see if they happen to regenerate the desired records for the keys in the dataset, and then encoding the index of the first successful mapping of keys to records. Using the entropy of the waiting time for first success (assuming each hash function is an independent trial producing all bit sequences of a given length equiprobably) it is easily shown that this index encoding requires approximately $n + 1/\ln(2)$ bits on average. By breaking down the search for hash functions into groups of k bits it is possible to do small combinatoric searches on subsets of k bits from n, so that the total overhead is approximately $(n/k)(k + 1/\ln(2))$. (Ramakrishna noted that brute force is effective in finding Perfect Address Hash functions and proposed a composition scheme for minimizing worse case evaluation time [5-6].) With current computer speeds $k = 16$ is easily achievable which implies .09 bits per binary record. A practical algorithm has been developed for finding Perfect Value Hash functions with an overhead of .1 bits per binary record. The average time required to evaluate the hash function is independent of n. The same technique can be used for non-binary records for increased speed in evaluation. This has been used to provide random access by key of any 4-bits from a highly compressed spelling dictionary.

References

[1] K. Melhorn, "On the program size of perfect and universal hash functions," *Proceedings of the 23th IEEE Symposium on Foundations of Computer Science*, (Chicago, Ill.), p.170–175, 1982.

[2] E.A. Fox et.al., "Practical Minimal Perfect Hash Functions for Large Databases," *Communications of the ACM*, vol. 35, 1. p.105–121, 1992.

[3] D. Slepian and J.K. Wolf, " Coding Theorem for Multiple Access Channels with Correlated Sources," *Bell Systems Tech. Journal*, vol. 52, p.1037–1076, 1973.

[4] Thomas M. Cover and Joy Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[5] M.V. Ramakrishna and P. Larson, "File Organization Using Composite Perfect Hashing," *ACM Transactions on Database Systems*, vol. 14, 2, p. 231–263, 1989.

[6] M.V. Ramakrishna, "Simple Perfect Hashing Method for Static Sets," *Proceedings of the 4th International Conference on Computing and Information*, p.401–404, 1992.

Information Retrieval from Databases

Nikolai K. N. Leung, John T. Coffey and Stuart Sechrest[1]
Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, MI 48109, U.S.A.

Abstract — **This study further investigates and generalizes the database model introduced in [1] by applying new techniques to the problem of data retrieval. The problems analyzed are representative of important issues involved in storing data for *context dependent* retrieval from databases. They arise when simple storage devices such as tapes and disks are used to store relatively more complex data structures such as large multi-dimensional images. The mismatch between the physical nature of the storage device and the data structure, i.e., the manner in which its elements are requested, prevents some requests from being instantaneously accessible on the database. Hence, we have the non-trivial problem of designing the database so as to minimize the expected access time EA.**

I. Introduction

The basic model in [1] is generalized to a large multi-dimensional image stored onto a lower dimensional tape where the sequence of user requests is modelled as a random walk on the image. It is found that careful use of redundancy in the storage scheme can reduce access time significantly over the no-redundancy case. As an interesting information theory problem, we examine what is the minimum expected access time that can be achieved under any system using redundancy, a cache, and multiple tapes (possibly implementing erasure-correcting codes).

II. Expected Access Time, EA

Under a linear cost function and no redundancy, we find that the minimum access time, EA^*, is dependent on the function ψ which we define as the **absolute central moment** of a graph. For a graph $\mathcal{B}$, $\psi(\mathcal{B}) = \sum_{b\in\mathcal{B}} d(b,c)$, where c is the center of $\mathcal{B}$ and $d(b,c)$ is the graph distance between the points b and c. It is found that when storing a d-dimensional toroidal image $\mathbf{I}$ onto a t-dimensional toroidal tape $\mathbf{T}$ of equal volume under a linear cost function without redundancy, the minimum access time EA^* is bounded by

$$EA^* \geq \frac{\psi(T)}{\psi(I)}$$

where each pixel/block of the image is represented as a node in the graph. In particular, the above bound is found to be tight when storing rectangular images onto 1-dimensional tape loops, i.e., $EA^* = \frac{n^{d-1}}{d} + \mathcal{O}(n^{d-2})$ where n is the length of the sides of a cube image.

We also introduce a slight variation to the problem where the tape is a loop and the read head is restricted to moving in only one direction. We begin with expressions for the minimum achievable access time for storing images onto the unidirectional tape under a linear cost, $EA^* = \frac{1}{2}\text{vol}(\mathbf{I})$ and capped cost function, $EA^* = \frac{1}{2}(\ell+1)$. It is then demonstrated that redundancy can be used to improve EA^* significantly under a capped cost function, $EA^* \leq \sqrt{2\ell}$. On the other hand, we find simple counter examples where redundancy only makes performance worse. In fact it is conjectured that redundancy can not improve performance under the linear cost function for this model. This is found to be an interesting question in its own right and can be modelled in a game theory context[2].

III. Caching and Multiple Tapes

When storing a one-dimensional image without redundancy, a cache of size $\mathcal{C}$ can be shown to reduce access time by a factor of at least $\mathcal{C}+1$. On the other hand EA^* cannot be reduced by more than $\frac{1}{4}\frac{\mathcal{C}^2+3\mathcal{C}}{1-2^{\mathcal{C}-n}} + 1$ where n is the size of the image. We then examine how a cache and redundancy can be applied together to further improve performance. In cases where exact reconstructions are not required to satisfy user requests, we model the problem in a rate-distortion context and explore achievable distortion-access time pairs and plan to relate this work to that described in [3].

The cache problem is extended to utilizing multiple tapes/heads to improve performance. Using T tapes under a block/file segmentation scheme, the access time is reduced by a factor of T. [4] [2] demonstrate problems where Reed-Solomon codes can be used to achieve a significant improvement over file segementation. Using such erasure-correcting codes for the multiple tape problem, it is found that accessed data elements that would incur high retrieval costs can be treated as *erasures*. Then using data from the other tape heads the erasure can be reconstructed using the code.

IV. Conclusions

The preliminary results from analyzing the representative models discussed suggest that the most effective means for improving EA^* is to use redundancy in the storage device. It is demonstrated however that this has to be done very carefully otherwise performance degrades. Deciding whether to use a cache or multiple tape heads is less significant and the relative importance of the two depends on the model used.

References

[1] John T. Coffey, Tal Herbsman, and Stuart Sechrest. *Communication Theory and Applications II*, chapter Information Theory Approaches to Retrieval from Databases. HW Communications Ltd., United Kingdom, 1993.

[2] M. Naor and R.M. Roth. Optimal file sharing in distributed networks. *Proceedings of 32nd Annual Symposium on Foundations of Computer Science*, pages 269-275, Puerto Rico, Oct. 1991.

[3] James R. Roche. Gambling for the mnemonically impaired. *Proceedings of the IEEE International Symposium on Information Theory*, page 184. Norway, June 1994.

[4] James R. Roche. *Distributed Information Storage*. PhD thesis, Stanford University, Department of Electrical Engineering, 1992.

[1] This work was supported in part by the National Science Foundation under grant NCR-9105832

[2] Proposed by Matthew Klimesh

Information Theory and Noisy Computation

William S. Evans[1]
Dept. of Computer Science, Univ. British Columbia

Leonard J. Schulman[2]
Dept. of Applied Mathematics, Weizmann Inst. Science
College of Computing, Georgia Inst. Technology

We report on two types of results. The first is a study of the rate of decay of information carried by a signal which is being propagated over a noisy channel. The second is a series of lower bounds on the depth, size, and component reliability of noisy logic circuits which are required to compute some function reliably. The arguments used for the circuit results are information-theoretic, and in particular, the signal decay result is essential to the depth lower bound.

Our first result can be viewed as a quantified version of the data processing lemma, for the case of Boolean random variables.

Theorem 1 (Signal Decay) *If X, Y are Boolean random variables and Z is the output of the channel $\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ on input Y then $\frac{I(X;Z)}{I(X;Y)} \leq \sin^2\theta$, where θ is the angle in the plane between the vectors $(\sqrt{1-a}, \sqrt{a})$ and $(\sqrt{b}, \sqrt{1-b})$.*

It is worth emphasizing that the bound holds regardless of the distribution on X and Y, and is a property of the channel alone. The bound is tight in that for any such channel, one can describe a joint distribution for X and Y so that $I(X;Z)/I(X;Y)$ is arbitrarily close to $\sin^2\theta$.

The previous theorem is a general result about mutual information. The remaining theorems concern the noisy circuit model of Von Neumann [7]. The signal decay theorem is useful in proving lower bounds on the structure of such circuits whose components (i.e. individual logic gates) fail with some probability. These results improve and simplify all previous lower bounds in this model.

Theorem 2 (Noisy Circuit Depth) *Let f be a Boolean function which depends on n inputs. Let C be a circuit of depth c using gates with at most k inputs, where each gate fails independently with probability $(1-\xi)/2$. Suppose C computes the function f correctly on all inputs with probability at least $1-\delta$ where $\delta < 1/2$. Let $\Delta = 1+\delta\log\delta+(1-\delta)\log(1-\delta)$.*

- *If $\xi^2 > 1/k$ then $c \geq \frac{\log(n\Delta)}{\log(k\xi^2)}$*
- *If $\xi^2 \leq 1/k$ then $n \leq 1/\Delta$*

To prove this theorem, we analyze the mutual information between the input to the noisy circuit and its output. This information must be large since the circuit reliably computes the function f; yet, according to the signal decay theorem, each noisy gate in the circuit, when viewed as a noisy channel, decreases information. Together, these observations imply the lower bound on circuit depth. This improves on the lower bounds of Pippenger [5] and Feder [1].

A similar technique, using a different measure of correlation than mutual information, provides a lower bound on noisy circuit size.

Theorem 3 (Noisy Circuit Size) *Let f be a Boolean function with sensitivity[1] s. Let C be a circuit using gates with at most k inputs, where each gate fails independently with probability $(1-\xi)/2$. Suppose C computes the function f correctly on all inputs with probability at least $1-\delta$ where $\delta < 1/2$, then the number of gates in C is at least $\frac{s\log s + 2s\log(2(1-2\delta))}{k\log t}$ where $t = \frac{\omega^3+(1-\omega)^3}{\omega(1-\omega)}$ and $\omega = \frac{1-\sqrt[k]{\xi}}{2}$.*

Previously, Gál [3], Reischuk and Schmeltz [6], and Gács and Gál [2] proved an $\Omega(s\log s)$ bound on reliable circuit size. Our improvement is in the bound's dependence on component reliability.

Finally, we establish a threshold for component reliability below which one cannot reliably compute all functions.

Theorem 4 (Component Reliability) *For k odd there exists $\delta < 1/2$ such that for all Boolean functions f there exists a formula[2] (using gates with at most k inputs, where each gate fails independently with probability ϵ) which computes f correctly on all inputs with probability at least $1-\delta$ if and only if*

$$\epsilon < \frac{1}{2} - \frac{2^{k-2}}{k\binom{k-1}{\frac{k-1}{2}}}.$$

This extends work done by Hajek and Weller [4], who showed the result for $k = 3$.

References

[1] T. Feder. Reliable computation by networks in the presence of noise. *IEEE Transactions on Information Theory*, 35(3):569–571, May 1989.

[2] P. Gács and A. Gál. Lower bounds for the complexity of reliable Boolean circuits with noisy gates. *IEEE Transactions on Information Theory*, 40(2):579–583, March 1994.

[3] A. Gál. Lower bounds for the complexity of reliable Boolean circuits with noisy gates. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, pages 594–601, 1991.

[4] B. Hajek and T. Weller. On the maximum tolerable noise for reliable computation by formulas. *IEEE Transactions on Information Theory*, 37(2):388–391, March 1991.

[5] N. Pippenger. Reliable computation by formulas in the presence of noise. *IEEE Transactions on Information Theory*, 34(2):194–197, March 1988.

[6] R. Reischuk and B. Schmeltz. Reliable computation with noisy circuits and decision trees — a general $n\log n$ lower bound. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, pages 602–611, 1991.

[7] J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 43–98. Princeton University Press, 1956.

[1]Supported by a Canadian International Fellowship and NSF grant CCR 92-01092. wevans@cs.ubc.ca

[2]Supported in part by an NSF Postdoctoral Fellowship. schulman@cc.gatech.edu

[1]The sensitivity of a function is the maximum (over all inputs) of the number of bits in the input which, when changed individually, change the function value.

[2]A formula is a circuit in which each gate has out-degree one.

Multiple repetition feedback coding for discrete memoryless channels

Thijs Veugen

Eindhoven University of Technology, Group on Information and Communication Theory, PO Box 513, 5600 MB Eindhoven, The Netherlands

Abstract — **It is shown that for a suitable choice of the parameters, multiple repetition feedback coding achieves a rate close to capacity for an arbitrary discrete memoryless channel. For wide- sense symmetric channels the difference between the rate of a multiple repetition feedback strategy and the channel capacity can be written as an informational divergence.**

I. Multiple repetition feedback coding

Consider a discrete memoryless channel with input symbols $0, \ldots, m-1$ and output symbols $0, \ldots, \hat{m}-1$. We assume w.l.o.g. [2] $\hat{m} \geq m$. The output symbols j, $m \leq j < \hat{m}$ can be seen as erasure symbols. The channel error probabilities are denoted by p_{ij} $(0 \leq i < m, 0 \leq j < \hat{m})$. The idea of repetition coding is the following: suppose during transmission of a message an $i \rightarrow j$ error occurs. The sender can detect this because of the feedback link and 'corrects' the error by repeating the symbol i a fixed number k_{ij} of times. The receiver scans the received sequence from right to left and replaces each subsequence $ji^{k_{ij}}$ by i. A consequence of repetition coding is that messages have to be precoded, since no subsequence $ji^{k_{ij}}$ may occur. For asymmetric channels precoding is also used to fix a symbol precoding distribution $\underline{q} = (q_0, \ldots, q_{m-1})$. This leads to a precoding rate $R_p(\underline{q})$. The expected number of transmissions to send symbol i such that all occuring transmission errors are corrected is $c_i = 1/(1 - \sum_j k_{ij} p_{ij})$. The rate of a repetition feedback strategy with repetition parameters k_{ij} $(0 \leq i < m, 0 \leq j < \hat{m})$ as a function of the symbol precoding distribution $\underline{q}$ is

$$R(\underline{q}) = \frac{R_p(\underline{q})}{\sum_i q_i c_i}$$

The rate R is the maximum of $R(\underline{q})$ over all symbol precoding distributions $\underline{q}$.

II. Wide-sense symmetric channels

A discrete memoryless channel is called wide-sense symmetric if the channel considered as a graph with labeled edges satisfies:

1. All input nodes have the same bag of outgoing edge labels.
2. All output nodes j, $0 \leq j < m$ have the same bag of incoming edge labels.
3. All output nodes j, $m \leq j < \hat{m}$ have the same bag of incoming edge labels.

Note that a bag is a set where elements can occur more than once. The labels of the edges that come in at output nodes j, $0 \leq j < m$, are (in arbitrary order) denoted by p_i $(0 \leq i < m)$. The labels of the edges that come in at output nodes j, $m \leq j < \hat{m}$, are (in arbitrary order) denoted by p_i $(m \leq i < \hat{m})$. A wide-sense symmetric channel has the property that capacity is achieved for a uniform input distribution.

Suppose a multiple repetition feedback strategy is used for such a channel. Each label p_i $(0 \leq i < \hat{m})$ will correspond to a repetition parameter k_i. Note that $k_i = 1$ for $m \leq i < \hat{m}$. For reasons of symmetry the symbol precoding distribution is no langer fixed during precoding, i.e. all messages without forbidden subsequences are allowed. The rate of the repetition strategy satisfies $R = (1 - \sum_{0 \leq j < \hat{m}} k_j p_j) \log_m x$, where $x > m$ is the solution of $\sum_{0 \leq i < m} x^{-k_i} = mx^{-1}$. From [1] follows that this rate is equal to the capacity of the channel when the channel error probabilities satisfy $p_i = \frac{1}{m} x^{1-k_i} \sum_{0 \leq j < m} p_j$ for $0 \leq i < m$, and $p_i = \frac{1}{\hat{m}-m} \sum_{m \leq j < \hat{m}} p_j$ for $m \leq i < \hat{m}$. Denote the solution of these equations by $\bar{p}_i$ $(0 \leq i < \hat{m})$. The following theorem shows how close the repetition strategy approaches channel capacity for an arbitrary wide-sense symmetric channel.

Theorem 1 *Consider an arbitrary wide-sense symmetric channel with characteristic channel error probabilities p_i $(0 \leq i < \hat{m})$ and capacity C. Let R be the rate of the multiple repetition feedback strategy with repetition parameters k_i $(0 \leq i < \hat{m})$. Then*

$$C - R = D_m((p_0, \ldots, p_{\hat{m}-1}) \| (\bar{p}_0, \ldots, \bar{p}_{\hat{m}-1}))$$

Here D_m denotes the m-ary informational divergence.

III. Arbitrary channels

For arbitrary channels it is difficult to obtain a simple expression indicating the exact distance between the rate of a multiple repetition feedback strategy and the channel capacity. However, it is possible to show that for a suitable choice of the repetition parameters, the rate will be close to capacity. When analysing the wide-sense symmetric case, it follows that the optimal channel error probabilities $\bar{p}_i$ $(0 \leq i < m)$ satisfy $k_i + \log_m(\bar{p}_i / \sum_{0 \leq j < m} p_j) \approx 0$ for large repetition parameters. Therefore, for arbitrary discrete memoryless channels with channel error probabilities p_{ij} $(0 \leq i < m, 0 \leq j < \hat{m})$, we suggest to use repetition parameters k_{ij} such that $k_{ij} \approx -\log_m(p_{ij} / \sum_{0 \leq s < m} p_{is})$ for $0 \leq j < m$. Note that k_{ij} should be equal to 1 for $m \leq j < \hat{m}$.

The rate of this suitably chosen multiple repetition feedback strategy as a function of the channel is denoted by $\bar{R}((p_{ij})_{ij})$. The following theorem indicates how close this rate approaches the channel capacity.

Theorem 2 *Consider an arbitrary discrete memoryless channel with channel error probabilities p_{ij} $(0 \leq i < m, 0 \leq j < \hat{m})$ and capacity C. If $p_{ij} \rightarrow 0$ for $0 \leq i \neq j < m$, then*

$$C - \bar{R} = \mathcal{O}\Big(\sum_{0 \leq i \neq j < m} p_{ij}\Big)$$

From Theorem 1 follows that the order of approximation in Theorem 2 is tight.

References

[1] Thijs Veugen Capacity achieving strategies for discrete memoryless channels with feedback In *1994 IEEE International Symposium on Information Theory*, page 466. June 1994

[2] Claude E. Shannon Some geometrical results in channel capacity *Nachrichtentechnische Zeit*, 10, 1957

EXTREMAL POLYPHASE SEQUENCES

Solomon W. Golomb

Department of Electrical Engineering–Systems, Communication Sciences Institute, University of Southern California, University Park, EEB-504A, Los Angeles, California 90089-2565, U.S.A.

SUMMARY

The unnormalized finite autocorrelation function $C(\tau)$ of the sequence $A = \{a_k\}_{k=1}^{n}$ of complex numbers on the unit circle is defined by $C(\tau) = \sum_{k=1}^{n-\tau} a_k a_{k+\tau}^*$, with $C(0) = n, C(-\tau) = C^*(\tau)$, and $|C(n-1)| = 1$, where z^* denotes the complex conjugate of z. We seek the sequence of length n, for each $n \geq 3$, which minimizes the value of

$$\max_{1 \leq \tau \leq n-2} |C(\tau)|,$$

and the value

$$T_n = \min_{all\ sequences} \max_{1 \leq \tau \leq n-2} |C(\tau)|$$

of this minimizing sequence.

As shown in [1], $\{|C(\tau)|\}_{\tau=-(n-1)}^{+(n-1)}$ is the same sequence for A, for $A^* = \{a_k^*\}_{k=1}^{n}$, for $A' = \{a_{n+1-k}\}_{k=1}^{n}$, and for each $A_{\alpha\beta} = \{\alpha\beta^k a_k\}_{k=1}^{n}$ where α and β are complex numbers with $|\alpha| = |\beta| = 1$.

A sequence A with $\max_{1 \leq \tau \leq n-2} |C(\tau)| \leq 1$ is called a *generalized Barker sequence* [1]. A clever hill-climbing program is described in [2], which reported empirical values of T_n for $3 \leq n \leq 25$, with the sequences attaining these values. In particular, generalized Barker sequences are claimed for all $n \leq 25$ except for $n = 20$. (A few of these examples are erroneous, although some or all of these errors may be typesetting mistakes.) No effort was made in [2] to describe the extremal sequences algebraically nor to make use of the group G of correlation-magnitude-preserving transformations to simplify the presentation of the data (as done in [1]). Here are best sequences $A_n = \{a_k\}_{k=1}^{n}$, and the corresponding correlation sequences $\{|C(\tau)|\}_{\tau=1}^{n}$, for $n = 3, 4$, and 6, expressed algebraically.

n	a_1	a_2	a_3	a_4	a_5	a_6
3	1	1	-1			
4	1	1	$-e^{i\gamma}$	$-e^{3i\gamma}$		
5						
6	1	1	$e^{\pi i/3}$	-1	1	$e^{4\pi i/3}$

(Here, $\gamma = cos^{-1}(1/4) = 75°.52248781\cdots$.)

n	$\|C(1)\|$	$\|C(2)\|$	$\|C(3)\|$	$\|C(4)\|$	$\|C(5)\|$	T_n
3	0	(1)				0
4	1/2	1/2	(1)			1/2
5						
6	1	1	1	1	(1)	1

The uniqueness (modulo the group G) of the sequence of length 6 was shown in [3]. It is believed that the methods used to obtain these results can be extended to other values of n.

REFERENCES

[1] S.W. Golomb and R.A. Scholtz, "Generalized Barker Sequences," *IEEE Trans. Info. Theory*, vol. IT-11, no. 4, October, 1965, 533-537.

[2] L. Bömer and M. Antweiler, "Polyphase Barker Sequences," *Electronic Letters*, vol. 25, no. 23, 9 November, 1989, 1577-1579.

[3] N. Zhang and S.W. Golomb, "Uniqueness of the Generalized Barker Sequence of Length 6," *IEEE Trans. Info. Theory*, vol. IT-35, no. 5, September, 1990, 1167-1170.

A Unified Construction of Perfect Polyphase Sequences

Wai Ho Mow[1]

Dept. of Elect. & Comp. Eng., Univ. of Waterloo, Waterloo, Ontario, CANADA N2L 3G1

Polyphase sequences over N-th complex roots of unity are considered. A sequence is *perfect* if all its out-of-phase periodic autocorrelation equal zero. Over the past 30 years, numerous constructions of *perfect polyphase sequences* (PPS) have been proposed due to their importance in various applications such as pulse compression radars, fast-startup equalization and channel estimation, and spread spectrum multiple access systems. We show that all previous PPS constructions, known to us, can be classified into four classes (c.f. [6]): (i) Generalized Frank sequences due to Kumar, Scholtz and Welch [4, Thm.3], (ii) Generalized chirp-like polyphase sequences due to Popović [7], (iii) Milewski sequences [5], (iv) PPS associated with the general construction of generalized bent function due to Chung and Kumar [1]. The key result here is a unified construction of PPS which includes the above four classes as special cases. Note, however, that only *explicit* constructions of PPS are considered in this work, since PPS obtainable by applying appropriate transformations to one or more previously explicitly constructed PPS are always obtainable from the unified construction in the same manner. Many useful transformations of this kind can be found in [1, Thm.1],[2],[3, Thm.2],[4].

For a polyphase sequence $\{\exp(2\pi\sqrt{-1}h(k)/L)\}_{k=0}^{L-1}$, we call $\{h(k)\}_{k=0}^{L-1}$ its index sequence, whose components need not be an integer.

Theorem 1 *Let $L = sm^2$, for $s, m \in \mathbf{Z}^+$. The polyphase sequence of length L defined by its index sequence*

$$\begin{aligned} f(km+l) &= \frac{m^2(s+1)}{2}\left(r_0 + n_0\frac{l(l+1)}{2}\right)k^2 \\ &\quad + m(r_1\pi(l) + n_1)k + f(l) \\ &\qquad \forall l \in \mathbf{Z}_m, \forall k \in \mathbf{Z}_{sm} \end{aligned} \tag{1}$$

where r_0 is any integer in $\mathbf{Z}_s$ coprime to s, n_0 is any integer in $\mathbf{Z}_s$ such that $(s+1)n_0$ is even and $r_0 + n_0 l(l+1)/2$ is coprime to s for all $l \in \mathbf{Z}_m$, r_1 is any integer in $\mathbf{Z}_{sm}$ coprime to m, n_1 is any integer in $\mathbf{Z}_{sm}$, π is an arbitrary permutation of the elements of $\mathbf{Z}_m$, and $f(l), \forall l \in \mathbf{Z}_m$, is an arbitrary rational-valued function, is perfect.

The number of distinct PPS for a large subset of (1) is determined below.

Theorem 2 *For the construction (1), the number of perfect polyphase sequences of length $L = sm^2$ and alphabet size N is $m!N^m$ for $s = 1$; and $sm(m!)\phi(s)N^m$ for $s > 1$, $n_0 = 0$ and $r_1 = 1$, where the Euler's function $\phi(u)$ is the number of integers in $\{1, 2, \cdots, u-1\}$ coprime to u.*

Comparing with exhaustive search results for all PPS satisfying $N \leq 15$, $L \leq 20$ and $N^L \leq 11^{11}$ (c.f. [6]), Theorem 2 predicts the exact numbers of all PPS found except for $(L, N) = (12, 6)$. Hence, Theorem 2 gives an excellent lower bound on the number of PPS for a given L and N. A computer program, which finds all PPS derivable from (1) proves that Theorem 1 in fact generates all possible PPS in the above search range [6].

We conjecture a simple relationship between the length and the minimum alphabet size.

Conjecture 1 *Let $L = sm^2$, for $s, m \in \mathbf{Z}^+$ and s is square-free. A perfect polyphase sequence of length L exists if and only if its alphabet size N is an integer multiple of $N_{\min}$ where $N_{\min}$ is the minimum alphabet size given by*

$$N_{\min} = \begin{cases} 2sm & \text{for even } s \text{ and odd } m, \\ sm & \text{else.} \end{cases} \tag{2}$$

This conjecture is closely related to some famous open problems such as the nonexistence of Barker sequences, circulant Hadamard matrices, and one-dimensional generalized bent functions [6].

With the unified construction (1), it is not difficult to build optimal sequence sets, with respect to the Sarwate bound [8], that generalizes all previously known constructions of this kind.

Theorem 3 *Denote the smallest prime divisor of L by p. Then the set of $p-1$ perfect polyphase sequences of length L as defined in Theorem 1 with $n_0 = 0$, π the identity map and $r_0 = r_1$ an element of $\{1, 2, \cdots, p-1\}$, n_1 an arbitrary integer, and $f(l), \forall l \in \mathbf{Z}_m$ an arbitrary rational-valued function, is optimal with respect to the Sarwate bound.*

References

[1] H. Chung and P. V. Kumar, "A new general construction for generalized bent functions," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 206–209, 1989.

[2] E. M. Gabidulin, "Further results on perfect auto-correlation PSK sequences," *Proceedings of the 1st International Symposium on Communication and Applications*, UK, 1991.

[3] E. M. Gabidulin, "Non-binary sequences with the perfect periodic auto-correlation and with optimal periodic cross-correlation," in *1993 IEEE International Symposium on Information Theory (ISIT' 93)*, pp. 412, Jan. 1993.

[4] P. V. Kumar, R. A. Scholtz and L. R. Welch, "Generalized bent functions and their properties," *J. Combin. Theory*, Series A, vol. 40, pp. 90–107, 1985.

[5] A. Milewski, "Periodic sequences with optimal properties for channel estimation and fast start-up equalization," *IBM J. Res. Develop.*, vol. 27, pp. 426–431, 1983.

[6] W. H. Mow, "A Study of Correlation of Sequences," PhD Thesis, Dept. of Information Engineering, the Chinese University of Hong Kong, Shatin, Hong Kong, May 1993.

[7] B. M. Popović, "Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 1406–1409, 1992.

[8] D. V. Sarwate, "Bounds on crosscorrelation and autocorrelation of sequences," *IEEE Trans. Inform. Theory*, vol. IT-25, pp.720–724, 1979.

[1]This work was supported by the Croucher Foundation Fellowship 1994/95.

Asymptotic Autocorrelation of Golomb Sequences

E. M. Gabidulin
Moscow Institute of Physics and Technology
Institutskii per. 9, 141700 Dolgoprudnyi
Russia, E-mail:gab@ippi.msk.su

P. Z. Fan and M. Darnell
Department of Electronic and Electrical Engineering
Leeds University, Leeds LS2 9JT
UK. Email: p.fan@ieee.org

Abstract — **Golomb sequences of length L form a class of polyphase sequences which have a perfect periodic autocorrelation. Given certain constraints, they also have a favorable aperiodic autocorrelation. This paper presents a comprehensive study of the asymptotic behavior of the aperiodic autocorrelation function of Golomb sequences.**

I. Introduction

In 1953, R. H. Barker [1] introduced binary sequences with particularly favorable aperiodic autocorrelation functions (ACFs). In 1965, Golomb and Scholtz [2] proposed a class of generalized polyphase Barker sequences which satisfy the original Barker constraint on aperiodic autocorrelation. In order to obtain a larger number of sequences with favourable aperiodic correlation, Golomb [3] defined a class of infinite classes of sequences. For Golomb sequences of arbitrary length L,

$$a_{r,k} = e^{i\frac{\pi r(k-1)k}{L}}, \quad 1 \le k \le L, \ (r, L) = 1, \tag{1}$$

Zhang and Golomb [3] proved that the maximum out-of-phase aperiodic autocorrelation value with $r = 1$ (and $r = L - 1$) is bounded by $\sqrt{L/4.438}$. When L is odd, $r = \frac{L\pm1}{2}$, Fan, Darnell and Honary [4] further showed that the out-of phase aperiodic autocorrelation value of Golomb sequences is asymptotically bounded by $\sqrt{L/2.174}$. In this paper, we study the general asymptotic behavior of Golomb sequences.

II. Basic Properties of Golomb Sequences

The maximum out-of-phase autocorrelation value of the sequence $a_{r,k}$ is given by

$$B_r = \max_{\tau=1,2,\ldots,L-1} |C_r(\tau)| = |C_r(I_m(L))|, \tag{2}$$

where $C_r(\tau) = \sum_{j=1}^{L-\tau} a_{r,j} a^*_{r,j+\tau}$, $I_m(L)$ is the value of $\tau (0 < \tau < L)$ which maximizes $|C(\tau)|$.

For Golomb sequences, it is simple to show that

Lemma 1

$$C_r(\tau) = -\frac{\sin\frac{\pi r}{L}\tau^2}{\sin\frac{\pi r}{L}\tau}. \tag{3}$$

$$C_r(\tau) = (-1)^{(L-1)r+1} C_r(L-\tau), \quad C_r(\tau) = C_{L-r}(\tau). \tag{4}$$

Thus we need only consider the values of $C_r(\tau)$ in the range of $1 \le \tau \le \left[\frac{L}{2}\right]$ and $1 \le r \le \left[\frac{L-1}{2}\right]$.

III. Asymptotic Behavior of the Aperiodic ACF of Golomb Sequences

Based on Lemma 1, we have the following asymptotic bound:

Theorem 1

$$B_r(b) \approx \begin{cases} 0.48\sqrt{\frac{b}{r}}\,L, & I_m(L) = \frac{(Lb-1)s_0}{r}, \\ & r \ge 2, \ 0 \le \frac{b}{r} \le 0.37, \\ \frac{L}{\pi}\sin\left(\frac{\pi b}{r}\right), & I_m(L) = \frac{Lb-1}{r}, \\ & r \ge 2, \ 0.5 \ge \frac{b}{r} \ge 0.37. \end{cases} \tag{5}$$

where $bL \equiv \pm 1 \ mod \ r$, $1 \le b \le \left[\frac{r}{2}\right]$, $s_0 = \sqrt{\frac{z_0 r}{\pi b}}$ *and* $z_0 = 1.1655$.

When $r = 1$, which is excluded from above derivation, we have the following result which is the same as in [3] but the derivation is simpler.

Theorem 2

$$B_1 = \sqrt{L/4.34}, \quad I_m(L) = \sqrt{L/2.68}. \tag{6}$$

The result given in [4] can be obtained directly from Eqn 5.

Corollary 1 *If L is odd and $r = \frac{L-1}{2}$, then*

$$B_{\frac{L-1}{2}} = \sqrt{L/2.17}, \quad I_m(L) = \sqrt{L/1.34}. \tag{7}$$

IV. Summary

In conclusion, we have considered the asymptotic maximum out-of-phase ACF of Golomb sequences of arbitrary length L and order r. It is shown that the B_r is bounded by $\sqrt{L/4.34}$ if $r = 1$; or $0.48\sqrt{b/r}L$, if $r \ge 2$, $b/r \le 0.37$; or $L/\pi \sin \pi b/r$, if $r \ge 2$, $b/r > 0.37$.

References

[1] R. H. Barker, *Communication Theory(Jackson, W., Ed.)*, ch. Group synchronising of binary digital systems, pp. 273 – 287. Butterworths, London, 1953.

[2] S. W. Golomb and R. A. Scholtz, "Generalised Barker sequences," *IEEE Trans. on Information Theory*, vol. IT-11, pp. 533–537, October 1965.

[3] N. Zhang and S. W. Golomb, "Polyphase sequence with low autocorrelations," *IEEE Trans. on Information Theory*, vol. IT - 39, pp. 1085–1089, May 1993.

[4] P. Z. Fan, M. Darnell, and B. Honary, "Polyphase sequences with good periodic and aperiodic autocorrelations," in *1994'IEEE Symposium on Information Theory*, (Trondheim, Norway), p. 145, June 1994.

Perfect Sequences Derived from M-sequences

M. Darnell and P. Z. Fan
Department of Electronic and Electrical Engineering
Leeds University, Leeds LS2 9JT
UK. Email: p.fan@ieee.org

Fan Jin
Dept of Computer Science & Engineering
Southwest Jiaotong University, Chengdu, 610031
P.R. of China

Abstract — **New classes of multi-level and complex sequences with perfect periodic autocorrelations are presented. The sequences are derived directly from certain m-sequences over rational and Gaussian integers.**

I. Quasi-perfect Multi-level Sequences

In their basic form, p-level m-sequences comprise the rational integers $0, 1, 2, \cdots, (p-1)$, where p is a prime. To derive a practical bipolar sequence from such an m-sequence, integer and sinusoidal level transformations can be used [1]. Both these transformations yield bipolar signals with useful periodic ACF properties. For $p = 3$ and 5, the integer level transformation gives bipolar IR sequences $A = \{a_j\}$ of length $L = 2N$ with quasi-perfect periodic ACFs of the form:

$$\theta_A(l) = \begin{cases} P, & l = 0 \\ -P, & l = N, \\ 0, & otherwise. \end{cases} \tag{1}$$

II. Quasi-perfect Complex Sequences

Let $h(x) = x^n + h_{n-1}x^{n-1} + \cdots + h_1 x + h_0$, $h_j \in G_\pi$, denote a primitive polynomial of degree n over residue class of Gaussian integer, G_π. A maximal length sequence $A = \{a_j\}$ over G_π can be obtained. It is shown that most of the properties of the complex m-sequences are similar to those of maximal length sequence over Galois fields; however, there are some particular properties which are distinct [2]. Specifically, two sub-classes of complex m-sequences of length $L = 4N$ with the following quasi-perfect autocorrelation function have been obtained by letting $\pi = 2 + i$ and $\pi = 3i$, which correspond to $p = 5$ and $p = 3$ respectively.

$$\theta_A(l) = \sum_{k=0}^{L-1} a_k a_{k+l}^* = \begin{cases} P, & l = 0; \\ i\,P, & l = N; \\ -P, & l = 2N; \\ -i\,P, & l = 3N; \\ 0, & otherwise. \end{cases} \tag{2}$$

III. Synthesis of Perfect Multilevel and Complex Sequences

If two component multi-level sequences $A = \{a_j\}$ of period $L = 2N$ and $B = \{(-1)^j\}$ of period 2 are combined using digit-by-digit multiplication, the periodic ACF of the resulting composite sequence C, $\theta_C(l)$, is given by

$$\theta_C(l) = (-1)^l \theta_A(l) = \begin{cases} \theta_A(l), & l = 0 \ mod\ 2 \\ -\theta_A(l), & l = 1 \ mod\ 2 \end{cases} \tag{3}$$

If sequence A is chosen as a transformed p-level m-sequence with quasi-perfect ACF, and the length of this sequence A is exactly divisible by 2 to give an odd integer N, then due to the inverse-repeat (IR) format of A, the digit-by-digit multiplication process yields a multi-level perfect sequence C' of period N: $C' = (c_0, c_1, \cdots, c_{N-1})$.

If the two component complex sequences $A = \{a_j\}$ of period $L = 4N$ and sequence $B = \{(i)^j\}$ of period 4 are combined using digit-by-digit multiplication, the periodic ACF of the resulting composite sequence C is given by

$$\theta_C(l) = (-i)^l \theta_A(l) = \begin{cases} \theta_A(l), & l = 0 \ mod\ 4 \\ -i\theta_A(l), & l = 1 \ mod\ 4 \\ -\theta_A(l), & l = 2 \ mod\ 4 \\ i\theta_A(l), & l = 3 \ mod\ 4 \end{cases} \tag{4}$$

Similarly, if the complex sequence A is a quasi-perfect sequence of period $L = 4N$, where N is an odd number, then the sequence C synthesised has a perfect ACF. Let $C' = (c_0, c_1, \cdots, c_{N-1})$, then C' is a perfect sequence of period N.

IV. Synthesis Examples

Firstly, consider the ternary m-sequence obtained using the integer level transformation. Given the values $p = 3$, $n = 3$ and $p^n - 1 = 26$, a perfect sequence of length 13 can be obtained: $C' = (0, -1, 1, -1, 0, 0, -1, 0, -1, -1, -1, 1, 1)$, $\theta_{C'} = (9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$

Secondly, consider a complex m-sequence of length $L = 5^3 - 1 = 124$ generated by primitive polynomial $h(x) = x^3 + ix^2 + x - i$ over G_{2+i}. Then a new perfect sequence C' can then be derived as follows: $C' = (0, 0, -1, -1, i, 1, i, -i, 1, 1, -1, 1, 1, 1, i, 0, -1, -i, 1, -i, 0, i, 0, i, 1, -1, i, i, 0, 1, -1)$, $\theta_{C'} = (25, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$

Acknowledgements

This work was supported in part by British Council and Chinese Government through the Academic Links with China Scheme (ALCS).

References

[1] M. Darnell, *Perturbation Signals for System Identification(Ed. K. R. Godfrey)*, ch. Periodic and non-periodic, binary and multi-level pseudo-random signals, pp. 176–208. Prentice-Hall, 1993.

[2] P. Z. Fan, M. Darnell, and B. Honary, "Maximal length sequences over gaussian integers," *Electron. Lett.*, vol. 30, pp. 762–764, August 1994.

Balanced Quadriphase Sequences with Near-ideal Autocorrelations

P. Z. Fan and M. Darnell
Dept of Electronic and Electrical Engineering,
Leeds University, Leeds, LS2 9JT, U.K., Email: p.fan@ieee.org

Abstract — **Given any prime $p \equiv 1$ *mod*4 and any positive integer m, a class of balanced quadriphase sequences of length $p^m - 1$ with near-ideal periodic autocorrelation properties is constructed. The quadriphase sequences are optimal under the condition of balanced sequence elements.**

I. Introduction

In 1977, Lempel, Cohn and Eastman describe a class of balanced binary sequences with optimal periodic autocorrelation properties [1]. Their work is related to the construction of orthogonal matrices [2, 3]. Given any odd prime p and any positive integer m, a balanced (± 1) binary sequence of length $p^m - 1$ whose out-of-phase autocorrelation function $R(\tau)$ satisfies $c(\tau) = +2$ or -2 for $(p^m - 1)/2$ odd, and $R(\tau) = 0$ or -4 for $(p^m - 1)/2$ even is obtained. It is shown that every balanced binary sequence must have at least two distinct out-of-phase correlation values which are at least as high as those obtained by Lempel et al. It is in this sense that their sequences are optimal.

In this paper we describe a generalization of balanced binary sequences to quadriphase sequences. It is shown that for any prime $p \equiv 1$ *mod*4 and any positive integer m, a class of balanced quadriphase sequences of length $p^m - 1$ with near-ideal periodic autocorrelation properties can be constructed. The quadriphase sequences obtained are also optimal under the condition of balanced sequence elements.

II. Main Result

Consider a finite field $F = GF(p^m)$, where $p \equiv 1$ *mod* 4 and m is a positive integer: let G denote the multiplicative group of F and α be any primitive element of F, i.e., α is a generator of the cyclic group G. Consider also the subset S_l of G defined by

$$S_l = \left\{\alpha^{4j+l} - 1\right\}_{j=0}^{k-1}, \quad l = 1, 2, 3;\ k = \frac{p^m - 1}{4} \tag{1}$$

$$S_4 = G \setminus (S_1 \cup S_2 \cup S_3) \tag{2}$$

Note that each S_l contains exactly one quarter of the elements of G and that every element of S_l is equal to some power of α.

Let f denote the mapping from G onto $\{1, i, -1, -i\}$ defined by

$$f(\alpha^t) = \begin{cases} 1, & if\ \alpha^t \in S_1 \\ i, & if\ \alpha^t \in S_2 \\ -1, & if\ \alpha^t \in S_3 \\ -i, & if\ \alpha^t \in S_4 \end{cases}, 0 \le t < 4k = p^m - 1 \tag{3}$$

or

$$f(\alpha^t) = i^l, \quad if\ \alpha^t \in S_l, \ \ 0 \le t < 4k = p^m - 1 \tag{4}$$

Based on the above, we can prove the following result:

Theorem 1 *The periodic autocorrelation function $R(\tau)$ of the quadriphase sequence $\mathbf{a} = a_0, a_1, \cdots, a_{4k-1}$, where $a_t = f(\alpha^t)$, $0 \le t \le 4k - 1$, satisfies $R(0) = 4k$ and, for $0 < \tau \le 4k - 1$,*

$$R(\tau) = \begin{cases} 0\ or \pm 2\ or \pm 2i\ or\ \ \pm 2 \pm 2i, & k = odd \\ 0\ or \pm 2\ or \pm 2i\ or\ \ \pm 2 \pm 2i & \\ or \pm 4\ or \pm 4i\ or\ \ \pm 4 \pm 4i, & k = even \end{cases} \tag{5}$$

Moreover, $\mathbf{a}$ is balanced and $R(\tau)$ is optimal, given the condition of balance.

III. Examples

Example 1: $p = 13$, $m = 1$, $k = \frac{p^m - 1}{4} = 3$, and $\alpha = 2$. For this set of parameters we obtain

$$\begin{array}{ll} \{\alpha^t: & 1, 2, 4, 8, 3, 6, 12, 11, 9, 5, 10, 7\ \} \\ \{f(\alpha^t): & 1, -i, 1, -i, i, -1, -i, i, i, 1, -1, -1\ \} \\ \{R(\tau): & 12, -2, -2, -2i, -2 + 2i, 2i, 0, -2i, -2 - 2i, 2i, \\ & -2, -2\ \} \end{array}$$

Example 2: $p = 5$, $m = 2$, $k = \frac{p^m - 1}{4} = 6$, and $\alpha = x = (0, 1)$. For this set of parameters we obtain

$$\begin{array}{ll} \{\alpha^t: & (1,0), (0,1), (2,2), (4,1), (2,1), (2,4), (3,0), (0,3), \\ & (1,1), (2,3), (1,3), (1,2), (4,0), (0,4), (3,3), (1,4), \\ & (3,4), (3,1), (2,0), (0,2), (4,4), (3,2), (4,2), (4,3)\ \} \\ \{f(\alpha^t): & i, -i, 1, 1, 1, -i, -i, i, -i, i, 1, i, -i, -1, -1, 1, -i, \\ & -1, i, -1, 1, i, -1, -1\ \} \\ \{R(\tau): & 24, -2 + 2i, 2i, -2 - 2i, 0, -2 + 2i, 0, 0, \\ & -2, -2 + 2i, 2i, 0, -4., 0, -2i, -2 - 2i, -2, \\ & 0, 0, -2 - 2i, 0, -2 + 2i, -2i, -2 - 2i\ \} \end{array}$$

References

[1] A. Lempel, M. Cohn, and W. Eastman, "A class of balanced binary sequences with optimal autocorrelation properties", *IEEE Trans. on IT*, vol. 23, no. 1, pp. 38–42, January 1977.

[2] J.M. Goethals and J. J. Seidel, "Orthogonal matrix with zero diagonal", *Can. J. Math*, vol. 19, pp. 1001–1010, 1967.

[3] R. J. Turyn, *Error Correcting Codes (Mann, H. B. Ed.)*, chapter Sequences with small correlation, pp. 195 – 228, Wiley, 1968.

Quasi-linear Synchronization Codes

A.J. van Wijngaarden

Institute for Experimental Mathematics, Ellernstr. 29, 45326 Essen, Germany

Abstract — **The use of quasi-linear synchronization (QLS) codes to provide synchronization of frames with fixed length n offers many advantages relative to comma-free codes and prefix synchronized codes. Easy frame location and the absence of data conversion enable a QLS-code to be implemented with very low complexity independent of the frame length. Another important aspect is the ability of error control in the presence of substitution errors. A list of optimal QLS-codes of length up to 40 obtained with elaborate computer search is presented. Several families of perfect and (sub) optimal QLS-codes with large word length n have been constructed, and also new upper bounds on the redundancy of the codes have been established.**

I. Introduction

A quasi-linear synchronization (QLS) code [1], being a coset of a linear code, allows easy encoding and decoding, easy frame location, and error control. Consider a code $\mathcal{C}$ of length n, being a subset of $\mathcal{A}_q^n$, where $\mathcal{A}_q$ denotes the q-ary alphabet $\{0, 1, \ldots, q-1\}$. The synchronization and error control properties of a code $\mathcal{C} \subset \mathcal{A}_q^n$ are determined by the code distance $d(\mathcal{C})$ (i.e. the minimal Hamming distance of the code), and by the code separation $\rho(\mathcal{C})$, defined by

$$\rho(\mathcal{C}) = \min_{\substack{0<i<n \\ X,Y,Z \in \mathcal{C}}} d(T_i(X,Y), Z) \ , \qquad (1)$$

with shift operator $T_i(X,Y) = x_i x_{i+1} \ldots x_{n-1} y_0 y_1 \ldots y_{i-1}$.

The code $\mathcal{C} \subset \mathcal{A}_q^n$ is called a quasi-linear synchronization code of length n and separation ρ, if for each code word $X \in \mathcal{C}$, a fixed set P of positions is used, establishing separation $\rho(\mathcal{C}) \geq \rho$ irrespective of the actual value of the other (data) positions. In this way, an arbitrary data word $D \in \mathcal{A}_q^m$ of length $m = n - |P|$ can be easily inserted at the data positions. A q-ary QLS-code of length n and separation ρ is called a QLS(q, n, ρ) code.

The use of distinct synchronization positions and unconstrained data positions allows easy encoding and decoding for any separation. QLS-codes with separation $\rho > 1$ can be used for error control coding. Correct synchronization and error correction can be guaranteed in the presence of no more than t substitution errors in n successive symbols for a code $\mathcal{C}$ with $d(\mathcal{C}) \geq (2t+1)$ and $\rho(\mathcal{C}) \geq (2t+1)$.

II. Bounds and Code Constructions

The redundancy R of a q-ary QLS-code of length n and separation ρ is, according to Levenshtein [1], bounded by $R \geq R_{\min}(q, n, \rho)$, with

$$R_{\min}(q, n, \rho) = \left\lceil \sqrt{\frac{q\rho(n-1)}{q-1}} \right\rceil . \qquad (2)$$

The construction of q-ary QLS-codes with arbitrary code separation is in general difficult, especially the construction of QLS-codes with minimal redundancy, so called optimal codes. Using constructions proposed by Levenshtein [1], optimal binary QLS-codes with separation $\rho \leq 2$ and redundancy $R_{\min}$ can always be obtained for any length n. For $\rho > 2$, two new upper bounds on the redundancy have been obtained for binary codes, based on construction methods. Firstly, a binary QLS-code can be constructed with redundancy $R_1(2, n, \rho)$, bounded by $R_1(2, n, \rho) \leq R_{\min}(2, n, \rho) + \varphi(\rho)$, for which $\rho - 2 \leq \varphi(\rho) \leq 3\rho - 2$. The term $\varphi(\rho)$ is independent of the length n, therefore the constructed codes are asymptotically optimal for $n \to \infty$ and $\rho/n \to 0$. Secondly, for $n > \rho(2\rho^2 + 2\rho + 1)$, a binary QLS-code can be constructed with redundancy $R_2(2, n, \rho)$, bounded by $R_2(2, n, \rho) \leq R_{\min}(2, n, \rho) + \rho - 1$.

Several search methods can be used to find optimal codes with larger separation ($\rho > 2$). In this way optimal QLS-codes with length n up to 40 have been found.

III. Combinatorial Construction Methods

The theory of difference sets [3] is sometimes applicable for the construction of QLS-codes. It is convenient to use the following combinatorial description of a q-ary QLS-code of length n and separation ρ. The index position set P is partitioned into q subsets $P_0, P_1, \ldots, P_{q-1}$ in such a manner that for each number $d \not\equiv 0 \pmod{n}$ there are at least ρ pairs (x_i, y_j), $x_i \in P_i$, $y_j \in P_j$, $i \neq j$, satisfying $x_i - y_j \equiv d \pmod{n}$. If there are exactly ρ pairs for each d, the code is called perfect.

It is apparently very difficult to construct perfect codes with arbitrary parameters n and ρ. Using the theory of difference sets, two families of perfect QLS-codes can be directly constructed. Firstly, for length $n = 4t^2 + 1$, t being an odd positive integer and n being prime, perfect QLS-codes with separation $\rho = (t^2+1)/2$ can be constructed. Secondly, for any length n, n prime, perfect QLS-codes with separation $\rho = (n-1)/2$ can be constructed. The redundancy is equal to $2t^2 + 1$ and $n - 1$ respectively. Both perfect codes turn out to be unique, i.e. there are no solutions with the same parameters which do not belong to this family. It is expected that several other families of perfect codes will be found for binary as well as q-ary codes.

IV. Conclusion

It has been shown that for binary QLS-codes of arbitrary length and separation constructions can be obtained which are close to optimal. For a large variety of QLS-codes, especially for the important category of small codes, optimal codes have been found. Using combinatorial methods, various families of perfect QLS-codes have been obtained as well.

References

[1] V.I. Levenshtein, "One method of constructing quasi-linear codes providing synchronization in the presence of errors", *Problems of Information Transmission*, vol. 7, no. 3, 1971, pp. 30–40.

[2] J.J. Stiffler, "Theory of Synchronous Communications", *Prentice Hall, Inc.*, Englewood Cliffs, New Jersey, 1971.

[3] M. Hall, "Combinatorial Theory", *Wiley-Interscience*, New York, second edition, 1986.

Extended Sonar Sequences

Oscar Moreno*[1], Solomon W. Golomb** and Carlos J. Corrada*

*Department of Mathematics and Computer Science, University of Puerto Rico, Rio Piedras, Puerto Rico 00931

**Communication Science Institute, Department of Electrical Engineering–Systems, University of Southern California, Los Angeles, CA 90089

Abstract — **Sonar as well as other related sequences were introduced by Golomb and Taylor in [2]. Following a similar approach,we introduce the concept of an extended sonar sequence. It is similar to that of a sonar sequence but blank columns are permitted. Several constructions of extended sonars are given. Our constructions are very close to ordinary sonar sequences. However they provide good improvements to the list of the best known constructions for sonar sequences up to 100 symbols given in [3].**

I. Introduction

Sonar sequences were introduced in [2] to deal with the following problem: "You have an object which is moving towards (or away) from you, and you want to know effectively your distance and velocity of the object."

The solution to the problem comes from using the Doppler effect: when a wave hits a moving object its frequency changes in direct proportion to the velocity of the object. In other words you send a wave, wait until it returns and from the time it takes you know the distance, from the new frequency you know the velocity. On the other hand since the world is noisy you might send out a wave that does not return. Consequently you send out m waves with frequencies ranging from 1 to n. Waves are sent out at times ranging from 1 to m. Once the whole pattern of waves returns, from the change in frequency you determine the velocity of your object and from the time change the distance. On the other hand if not all the frequencies return there might be some ambiguities as to what is the whole pattern. Sonars are those patterns for which you send out exactly one wave at every time and also for which even if only two waves return you can reconstruct the whole pattern. This last point means that there is no ambiguity. The problem for sonars is, given n frequencies, construct an n by m sonar sequence for m as large as possible.

II. Extended Sonar sequences

The point of this talk is that for the sonar application an alternative to sending exactly one wave at every time (the sonar case) is that of sending at most one wave, or in other words, choose not to send any wave in some time slots. This is done to achieve a larger number of waves sent for a given number of frequencies, while increasing the probability of receiving at least the two frequencies needed to reconstruct the whole sequence. Because of the similarity with the common sonar sequences, our sequences will be using the same equipment, in other words, it will be more cost effective than common sonar sequences. Again we would send m waves with frequencies ranging from 1 to n, and let us say that there are k blank (no wave) time slots, then we would call it an (n, m, k) extended sonar.

The case when $k = 0$ is that of a sonar, and the case of $n = 1$ reduces to what has been studied previously under the name of rulers, which have other applications besides radar and sonar to synchronization, crystallography, etc (see [1]). In other words extended sonar sequences are a natural generalization of sonars and also of rulers. The main point of the present talk is to give several constructions of extended sonars.

III. The Constructions

We will show how some of the constructions used in [3] to generate Costas and Sonar sequences, have a circular periodicity property that is the basis of our constructions of extended sonar sequences, namely the Extended Logarithmic Welch, the Extended Shift Sequence and the Extended Lempel-Golomb. This three constructions with $k = 1$, are very similar to ordinary sonars but for which the table of our constructions for n up to 100, outperforms the corresponding table of the best known construction for sonars given in [3]. For example for $n = 46$ and $n = 75$ it fills 7 more slots that common sonar sequences.

Also we have tested the performance of this constructions comparing them with the best possible extended sonar sequences obtained doing an extensive search. The problem of generating extended sonar sequences exhaustively with the computer resides in the fact that the time of computation increases exponentially. The only practical way to obtain a sonar or extended sonar sequence for large lengths is therefore by generating it with some particular construction. At the moment we have done the extensive search for up to $m = 10$. The constructions obtained the best possible value 60% of the time.

We will define the Circular extended sonar sequences and then we will prove that the Logarithmic Welch, the Shift Sequences and the Lempel-Golomb constructions give us a circular extended sonar sequences. We will show then that from any circular extended sonar sequences, we can obtain n extended sonar sequences.

Then we will apply a series of transformations to the resulting extended sonar sequence to obtain a sequence with a reduced number of symbols obtaining the best known extended sonar sequences.

[1] Work partially supported by NSF grants RII-9014056, component IV of the EPSCoR of Puerto Rico Grant and the ARO grant for Cornell MSI.

References

[1] G. S. Bloom and S. W. Golomb, *Applications of numbered undirected graphs*, Proceedings of the IEEE, 65:4 (1977), pp. 562-570.

[2] S. W. Golomb and H. Taylor, *Two -Dimensional Synchronization Patterns for Minimum Ambiguity*, IEEE Transactions on Information Theory, 28:4 (1982), pp. 600–604.

[3] O. Moreno, R. A. Games and H. Taylor, *Sonar Sequences from Costas arrays and the best known sonar sequences with up to 100 symbols*, IEEE (1992)

Extended Prefix Synchronization Codes

A.J. van Wijngaarden and H. Morita*
Institute for Experimental Mathematics, Ellernstr. 29, 45326 Essen, Germany
* Dept. of Comp. Science & Information Math., U.E.C., Chofu, Tokyo 182, Japan

Abstract — **A new synchronization code construction technique is presented which uses a so called extended prefix containing positions with fixed symbols and unconstrained data positions, followed by a constrained data sequence. In this way a set of prefixes is used to identify the frame, instead of only one prefix, as for normal prefix synchronized (PS) codes. This enlarges the code size, while the advantages of PS-codes, i.e. easy frame recognition and the availability of data mapping procedures, are maintained.**

I. Introduction

Synchronization of fixed length frames can be performed using comma-free codes [1]. Several maximal comma-free codes can be constructed [1], but both frame recognition and data mapping tend to be very complex. One solution is to use so-called prefix synchronized (PS) codes, introduced by Gilbert [2], and further analyzed by Guibas and Odlyzko [3]. A PS-code $\mathcal{C}_P(k+m)$ is defined as a set of code words of length $n = k + m$ with q-ary symbols of the alphabet $\mathcal{A}_q$, with the property that for any code word $p_1 p_2 \dots p_k c_1 c_2 \dots c_m$ the prefix $P = p_1 p_2 \dots p_k$ does not appear anywhere in the sequence $p_2 \dots p_k c_1 \dots c_m p_1 \dots p_{k-1}$.

We will modify the marker by lifting the condition to use consecutive fixed symbols. The modified marker is called an *extended* prefix. After a formal definition of extended prefix synchronized (EPS) codes, the construction of extended prefixes will be described, and expressions for the cardinality will be derived in order to compare the EPS-codes with PS-codes. Finally, a data mapping procedure will be presented.

II. Code description

Prior to giving an exact definition of prefix synchronized codes, the correlation between two sequences will be defined. For two sequences X and Y of length n the *correlation* X over Y, denoted by $X \circ Y$, is a binary vector $r_1 r_2 \dots r_n$, with r_i is 1 if the subsequence $x_i x_{i+1} \dots x_n$ equals $y_1 y_2 \dots y_{n-i}$, and 0 otherwise.

For q-ary PS-codes with $q \leq 4$, prefixes P of size k with correlation $P \circ P = 10^{k-1}$ maximize the code set [3]. These PS-codes have the following form: $\mathcal{C}_P(k+m) = P\, \mathcal{F}_P(m)$, where $\mathcal{F}_P(m)$ denotes the set of constrained sequences $c_1 \dots c_m$ in which P does not appear as a subsequence.

As an example of extended prefixes, let us consider the set of two patterns 11000 and 11010, denoted by 110*0. The code set $\mathcal{C}_{110*0}(5+m)$ is the union of the sets $11000\, \mathcal{F}_{110*0}(m)$ and $11010\, \mathcal{F}_{110*0}(m)$, where $\mathcal{F}_{110*0}(m) = \mathcal{F}_{11000}(m) \cap \mathcal{F}_{11010}(m)$. We notice that each code word in $\mathcal{C}_{110*0}(5+m)$ belongs to a PS-code of prefix 11000 or to a PS-code of prefix 11010. The advantage of this extended marker is to obtain one unconstrained position (fourth position) while the disadvantage is to force the remaining part of each code word to be more constrained, since neither 11000 nor 11010 are allowed to occur in the constrained sequence. The cardinality of $\mathcal{C}_{110*0}(5+m)$ is equal to $2\,|\,\mathcal{F}_{11000}(m) \cap \mathcal{F}_{11010}(m)\,|$. According to Gilbert [2], prefixes of length 4, e.g. 1100, have the maximal code size among PS-codes of length 11 upto 21. In this range the inequality $|\mathcal{C}_{110*0}(n)| > |\mathcal{C}_{1100}(n)|$ always holds, e.g. for $n = 13$, EPS-code $\mathcal{C}_{110*0}(13)$ with 384 code words is 17.8% larger than PS-code $\mathcal{C}_{1100}(13)$ with 326 code words.

An extended prefix synchronization code uses an extended marker $\mathcal{P}$ of length h with k fixed positions and $h-k$ unconstrained data positions. In fact, $\mathcal{P}$ is a set of q^{h-k} different prefixes. For every pair of prefixes $P_i, P_j \in \mathcal{P}$ the correlation $P_i \circ P_j$ is equal to 10^{h-1} if $i = j$, and 0^h otherwise. In this case the code $\mathcal{C}_{\mathcal{P}}(h+m)$ with extended prefix $\mathcal{P}$ is defined by

$$\mathcal{C}_{\mathcal{P}}(h+m) = \mathcal{P}\, \mathcal{F}_{\mathcal{P}}(m) = \bigcup_{P_i \in \mathcal{P}} \left\{ P_i \left\{ \bigcap_{P_i \in \mathcal{P}} \mathcal{F}_{P_i}(m) \right\} \right\}$$

We will show that for each $P_i \in \mathcal{P}$, $\mathcal{C}_{\mathcal{P}}(h+m)$ is a PS-code with prefix P_i. The cardinality of an EPS-code, $\mathcal{C}_{\mathcal{P}}(h+m)$, equals $q^{h-k} F_{\mathcal{P}}(m)$, where $F_{\mathcal{P}}(m)$ denotes the size of $\mathcal{F}_{\mathcal{P}}(m)$.

Theorem 1 *An extended prefix synchronized code $\mathcal{C}_{\mathcal{P}}(h+m)$ with extended marker $\mathcal{P}$ of length h and k fixed positions, has generating function*

$$F^g_{\mathcal{P}}(z) = \frac{1}{1 - qz + q^{h-k} z^h}, \tag{1}$$

which provides the following recursive formula for $F_{\mathcal{P}}(m)$:

$$F_{\mathcal{P}}(m) = \begin{cases} q^m & 0 \leq m < h \\ q\, F_{\mathcal{P}}(m-1) - q^{h-k} F_{\mathcal{P}}(m-h) & m \geq h. \end{cases}$$

We found that, for given k, binary EPS-codes having an extended prefix of the form $P = 1^t 0\, (*^{t-1} 0)^{k-t-1}$ with $t = \lceil k/2 \rceil$ have maximal cardinality. Mapping procedures have been developed which associate each number x in the range $0 \leq x < F_{\mathcal{P}}(m)$ with a unique word of the set $\mathcal{F}_{\mathcal{P}}(m)$ and vice versa.

III. Conclusion

A new, so called extended prefix synchronized code has been presented, as well as methods to construct extended prefixes, an expression to exactly determine the cardinality of an arbitrary q-ary EPS-code, and a mapping procedure to generate codes with maximal code size. EPS-codes allow easy frame detection and have a coding complexity which is roughly equivalent to PS-codes, and prove to have a larger code size compared to the traditional PS-codes.

References

[1] S.W. Golomb, B. Gordon, L.R. Welch, "Comma-free codes", *Can. J. Mathematics*, vol. 10, no. 2, pp. 202-209, 1958.

[2] E.N. Gilbert, "Synchronization of binary messages", *IRE Trans. on Information Theory*, Sept. 1960, pp. 470–477.

[3] L.J. Guibas, A.M. Odlyzko, "Maximal prefix-synchronized codes", *SIAM J. Appl. Math.*, vol. 35, no. 2, Sept. 1978, pp. 401–418.

Maximal Prefix Synchronized Codes by means of Enumerative Coding

Hiroyoshi Morita†, Adriaan van Wijngaarden * and A.J. Han Vinck *
†Graduate School of Information Systems, University of Electro-Communications, Chofu, Tokyo 182, Japan
*Institute for Experimental Mathematics, Ellernstr. 29, 45326 Essen, Germany

Abstract — **A systematic procedure for mapping data sequences into code words of a binary maximal prefix synchronized (MPS) code as well for the inverse mapping is presented. The complexity of the proposed scheme is proportional to the code word length. In order to be able to choose another prefix, e.g. a Barker sequence, methods will be presented which convert an MPS code into other MPS code with a different prefix. Both the mapping algorithm and the conversion algorithm can be generalized for q-ary prefix synchronized codes.**

I. Introduction

A prefix-synchronized (PS) code, introduced by Gilbert [1] and further analyzed by Guibas and Odlyzko [2], is a collection of code words of length $k+n$ over an alphabet $\mathcal{A}_q$ of size q whose first k symbols equal the prefix $P = p_1 p_2 \dots p_k$, and in addition, any code word $p_1 \dots p_k c_1 \dots c_n$ satisfies the constraint that P does not appear as a block of k consecutive symbols anywhere in $p_2 \dots p_k c_1 \dots c_n p_1 \dots p_{k-1}$. Let $\mathcal{G}_P^{(k+n)}$ be a maximal PS (MPS) code which maximizes the code size among all PS codes with the same parameters n and P of length k.

The advantage of PS codes relative to maximal comma-free codes [3] is easy word synchronization recovery, since the decoder only has to search for the appearance of P in the incoming stream symbols. In this paper, we propose simple encoding and decoding algorithms for an MPS code $\mathcal{G}_P^{(k+n)}$ with *self-uncorrelated* prefix P such that no prefix of P matches any suffix of P.

II. Recursive Subdivision of MPS-codes

For a prefix P of length $k \geq 1$, let $\mathcal{F}_P^{(n)}$ denote the set of sequences of length n such that no P appears at any position as a string of k consecutive symbols. The autocorrelation of a sequence $X = x_1 x_2 \dots x_m$ of length m is defined as a binary sequence Y of length m, satisfying the property that $y_i = 1$ if the prefix $x_1 \dots x_{m-i+1}$ of X and the suffix $x_i \dots x_m$ of X are identical, otherwise zero. The autocorrelation of X is denoted by $X \circ X$. As an example, if $P = 1110$, then $P \circ P = 1000$. Note that $P \circ P = 10^{k-1}$ *iff* P is self-uncorrelated. For two strings X and Y, let XY be the concatenation of X and Y. Moreover, for a string X and a set of strings $\mathcal{F}$, let $X\mathcal{F}$ be $\{XY | Y \in \mathcal{F}\}$. The followings are main results we obtained.

Theorem 1 *For a prefix P of length k, $P\mathcal{F}_P^{(n)}$ is an MPS code $\mathcal{G}_P^{(k+n)}$ if and only if P is self-uncorrelated.*

Theorem 2 *Let P_G be $1^{k-1}0$. Then it follows that*

$$\mathcal{F}_{P_G}^{(n)} = \{1^n\} \cup \bigcup_{i=1}^{k-1} \left\{ 1^{i-1} 0 \, \mathcal{F}_{P_G}^{(n-i)} \right\}.$$

Note that the recursion on the size of $\mathcal{F}_{P_G}^{(n)}$ obtained from theorem 2 gives us a strictly larger code than Mandelbaum's code [4] in which Fibonacci recursions [5] are used.

III. Mapping Algorithm for MPS-codes

Theorem 2 shows the division of $\mathcal{F}_{P_G}^{(n)}$ into k distinct subsets. By recursively applying the theorem to each subset except the singleton set (consisting of only one element), $\mathcal{F}_{P_G}^{(n)}$ can be represented as a collection of $G_{k,n}$ singleton sets where $G_{k,n}$ is the cardinality of $\mathcal{G}_{P_G}^{(n+k)}$. We present an algorithm to find a singleton set corresponding to an input x with $0 \leq x < G_{k,n}$ and the inverse algorithm as well. The scheme is based on enumerative coding [6].

IV. MPS-code Conversion

In practical situations, one might use another prefix than $P_G = 1^{k-1}0$, e.g. a Barker sequence [7]. We show an algorithm to transform a code word in $\mathcal{G}_{P_G}^{(k+n)}$ to another in $\mathcal{G}_Q^{(k+n)}$ if $Q \circ Q = 10^{k-1}$ holds: Scanning $V \in \mathcal{G}_{P_G}^{(k+n)}$ from the left to the right, replace Q with P_G when Q is found on V. Then the sequence obtained must belong to $\mathcal{G}_Q^{(k+n)}$ if the last symbol q_k of Q is 0. In case $q_k = 1$, negating V and replacing P_G with the negation of P_G, the above conversion procedure transforms $V \in \mathcal{G}_{P_G}^{(k+n)}$ to a code word in $\mathcal{G}_Q^{(k+n)}$.

V. Conclusion

In conclusion of his paper [4], Mandelbaum makes comment on his codes as follows: "These codes seem to have the best efficiency of all comma-free codes that can be constructed systematically (no table lookup)." We disprove his statement. A systematic procedure for mapping data sequences into code words of a binary MPS code of prefix P_G as well for the inverse mapping is presented. The complexity of the proposed scheme is proportional to the code word length n. In order to enable the choice of another prefix, methods are presented which transform $\mathcal{G}_{P_G}^{(k+n)}$ into any $\mathcal{G}_Q^{(k+n)}$ of self-uncorrelated prefix Q. The mapping procedure and the conversion method can be generalized for q-ary prefix synchronized codes.

References

[1] E. Gilbert, "Synchronization of Binary Messages," *IRE Trans. on Information Theory*, vol. IT-6, pp. 470 – 477, 1960.

[2] L.J. Guibas, A.M. Odlyzko, "Maximal prefix-synchronized codes," *SIAM J. Appl. Math.*, vol. 35, pp. 401–418, 1978.

[3] S.W. Golomb, B. Gordon, L.R. Welch, "Comma-free codes," *Can. J. Mathematics*, vol. 10, no. 2, pp. 202–209, 1958.

[4] D.M. Mandelbaum, "Synchronization of codes by means of Kautz's Fibonacci encoding," *IEEE Trans. on Information Theory*, vol. IT-18, pp. 281–285, 1972.

[5] W.H. Kautz, "Fibonacci codes for synchronization control," *IEEE Trans. on Information Theory*, vol. IT-11, pp. 284–292, 1965.

[6] T. Cover, "Enumerative source encoding," *IEEE Trans. on Information Theory*, vol. IT-19, pp. 73–77, 1973.

[7] R.H. Barker, "Group Synchronizing of Binary Digital Systems," *Communication Theory*, C. Cherry, ed., pp. 273–287, 1953.

Partial Classification of Sequences with Perfect Auto-Correlation and Bent Functions

Ernst M. Gabidulin[1]
Moscow Institute of Physics and Technology
Institutskiĭ per., 9; 141700 Dolgoprudnyi, Russia
E-mail: gab@ippi.ac.msk.su

Abstract — **Complex valued sequences of length n are considered. A sequence is said to be a perfect sequence if all the out-of-phase periodic autocorrelation coefficients are equal to zero. A sequence is said to be a phase shift keyed (PSK) sequence if all the coordinates are on the unit circle. A sequence is said to be a polyphase sequence if all the coordinates are n'th roots of unity. For the case n is a power of a prime integer, the partial classification of perfect PSK sequences is given. As a consequence, the full classification of one dimensional bent functions is presented.**

I. Introduction

Let $\mathbf{x} = (x_0, x_1, \ldots, x_{n-1})$ be a complex valued sequence of length n containing at least one non-zero component. The periodic cross-correlation function of sequences $\mathbf{x}$ and $\mathbf{y}$ is given by $R_{\mathbf{x},\mathbf{y}}(\tau) = \sum_{s=0}^{n-1} x_s y_{s+\tau}^*$, $\tau = 0, 1, \ldots, n-1$, where all the indices are calculated mod n and x^* denotes the complex conjugation of x. The periodic autocorrelation function (PAF) of $\mathbf{x}$ is defined by $R_{\mathbf{x}}(\tau) := R_{\mathbf{x},\mathbf{x}}(\tau)$. The "energy" of the sequence $\mathbf{x}$ is given by $R_{\mathbf{x}}(0) = \sum_{s=0}^{n-1} |x_s|^2 > 0$. .

Consider a set of sequences $\mathcal{M} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$, where $\mathbf{x}_i = (x_{i,0}, x_{i,1}, \ldots, x_{i,n-1})$. Let $\mathcal{R}(\mathcal{M}) := \left\{|R_{\mathbf{x}_i,\mathbf{x}_j}(\tau)|, \tau = 0, 1, \ldots, n-1,\ i, j = 1, 2, \ldots, M\right\}$ be the set of absolute values of periodic auto and cross correlation coefficients.

The following simultaneous linear transformations of the $\mathbf{x}_i$'s do not change the set $\mathcal{R}(\mathcal{M})$: 1) **Projectivity:** $\mathbf{z}_i = a\mathbf{x}_i$, $i = 0, 1, 2, \ldots, M$, where $|a| = 1$ is a complex number on the unit circle. 2) **Cyclic shift:** $z_{i,j} = x_{i,j+1}$, $i = 0, 1, 2, \ldots, M$; $j = 0, 1, 2, \ldots, n-1$. 3) **Permutation group:** $z_{i,j} = x_{i,kj \pmod n}$, $i = 1, 2, \ldots, M$; $j = 0, 1, \ldots, n-1$, where $\gcd(k, n) = 1$. 4) **"Linear frequency modulation":** $z_{i,j} = x_{i,j}\zeta^{sj}$, $i = 0, 1, 2, \ldots, M$; $j = 0, 1, 2, \ldots, n-1$, where s is an integer, ζ is a primitive root of unity of degree n. 5) **Conjugation:** $z_{i,j} = x_{i,j}^*$, $i = 0, 1, 2, \ldots, M$; $j = 0, 1, 2, \ldots, n-1$.

We refer to sequences $\mathbf{z}$ and $\mathbf{x}$ as *equivalent* sequences if the sequence $\mathbf{z}$ can be obtained from the sequence $\mathbf{x}$ using a number of these transformations.

II. Perfect Sequences

Let $\mathbf{P} = (1/\sqrt{n})\left[\zeta^{ij}\right]$, $i, j = 0, 1, \ldots, n-1$ be the matrix of the Discrete Fourier Transform (DFT) of dimension n. Let $\mathbf{y} = \mathbf{xP}$ be the DFT of a sequence $\mathbf{x} = (x_0, x_1, \ldots, x_{n-1})$.

Theorem 1 [1] *A sequence* $\mathbf{x}$ *is a* **perfect** *sequence if and only if all the Fourier components, i.e., the components of* $\mathbf{y}$*, have the same magnitude .*

Theorem 1 gives the complete description of the set of all general perfect sequences. An important problem in the theory of perfect sequences is finding non-equivalent sequences or finding the dimension of a set of perfect sequences.

[1]This work was supported by Grant NKF 000

Theorem 2 [1] $\dim \mathcal{P}_n = n - 1$, *where* $\mathcal{P}_n$ *denotes the set of non-equivalent perfect sequences of energy* n.

Theorem 3 [2] *The number of non-equivalent perfect PSK sequences of length* $n = p_1 p_2 \ldots p_m$, *where* p_i*'s are distinct primes, is finite.*

The situation is quite different when n is not square free. In this case, there are infinitely many non-equivalent perfect PSK sequences.

Theorem 4 *The maximal dimension of a set of* **perfect** *PSK sequences of lenght* $n = p^{2m}$ *or* $n = p^{2m+1}$ *is equal to* $k = p^m - 1$. *Such sets can be constructed in the explicit form.*

III. Perfect Polyphase sequences

Polyphase sequences is a special subset of PSK sequences.

An index function $f(x)$ *is known as a one-dimensional* **bent function** *if and only if the corresponding sequence* $\mathbf{x} = (\zeta^{f(0)}, \zeta^{f(1)}, \ldots, \zeta^{f(n-1)})$ *is perfect.*

The number of different bent functions is finite. A general construction of bent functions is given in [3]. Nevertheless, this construction does not describe *all* the bent functions. We give the full classification of bent functions.

Theorem 5 *If* $n = 2m$, m *odd, then a bent function does not exist.*

Theorem 6 *Let* $n = p$, $p \geq 3$ *is a prime. All the bent functions are quadratic polynomials* $f(x) = ax^2 + bx + c$, $a, b, c \in \mathbf{Z}_p$, $a \neq 0$.

Theorem 7 *Let* $n = p^{2k}$. *Let* x_0 *and* x_1 *be the unique representation of* x *given by* $x = x_0 + x_1 p^k$, *where* $0 \leq x_0 \leq p^k - 1$, $0 \leq x_1 \leq p^k - 1$. *Then all the bent functions of length* n *are given by*

$$f(x) = F(x_0) + x_1 G(x_0) p^k, \qquad (1)$$

where $F(x_0)$ *is a function taking values in* $\mathbf{Z}_n$ *and* $G(x_0)$ *is a function taking values in* $\mathbf{Z}_{p^k}$ *such that* $G(a) \neq G(b)$, *if* $a \neq b$, $a, b \in \mathbf{Z}_{p^k}$.

For the case $n = p^{2k+1}$, there exists a similar theorem.

References

[1] 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.

[2] E.M. Gabidulin, "Further Results on Perfect Auto-Correlation PSK Sequences," *Proceedings of the 1st International Symposium on Communication & Applications*, UK, 1991.

[3] H. Chung and P.V. Kumar, "A New General Construction for Generalized Bent Functions," *IEEE Trans. Inform. Theory*, vol.IT-35, pp. 206-209, 1989.

Codes and Iterative Decoding on General Graphs

Niclas Wiberg, Hans-Andrea Loeliger, and Ralph Kötter
ISY, Linköping University, S-58183 Linköping, Sweden

Keywords: **turbo (de-)coding, low-density parity-check codes, Tanner graphs, Markov random fields, "trellises" with a generalized "time" axis, generalized Viterbi and BCJR decoding.**

I. Introduction

Until recently, most known decoding procedures for error-correcting codes were based either on algebraically *calculating* the error pattern or on some sort of *tree or trellis search.* With the advent of turbo coding [1], a third decoding principle has finally had its breakthrough: *iterative decoding.*

(Iterative decoding is not a new idea, though: most of the key ideas were already present in Gallager's work on low-density parity-check codes [2].)

With respect to Viterbi decoding, a code is most naturally described by means of a trellis diagram. The main thesis of the present paper is that, with respect to iterative decoding, the natural way of describing a code is by means of a Tanner graph [3], which may be viewed as a generalized trellis. More precisely, it is the "time axis" of a trellis that is generalized to a Tanner graph.

Trellises yield Tanner graphs of the type shown in Fig. 1; in particular, the graph has no cycles. The complexity reduction in turbo codes (and low-density parity-check codes, and many new codes to be discovered) comes from allowing Tanner graphs with cycles, cf. Fig. 2.

II. Decoding

Both Viterbi decoding and BCJR decoding [4] are easily generalized to arbitrary Tanner graphs *without cycles,* where these algorithms are still optimal (in the same sense as for trellises). The basic idea of iterative decoding is simply to apply these algorithms even to Tanner graphs *with* cycles, ignoring the fact that the algorithms are no longer optimal. The empirical success of turbo coding (as well as our own experiments with other types of codes) confirm the validity of this approach.

Of course, analytical understanding of the decoder operation is also desirable. Our main result here applies to "cycle codes" (a subclass of low-density parity-check codes): we give a complete algebraic characterization of all error patterns that are corrected by the generalized Viterbi algorithm after infinitely many iterations.

III. Realization Theory on General Graphs

Much recent work was devoted to finding, and bounding the size of, the "smallest" trellis for a given code. This problem is significantly generalized by considering general Tanner graphs.

In the traditional setting, the only degree of freedom (for a given code) was the ordering of the "time axis". For a given ordering, every linear code has a well-defined unique minimal trellis, and every other trellis for the same code may be collapsed to the minimal trellis by state merging.

In our more general setting, the "time axis" need not be ordered, but may be an arbitrary Tanner graph. Even for a fixed Tanner graph, there is, in general, no unique minimal trellis. (The simplest example are tail-biting trellises.) Nevertheless, bounds on the "size" of the realization may be obtained from the ("abstract") state spaces of the code.

IV. A Priori Probabilities

Our careful derivation of the two basic iterative decoding algorithms clarifies, in particular, what a priori distributions are admissible and how they are properly dealt with. As it turns out, these distributions are closely related to Markov Random fields.

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes (1)," *Proc. ICC'93,* Geneva, Switzerland, 1993, pp. 1064–1070.

[2] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory,* vol. 8, pp. 21–28, Jan. 1962.

[3] R. M. Tanner, "A recursive approach to low-complexity codes," *IEEE Trans. Inform. Theory,* vol. 27, pp. 533–547, Sept. 1981.

[4] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory,* vol. 20, pp. 284–287, March 1974.

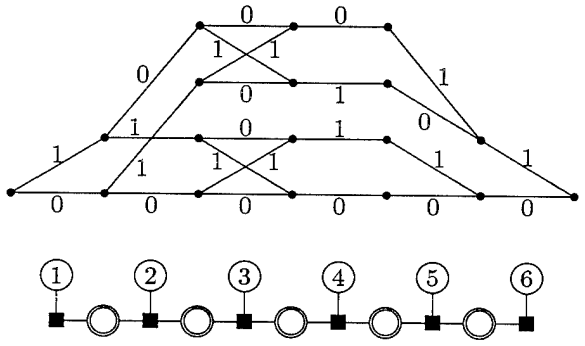


Figure 1: A trellis (top) and its Tanner graph (bottom).

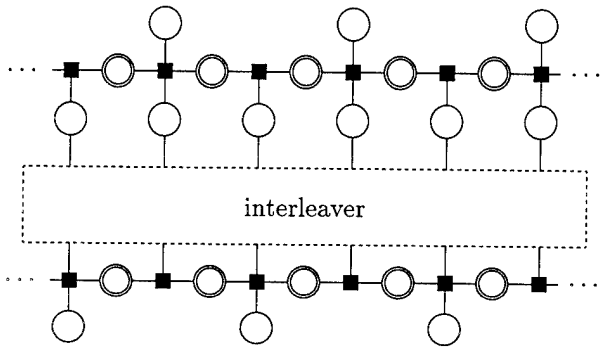


Figure 2: The Tanner graph of turbo coding [1].

Concatenated Coding System with Iterated Sequential Inner Decoding

Ole Riis Jensen (riis@it.dtu.dk) & Erik Paaske (ep@it.dtu.dk)

Institute of Telecommunication, Technical University of Denmark, DK-2800 Lyngby, Denmark

Abstract — We describe a concatenated coding system with iterated sequential inner decoding. The system uses convolutional codes of very long constraint length and operates on iterations between an inner Fano decoder and an outer Reed-Solomon decoder.

I. Introduction

We consider a concatenated system with a convolutional inner code, a block interleaver of degree I, and I outer RS codes of the same length, but with different redundancies/error correcting capabilities. After encoding by the outer codes and interleaving, the frame is split into a number of subframes. These are encoded by a memory M convolutional code, which is terminated by M input zeroes. The decoding for the inner code is performed by a number of sequential Fano decoders, which perform forward and backward (i. e. starting from the end of the subframe) decoding simultaneously, and on all the subframes in parallel. The process is monitored such that decoded symbols from the inner code in each of the RS words are counted. The non-decoded symbols are treated as erasures.

In a chosen implementation we use I=8, the error correctional profile for the outer code [16 50 16 6 6 16 6 6], and M=23. The three first decoders, and the 6th RS decoders are errors-and-erasure decoders which can correct e erasures as long as $e + 2t \le 100$ and $e \le 68$, respectively $e + 2t \le 32$ and $e \le 16$. The other RS words are errors-only decoders.

The first RS decoding attempt is then performed on the second RS word when 187 decoded symbols in this word are available from the inner decoders, and in case of a decoding failure (more than 50 errors detected) a new attempt is performed each time a new decoded symbol is available from the inner decoders. When the word is decoded the result is fed back and used to guide the sequential decoders in the continued decoding, i.e. the sequential decoders are forced to follow paths in the tree which agree with the RS decoded data. Decoding more and more RS words and feeding the results back to the inner decoders will in this way iterate the process towards a succesful decoding of the full frame. If 3 consequetive RS words (24 bits >M) are decoded, the forced inner decoding will effectively split the full frame into sub-sub-frames of length 48 bits, which can be decoded independently in both directions. If the decoding in one of these is stuck, a jump to the next sub-sub-frame can be made with only a small penalty.

II. Results for Iterated Sequential Decoding

In a system where sequential decoding is used the code should have a good (or optimum) distance profile (DP) together, of course, with a large free distance. However, if decoding is performed forward and backward on a frame (or subframe) a suitable code must also have a good distance profile in its reversed form, since this is the code used in the backward decoding. From a code search we obtained the following ODP memory M = 23, d_f=54 code written in hexadecimal form

$$G = [\ 96A77B\ \ B7EA67\ \ D0A25D\ \ E1C4D9\ ]$$

which also has a very good DP in reversed form.

In the simulations we have used an AWGN channel quantized into 16 levels, and the quantizing thresholds are E_s/N_o-dependent as is common practice. The Fano decoders use an ordinary FANO metric, which in our case is unquantized (32 bit floating-point words). As a preliminary value of Δ we have used the ratio $\Delta/bm_{max} = 6$, where bm_{max} is the maximum branch metric. We have chosen to use interleaving degree I = 8, but a further gain may be available by increasing the interleaving degree. A good choise for the number of subframes was determined through simulations to be 15. The choice of profile (and rate) for the outer codes is by no means obvious and deserves further investigation. Our simulation approach includes a number of different profiles, and the one chosen here, has proven the best results. Very good profiles with only two different outer codes does also exist.

In Figure 1 we have shown the average number of computations C_{av} found by simulation runs of 1000 frames (a total of 14,368,000 information bits) with different signal to noise ratios. No errors appeared at all. The E_b/N_o values specified are the net values for the entire system. Since the overall rate of the system is $R_{overall} = 0.216$ (including the small loss introduced by the termination of subframes) we notice that the inner sequential decoder operates at an E_s/N_o which is 6.66 dB below E_b/N_o. With a computational cut-off rate R_{comp} that falls below the convolutional code rate of 1/4 for $E_b/N_o < 2.55$ dB these results support our claim that a sequential decoder can operate well above R_{comp} if some kind of side information is available. In this case the side information is achieved by using an outer code.

We notice that for $E_b/N_o = 1.0$ dB we can build a decoder with a C_{av} that is at least 100 times smaller than the 16,384 decoding operations used by the Viterbi decoder for the n=4, M=14 code used in the Galileo mission. For $E_b/N_o = 1.0$ dB no frame required more than 2000 computations per bit and very few frames required more than 500 computations per bit. When we consider the $E_b/N_o = 0.6$ dB case 3% of the frames requires more than 10,000 computations, and 2% of the frames requires more than the Galileo code Viterbi decoder.

III. Conclusion

We have described a very efficient scheme utilizing iterated sequential decoding. However, the number of computations depends on the profile chosen and on the strategy used for the inner decoding, but as demonstated, very good results can be obtained.

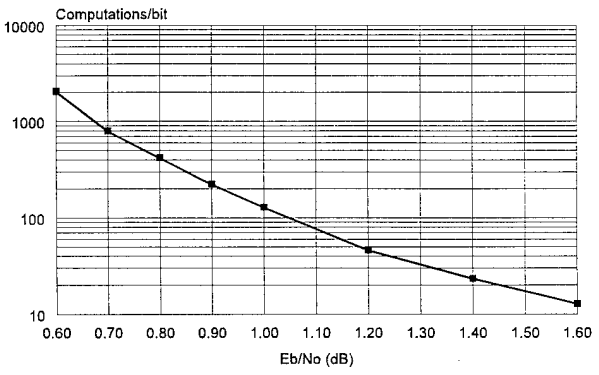


Figure 1: Average number of computations per bit as function of E_b/N_o.

The Least Stringent Sufficient Condition on the Optimality of Suboptimally Decoded Codewords[1]

Tadao Kasami[†] Takuya Koumoto[†] Toyoo Takata[†] Toru Fujiwara[††] Shu Lin[†††]

[†] Graduate School of Information Science, Nara Institute of Science and Technology
[††] Faculty of Engineering Science, Osaka University
[†††] Department of Electrical Engineering, University of Hawaii at Manoa

I. Introduction

The number of iterations of an iterative optimal or suboptimal decoding scheme [1–5] for binary linear block codes without any effect on its error performance can be reduced by testing a sufficient condition on the optimality of a candidate codeword. In this paper, the least stringent sufficient condition on the optimality of a decoded codeword is investigated under the assumption that the available information on the code is restricted to (1) the minimum weight or the distance profile and (2) for a given positive integer h, h or fewer already generated candidate codewords. The least stringent sufficient conditions of optimality for $1 \leq h \leq 3$ are presented. $Cond_1$ is the same as the one given in [2], $Cond_2$ is less stringent than the one given in [3], and $Cond_1$ and $Cond_2$ are derived from $Cond_3$ as special cases. These conditions can be used effectively to save computer simulation time for evaluating error probability for maximum likelihood decoding.

As examples, we consider Chase Algorithm II [1] and two iterative decoding algorithms [5,6] for $RM_{5,1}$, $RM_{5,2}$, $RM_{6,2}$, and $RM_{6,3}$, where $RM_{m,r}$ denotes the r-th order Reed-Muller codes of length 2^m. Majority-logic decoding with randomly breaking ties is used to generate candidate codewords. For an AWGN channel and BPSK, the effectiveness of $Cond_h$ for $1 \leq h \leq 3$ is evaluated by simulation.

II. Sufficient Conditions on the Optimality of a Decoded Codeword

Suppose a binary block code C of length N with distance profile W is used for error control over the AWGN channel using BPSK. A codeword $\boldsymbol{c}$ is mapped into $\boldsymbol{x} \in \{-1,1\}^N$. Suppose $\boldsymbol{x}$ is transmitted and $\boldsymbol{r}$ is received sequence at the output of matched filter of the receiver. Let $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$ be the binary hard-decision sequence obtained from $\boldsymbol{r}$.

Let V_N denote the set of all binary N-tuples. For $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$ in V_N, $D_1(\boldsymbol{u}) \triangleq \{i : u_i \neq z_i, \text{ and } 1 \leq i \leq N\}$, $D_0(\boldsymbol{u}) \triangleq \{1,2,\ldots,N\} - D_1(\boldsymbol{u})$, $n(\boldsymbol{u}) \triangleq |D_1(\boldsymbol{u})|$, and $L(\boldsymbol{u}) \triangleq \sum_{i \in D_1(\boldsymbol{u})} |r_i|$. For MLD, the decoder finds the optimal codeword $\boldsymbol{c}_{\text{opt}}$, for which $L(\boldsymbol{c}_{\text{opt}}) = \min_{\boldsymbol{c} \in C} L(\boldsymbol{c})$. If there exists $\boldsymbol{c}^* \in C$ for which $L(\boldsymbol{c}^*) \leq \alpha(\boldsymbol{c}^*) \triangleq \min_{\boldsymbol{c} \in C, \boldsymbol{c} \neq \boldsymbol{c}^*} L(\boldsymbol{c})$ then $\boldsymbol{c}^* = \boldsymbol{c}_{\text{opt}}$. If it is possible to determine a tight lower bound on $\alpha(\boldsymbol{c}^*)$, we have a sufficient condition on the optimality of a candidate codeword.

For simplicity, assume that the bit positions are ordered with $|r_1| \leq |r_2| \leq \cdots \leq |r_N|$. For a subset X of $\{1,2,\ldots,N\}$ and a positive integer $j \leq |X|$, let $X^{(j)}$ denote the set of j smallest integers in X. For $j \geq 0$, $X^{(j)} \triangleq \phi$ (empty set) and for $j \geq |X|$, $X^{(j)} \triangleq X$. Let $d_H(\cdot,\cdot)$ denote the Hamming distance.

For $h > 0$, let B^h denote the set of binary sequences of length h. For $\alpha \in B^h$ and $1 \leq i \leq h$, let $pr_i\alpha$ denote the i-th bit of α. For $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_h$ and $\boldsymbol{u}$ in V_N, $D_\alpha \triangleq \bigcap_{i=1}^h D_{pr_i\alpha}(\boldsymbol{u}_i)$, $n_\alpha \triangleq |D_\alpha|$ and $q_\alpha \triangleq |D_1(\boldsymbol{u}) \cap D_\alpha|$. For $d_1, d_2, \ldots, d_h$ in $W - \{0\}$, $V_{N,d_1,d_2,\ldots,d_h} \triangleq \{\boldsymbol{u} \in V_N : d_H(\boldsymbol{u}, \boldsymbol{u}_i) \geq d_i \text{ for } 1 \leq i \leq h\}$. Then, $\boldsymbol{u} \in V_{N,d_1,d_2,\ldots,d_h}$ if and only if

$$\sum_{\alpha \in B^h} (-1)^{pr_i\alpha} q_\alpha \geq \delta_i \triangleq d_i - n(\boldsymbol{u}_i), \text{ for } 1 \leq i \leq h. \quad (1)$$

Let Q denote the set of those 2^h-tuples over nonnegative integers which satisfy (1). We say, $\boldsymbol{q} = (q_{00\cdots0}, q_{0\cdots01}, \ldots, q_{11\cdots1}) \in Q$ is minimal if and only if there is no $\boldsymbol{q}' = (q'_{00\cdots0}, q'_{0\cdots01}, \ldots, q'_{11\cdots1})$ such that $\boldsymbol{q} \neq \boldsymbol{q}'$ and $q_\alpha \geq q'_\alpha$ for any α in B^h. Let $Q_{\min}$ denote the set of minimal tuples in Q. Then

$$\min_{\boldsymbol{u} \in V_{N,d_1,d_2,\ldots,d_h}} L(\boldsymbol{u}) = \min_{\boldsymbol{q} \in Q_{\min}} \sum_{i \in \bigcup_{\alpha \in B^h} D_\alpha^{(q_\alpha)}} |r_i|. \quad (2)$$

Example: Let $h = 2$. It is proved in [4] that

$$\min_{\boldsymbol{u} \in V_{N,d_1,d_2}} L(\boldsymbol{u}) = \sum_{i \in (D_{00} \cup D_{01}^{\lfloor \lfloor (\delta_1 - \delta_2)/2 \rfloor \rfloor (\delta_1)})^{(\delta_1)}} |r_i|. \quad (3)$$

For $\boldsymbol{u}_1 = \boldsymbol{u}_2$, that is, $h = 1$ and $d_1 = d_2 =$ the minimum distance, equality (3) reduces to the formula given in Theorem 1 in [2].

For $\boldsymbol{u}_1 \neq \boldsymbol{u}_2$, the right-hand side of (3) is shown to be tighter than the lower bound $\sum_{i \in D_0(\boldsymbol{u}_2)^{(\lceil (\delta_1 + \delta_2)/2 \rceil)}} |r_i|$ in [3].

For $h = 3$, we derive a formula for $\min_{\boldsymbol{u} \in V_{N,d_1,d_2,d_3}} L(\boldsymbol{u})$ in [4].

III. Simulation Result

Simulation results show that $Cond_2$ is effective in all cases. $Cond_3$ is slightly more effective than $Cond_2$. The effectiveness of $Cond_2$ over $Cond_1$ is relatively small for $RM_{6,2}$ and $RM_{6,3}$ The details are shown in [4,6].

[1] This research is partially supported by NSF Grants NCR-9115400 and BCS-9115400, NASA Grant NAG 5-931 and the Ministry of Education, Japan, Grant No. (C) 06650416.

References

[1] D.Chase, "A New Class for Decoding Block Codes with Channel Measuremnt Information," *IEEE Trans. Inform. Theory*, Vol.IT-18, pp.170–182, Jan. 1972.

[2] D.J.Taipale and M.B.Pursley, "An Improvement to Generalized Minimum-Distance Decoding," *IEEE Trans. Inform. Theory*, Vol.IT-37, pp.167–172, Jan. 1991.

[3] T.Kaneko, T.Nishijima, H.Inazumi and S.Hirasawa, "An Efficient Maximum-Likelihood-Decoding Algolithm for Linear Block Codes with Algebraic Decoder," *IEEE Trans. Inform. Theory*, Vol.IT-40, pp.320–327, March 1994.

[4] T.Kasami, T.Takata, T.Koumoto, T.Fujiwara, H.Yamamoto and S.Lin, "The Least Stringest Sufficient Condition on Optimality of Suboptimal Decoded Codewords," *Technical Report of IEICE*, IT-94-82, IEICE, Japan, Jan. 1995.

[5] S.Lin, H.T.Moorthy and T.Kasami, "An Efficiente Soft-Decision Decoding Scheme for Binary Liner Block Codes," *Proc. 3rd Intern. Symp. Commun. Theory & Appl.*, 10–14 July 1995, Ambleside, UK.

[6] T.Kasami, T.Koumoto, T.Takata and S.Lin, "The Effectiveness of the Least Stringent Sufficient Condition on the Optimality of Decoded Codewords," ibid.

Implementation and Performance of a Serial MAP Decoder for use in an Iterative Turbo Decoder

Steven S. Pietrobon
Satellite Communications Research Centre, University of South Australia, The Levels SA 5095, Australia

Abstract —**A MAP decoding algorithm is described that can greatly speed up computer simulations of turbo coding schemes and which allows the practical implementation of turbo codes in their most powerful form.**

I. Introduction

The discovery of Turbo codes and the claim that they can perform within 0.7 dB of Shannon capacity for 1 bit/sym [1] has generated considerable interest within the coding community. The heart of an iterative decoding algorithm for Turbo codes described in [1] is the use of a Maximum a Posteriori (MAP) decoding algorithm derived from [2]. This MAP decoding algorithm is extremely complicated and greatly limits the decoding speed possible (since two MAP decoders are required in each iteration stage of the Turbo decoding algorithm, which may be up to 18 stages).

A great simplification of the MAP decoding algorithm is given in [3]. For a rate 1/n systematic convolutional code with memory v (and $M = 2^v$ states) this algorithm involves 4M additions, $6M + 2^n - 1$ multiplications, one division, and n exponentials (for an additive white Gaussian noise channel) per decoded bit. By taking the $-$ logarithm of the algorithm (the logarithm is used in [3]) we can convert the multiplications, divisions, and exponentials to additions and subtractions only (the exponentials conveniently disappear). However the addition operand becomes the E operand defined below:

$$x \; E \; y = -\ln(e^{-x} + e^{-y}). \tag{1}$$

We can simplify (1) to

$$x \; E \; y = \min(x, y) - \ln\left(1 + e^{-|y-x|}\right). \tag{2}$$

The E operand is then reduced to finding the minimum of x and y and a function dependent only on the difference between x and y.

We can see that the maximum of $f(z) = \ln(1+e^{-z})$, $z \geq 0$, is small, equal to $\ln(2) \simeq 0.693$. With increasing z, f(z) quickly decays to near zero for $z \geq 7$. In a computer simulation z can be quantised to some maximum value and a look up table used to find f(z). This greatly speeds up the MAP decoding algorithm with almost no degradation in performance. This technique can also be used in a hardware implementation of the MAP algorithm using very small look up tables.

II. A MAP Decoding Algorithm

The log likelihood ratio of a transmitted information bit at time k (d_k) can be shown to be [3] (the division in [3] should actually be a subtraction)

$$L(d_k) = \mathop{E}_{m=0}^{M-1} A_k^0(m) + B_k^0(m) - \mathop{E}_{m=0}^{M-1} A_k^1(m) + B_k^1(m), \tag{3}$$

where we define

$$\mathop{E}_{m=0}^{M-1} g(m) = g(0) \; E \; g(1) \; E \; \cdots \; E \; g(M-1). \tag{4}$$

$A_k^i(m)$ and $B_k^i(m)$ can be computed iteratively as

$$A_k^i(m) = D_i(R_k, m) + \mathop{E}_{j=0}^{1} A_{k-1}^j(S_b^i(m)), \tag{5}$$

$$B_k^i(m) = \mathop{E}_{j=0}^{1} B_{k+1}^j(S_f^i(m)) + D_j(R_{k+1}, S_f^i(m)), \tag{6}$$

where $S_f^i(m)$ and $S_b^i(m)$ is the state you go to from state m along the path $d_k = i$ forwards and backwards in time, respectively. R_k is the received length n vector at time k. The branch metric $D_i(R_k,m)$ is defined for a rate 1/2 code as

$$D_i(R_k, m) = -\frac{2}{\sigma^2}(x_k i + y_k Y^i(m)), \tag{7}$$

where $R_k = (x_k, y_k)$, $x_k = (2d_k - 1) + p_k$, $y_k = (2Y_k - 1) + q_k$, p_k and q_k are two independent normally distributed random variables with variance σ^2, Y_k is the coded bit at time k, and $Y^i(m)$ is the coded bit for state m and $d_k = i$.

For a length N coded sequence (starting and finishing in state 0) the algorithm follows these steps:

1) Starting at time k = 0, compute $D_i(R_k,m)$ using (7) for all received symbols and store in an array of size $2^n N$.
2) Initialise $B_{N-1}^i(S_b^i(0)) = 0$ for i = 0,1 and $B_{N-1}^i(m) = \infty$ for all other m and i. Starting at time k = N–2, iteratively compute $B_k^i(m)$ using (6) and store in an array of size MN (since $B_k^1(m) = B_k^0(m')$ where $S_f^1(m) = S_f^0(m')$ we can reduce the array size by half).
3) Initialise $A_0^i(0) = D_i(R_0, 0)$ for i = 0,1 and $A_0^i(m) = \infty$ for all $m \neq 0$ and i = 0,1. Starting at time k = 1, iteratively compute $A_k^i(m)$ using (5). For each k compute $L(d_k)$ using (3).

The "state metrics" $A_k^i(m)$ and $B_k^i(m)$ need to be renormalised after each iteration to prevent the metrics from overflowing. This is achieved by subtracting the smallest state metric at each k (previously this would have been done by division).

A serial MAP decoder is being designed which uses a modified form of the above algorithm. The received samples are quantised into six bits with eight bit state metrics. We have $N = 2^{16}/M$ where M is programmable from 4 to 512 states. The decoder is able to decode any systematic code with rates from 1/2 to 1/4. Limiting is used to prevent overflow and small 64x4 lookup tables are used to implement f(z). The maximum bit rate is $10^7/(M+14)$ bit/s (19 to 556 kbit/s for M = 4 to 512 states, respectively).

Four Xilinx XC3100A programmable gate arrays are being used, together with several 64Kx4 static RAMs and 1Kx8 dual port static RAMs. With an additional XC3100A chip, two MAP decoders with depth 64K random interleaving can be implemented on a single board as one stage of a turbo decoder (the inner and outer code must be the same). Each board can have its data fed back or passed onto another board depending on the required speed/complexity requirements.

This work was supported by the Australian Research Council under an Australian Postdoctoral Research Fellowship and University of South Australia Research Development Grants 85413 78 and 85460 78.

References

[1] Berrou, C., Glavieux, A., and Thitimajshima, P., "Near Shannon limit error–correcting coding and decoding: Turbo–Codes," *ICC'93*, Geneva, Switzerland, pp. 1064–1070, May 1993.

[2] Bahl, L., Cocke, J., Jelinek, F., and Raviv, J., "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT–20, pp. 284–287, Mar. 1974.

[3] S. S. Pietrobon and A. S. Barbulescu, "A simplification of the modified Bahl decoding algorithm for systematic convolutional codes," *Int. Symp. Inform. Theory & its Applic.*, Sydney, Australia, pp. 1073–1077, Nov. 1994.

On the convergence of the iterated decoding algorithm

Giuseppe Caire, Giorgio Taricco, and Ezio Biglieri[1]
Dipartimento di Elettronica • Politecnico • Corso Duca degli Abruzzi 24 • I-10129 Torino (Italy)
fax: +39 11 5644099 • e-mail: <name>@polito.it •

Abstract — **Recently, the concept of *iterated decoding* of concatenated codes has been developed . Successful applications of this concept include *turbo-codes* and soft-decoding of product codes. After stating the iterated decoding algorithm formally, we provide a conjecture on the convergence and the asymptotic optimality of this algorithm.**

I. Introduction

Consider the following decoding strategy for block codes over a memoryless channel [1, 4]. The receiver observes the "unconstrained a posteriori" (UAP) distribution Q, and feeds it to the decoder to obtain a new distribution P which satisfies the code constraints and minimizes the *directed divergence* (or *cross-entropy*) $D(P \parallel Q)$ between Q and P. If I_C denotes the set of constraints introduced by code C, we write

$$P = Q \circ I_C \tag{1}$$

to describe the operation of the decoder.

Decoding by cross-entropy minimization consists of computing P given Q (i.e., given the observed channel output and the knowledge of the channel transition distribution), and then selecting the code word x which maximizes P. A "soft-output ML decoder" can be thought of as a device performing operation (1) [1]. Standard variational-calculus techniques provide the solution

$$Q \circ I_C = Q(x) I_C(x) \left[\sum_{x \in C} Q(x) \right]^{-1}. \tag{2}$$

II. Iterated decoding

Consider binary block codes which can be described as the intersection of two supercodes, i.e. $C = C_1 \cap C_2$ (two-fold product codes and the turbo-codes can be expressed in this way.) The iterated decoding schemes proposed for these codes [2, 3] can be formally described as follows.

Given a distribution P, let $\tilde{P}$ denote the distribution obtained as the product of the marginals of P. Let Q denote as usual the UAP distribution. Then let $\tilde{P}^{2,0} = Q$, and for $\ell = 1, 2, 3, \ldots$, iterated decoding consists of the sequence of minimization problems

$$P^{1,\ell} = \tilde{P}^{2,\ell-1} \circ I_{C_1} \qquad P^{2,\ell} = \tilde{P}^{1,\ell} \circ I_{C_2}. \tag{3}$$

III. Some conjectures

We conjecture that *iterated decoding is equivalent to typical-set decoding*. It is known that typical set decoding is asymptotically optimal, in the sense that it achieves channel capacity. Given the received word y, the typical-set decoder outputs code word x_0 if this is the unique code word $x \in C$ such that (x, y) are jointly typical, otherwise it outputs an error message.

Assume that we receive a typical y (if the received sequence is not typical we are done, since in any case we have an error). The number of x code words jointly typical with y is, on the average, $2^{nH(\mathcal{X}|\mathcal{Y})}$, where $H(\mathcal{X}|\mathcal{Y})$ denotes the conditional entropy rate of x given y. These are the words "of high probability" when y is given, and, by the asymptotic equipartition property, they have roughly the same probability $\simeq 2^{-H(Q)}$, where $H(Q)$ denotes the entropy of the UAP distribution Q. Let $A(y)$ denote the set of those sequences, and $A^{i,\ell}(y)$ the set of the sequences x jointly typical with y under the new conditional distribution $\tilde{P}^{i,\ell}$ obtained at the ℓ-th step of the iteration ($i = 1$ or 2). The three parts of the conjecture are as follows:

1. If $C_1 \cap A(y) \neq \emptyset$, then $|A^{1,1}(y)| \leq |A(y)|$. This is equivalent to saying that, if some sequences $x \in C_1$ are typical under distribution Q, then $H(Q) \geq H(\tilde{P}^{1,1})$.

2. Suppose that both $A(y) \cap C_1$ and $A(y) \cap C_2$ are not empty. Then

$$\begin{aligned} |A^{1,1}(y) \cap C_1| &\simeq |A(y) \cap C_1| \\ |A^{1,1}(y) \cap C_2| &\leq |A(y) \cap C_2| \end{aligned}$$

In other words, the typical sequences in C_1 under Q are still typical under $\tilde{P}^{1,1}$. Hence, since the size of the set must decrease, we must delete some sequences from $A(y)$ which are not in C_1. In particular, we can throw away some sequences of C_2 (which are not also in C_1).

3. If C is a good code, there will be at most one sequence $x_0 \in C$ jointly typical with y (otherwise we would have an error in any case). If $A(y) \cap C = \{x_0\}$, then $A^{1,\infty}(y) = A^{2,\infty}(y) = \{x_0\}$ and the iterations converge to the distribution

$$\tilde{P}^{1,\infty} = \tilde{P}^{2,\infty} = I_{x_0}(x).$$

[1]This research was sponsored by the Italian National Research Council (CNR) under "Progetto Finalizzato Trasporti."

References

[1] G. Battail, "Le décodage pondéré en tant que procédé de réévaluation d'une distribution de probabilité," *Annales des Télécommunications*,Vol. 42, No. 9–10, pp. 499–509, Sept.–Oct. 1987.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," *ICC'93*, Geneva, Switzerland, May 1993.

[3] J. Lodge, R. Young, P. Hoeher, and J. Hagenauer, "Separable MAP 'filters' for the decoding of product and concatenated codes," *ICC'93*, Geneva, Switzerland, pp. 1740–1745, May 1993.

[4] M. Moher, "Decoding via cross-entropy minimization," *Proc. IEEE Globecom'93*, Houston, TX, pp. 809–813, Nov. 29–Dec. 2, 1993.

An Iterative Decoding Scheme for Serially Concatenated Convolutional Codes

Mohamed Siala, Eckhard Papproth[1], Kaïs Haj Taieb, and Ghassan Kawas Kaleh
Ecole Nationale Supérieure des Télécommunications,
46 rue Barrault, 75634 Paris 13, France

Abstract — **We present an iterative soft-output decoding algorithm for serially concatenated convolutional codes which has better performance than the conventional noniterative decoding algorithm. The proposed decoding scheme can be used whenever some form of serial concatenation of encoders and channels with memory is applied.**

The figure shows the block diagram for a serially concatenated coding system with iterative soft-output Viterbi decoding. The binary data sequence $\{a_i\}$ is fed into the outer encoder. The binary sequence $\{b_j\}$ at the output of this encoder is interleaved to result in $\{c_k\}$. This sequence is then *serially* encoded by the inner encoder into the sequence $\{d_l\}$. The sequence $\{d_l\}$ is sent over a Gaussian channel and produces at its output the noisy sequence $\{y_l\}$, where $y_l = d_l + n_l$. The $\{n_l\}$ denotes an additive white Gaussian noise sequence with zero mean and variance σ^2.

In the first stage of the m-th iteration of the decoding algorithm, soft information, $\hat{\Lambda}_k^{(m)}$, $k = \ldots, -1, 0, 1, \ldots$, associated with the estimated symbols $\hat{c}_k^{(m)}$, $k = \ldots, -1, 0, 1, \ldots$, are computed by the simplified version of the SOVA [1] taking into account the intrinsic contributions of the outer decoder from the previous iterations. This algorithm is referred to as the *inner* SOVA. Let $\mathcal{T}_i$ denote the trellis which represents the structure of the inner encoder. The metric adopted by this stage is the Euclidean metric

$$\sum_l (y_l - \xi_l)^2 - \sum_k (2\chi_k - 1) \sum_n (\epsilon^{(n)} \Delta_k^{(n)}),$$

where χ_k and ξ_l are respectively the inputs and noiseless outputs in an arbitrary path with the same position as y_l in $\mathcal{T}_i$ and n goes over all previous iterations. For a given SNR, the positive parameters $\epsilon^{(n)}$ are arbitrary and should be chosen to minimize the bit-error-probability of the sequence $\{\hat{a}_i\}$ at the end of the iterative decoding process. $\Delta_k^{(n)}$ are the intrinsic contributions of the outer SOVA as defined below. The soft-information variables $\hat{\Lambda}_k^{(m)}$ represent, up to a multiplicative factor, an approximation of the log-likelihood ratios

$$\left| \ln \frac{P(c_k = \text{“0”} | \{y_l\})}{P(c_k = \text{“1”} | \{y_l\})} \right|.$$

The sequence of estimated symbols and the associated reliability information are deinterleaved using the reverse procedure of the block interleaver. The resulting sequences are denoted by $\{\hat{b}_j^{(m)}\}$ and $\{\hat{\Gamma}_j^{(m)}\}$, respectively.

The outer decoder uses also a simplified SOVA. It applies the structure of the outer encoder trellis to the sequence delivered by the inner decoder. Therefore, it provides enhanced soft outputs for the sequence received from the inner decoder $\tilde{\Gamma}_j^{(m)}$ associated with the new estimated symbols $\tilde{b}_j^{(m)}$. This algorithm is referred to as the *outer* SOVA. The metric adopted by this stage is the simple correlation metric

$$-\sum_j (2\beta_j - 1)(2\hat{b}_j^{(m)} - 1)\tilde{\Gamma}_j^{(m)},$$

where $\beta_j \in \{\text{“0”}, \text{“1”}\}$ is an arbitrary path symbol with the same position as $\hat{b}_j^{(m)}$ in the outer encoder trellis $\mathcal{T}_o$. Once again, the enhanced soft-information variables $\tilde{\Gamma}_j^{(m)}$ represent, up to a multiplicative factor, an approximation of the log-likelihood ratios

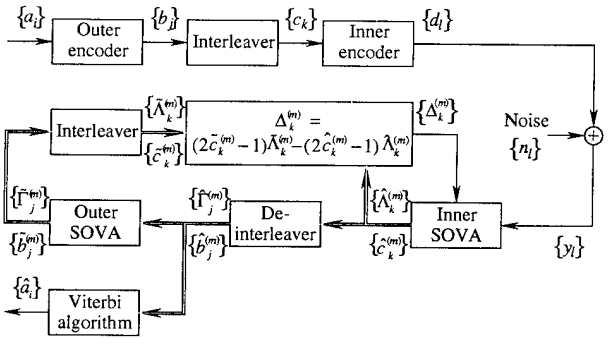
$$\left| \ln \frac{P(b_j = \text{“0”} | \{\hat{b}_{j'}^{(m)}\}, \{\tilde{\Gamma}_{j'}^{(m)}\})}{P(b_j = \text{“1”} | \{\hat{b}_{j'}^{(m)}\}, \{\tilde{\Gamma}_{j'}^{(m)}\})} \right|.$$

This second stage exploits the fact that the sequence $\{\hat{b}_j^{(m)}\}$, when correct, must be a codeword sequence of the outer encoder. Denote respectively by $\{\tilde{c}_k^{(m)}\}$ and $\{\tilde{\Lambda}_k^{(m)}\}$ the interleaved versions of the sequences $\{\tilde{b}_j^{(m)}\}$ and $\{\tilde{\Gamma}_j^{(m)}\}$. The intrinsic contribution of the outer SOVA in comparison with the inner one is measured by the difference

$$\Delta_k^{(m)} = (2\tilde{c}_k^{(m)} - 1)\tilde{\Lambda}_k^{(m)} - (2\hat{c}_k^{(m)} - 1)\hat{\Lambda}_k^{(m)}.$$

For the following iteration, the new soft-information variables generated by the inner SOVA and its associated decisions are denoted by $\hat{\Lambda}_k^{(m+1)}$ and $\hat{c}_k^{(m+1)}$, respectively.

At the final f-th iteration of the iterative decoding process, an outer conventional Viterbi decoder delivers the estimated sequence, $\{\hat{a}_i\}$, of the information sequence. The metric used by this decoder is $-\sum_j (2\beta_j - 1)(2\hat{b}_j^{(f)} - 1)\tilde{\Gamma}_j^{(f)}$.



References

[1] J. Hagenauer and P. Hoeher, "A Viterbi Algorithm with Soft-Decision Outputs and Its Applications," *GLOBECOM'89*, Dallas, Texas, pp. 47.1.1–47.1.7, November 1989.

[1] Author to whom correspondence may be addressed. E. Papproth was supported by a DAAD-fellowship HSP II financed by the German Federal Ministry for Research and Technology (BMFT).

Soft-Decision Decoding of Binary Linear Block Codes Based on An Iterative Search Algorithm

Hari T. Moorthy, Shu Lin and Tadao Kasami[1]
Dept. Electrical Eng., 2540 Dole Street, 483,
University of Hawaii, Honolulu, Hawaii, USA

Abstract — **This paper presents a suboptimum soft-decision decoding scheme for binary linear block codes based on an iterative search algorithm using a purged trellis diagram. The scheme achieves near optimum error performance with a significant reduction in computation complexity.**

I. Summary

The proposed scheme uses a hard-decision decoder to produce a candidate codeword and exploits the fact that the hard-decision decoded codeword is either the optimum maximum likelihood decoding (MLD) solution or at a distance not too far away from the optimum MLD solution. As a result, the optimum MLD solution may be found by searching through those codewords that are close to the candidate codeword. The search is conducted through a **purged trellis diagram** for the given code. If the optimum MLD solution is not found, a new candidate codeword is generated by using a new test error pattern to modify the hard-decision received sequence. Then **optimality test** is repeated and a new search begins. Generation of new candidate codewords and searches repeat until either the optimum MLD solution is found or the decoding process is terminated by exhausting all possible test error patterns.

Sufficient conditions for optimality are proved. Upper bounds on the Hamming distance between a hard-decision decoded candidate codeword and the optimum MLD solution are derived [1]. These upper bounds define a **search region** for the optimum MLD solution. The proposed decoding scheme is simulated for some well known codes. The simulation results show that the proposed decoding scheme achieves either practically optimum performance or a performance only a fraction of a dB away from MLD with a significant reduction in decoding complexity compared with the Viterbi decoding based on the full trellis diagrams of the codes [3].

II. Examples

Consider the (23,12,7) Golay code. The proposed decoding algorithm achieves practically optimum performance. At SNR = 4 dB, the proposed decoding algorithm achieves the bit-error-rate (BER) of 10^{-3} and requires less than 100 binary operations (including additions and comparisons for the purged trellis search). The average number of iterations required to decode a received sequence at SNR = 4 dB is 0.9. At SNR = 6 dB, the proposed decoding algorithm achieves BER of 10^{-6} with average number of binary operations less than 15 and the average number of iterations required to decode a received sequence is 0.2. However, the optimal Viterbi decoding algorithm based on the full trellis diagram of the code requires a fixed number of 2,559 binary operations. The most efficient optimum decoding algorithm for the (24,12,8) extended Golay code proposed so far requires at least 550 but no more than 651 binary operations [2]. For SNR greater than 3 dB, the proposed decoding algorithm requires much less computations than that of the optimum decoding proposed in [2].

Next, we consider the $(32, 21, 6)$ extended primitive BCH code. The proposed decoding algorithm again achieves practically optimum error performance. It achieves the BER of 10^{-5} at SNR = 5.4 dB. At this SNR, the average number of binary operations and the average number of iterations required to decode a received codeword are 25 and 0.5 respectively, whereas the optimum Viterbi decoding based on the full trellis diagram of the code would require 30,156 binary operations. Even for the worst case, the proposed decoding algorithm requires a maximum of no more than 6,350 binary operations. We see that for the (32,21,6) extended BCH code, the proposed decoding algorithm achieves practically optimum performance with a significant reduction in computation complexity.

Using the proposed algorithm to decode the (64,45,8) extended BCH code, there is a 0.5 dB loss in coding gain at the BER 10^{-5} compared with the optimum MLD. The SNR required to achieve BER 10^{-5} is 5.3 dB. At this SNR, the average number of binary operations and the average number of iterations required to decode a received word are 300 and 0.7 respectively. However, optimum Viterbi decoding based on the full trellis diagram of the code requires 4,301,823 binary operations (this number can be reduced by certain permutations of the order of bits in the trellis). Even for the worst case, the proposed decoding algorithm requires only a maximum of 57,182 binary operations. We see that a tremendous reduction in computation complexity is achieved with only a small performance degradation.

References

[1] Hari T. Moorthy, Shu Lin and Tadao Kasami, "Soft-Decision Decoding of Binary Linear Block Codes Based on an Iterative Search Algorithm," *submitted to IEEE Transactions on Information Theory*, Jan 1995.

[2] A. Vardy and Y. Be'ery, "More Efficient Soft Decoding of the Golay Codes," *IEEE Transactions on Information Theory*, Vol. 37, pp. 667-672, May 1991.

[3] S. Lin and D.J. Costello, "Error Control Coding: fundamentals and applications," *Prentice-Hall*, 1983.

[1]This research was supported by NSF Grant NCR-9115400 and NASA Grant NAG 5-931.

A Soft Output Decoding Algorithm for Concatenated Systems

Xiao-an Wang and Stephen B. Wicker
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

Concatenated codes are a powerful means for improving the performance of digital communication systems over extremely noisy channels. Such a code contains an inner code, which is often a convolutional code, and an outer code. Since the conventional inner decoder only gives hard outputs (i.e., 0 or 1 for binary codes), the outer decoder is forced to work in the HDD (hard decision decoding) fashion. Reliability information is needed to fully utilize the SDD capacity of the outer code. Even in cases where no practical SDD algorithms exist for the outer codes (e.g., Reed-Solomon codes), reliability information can help to erase highly unreliable bits, and the performance can be improved through errors and erasures decoding [1]. A decoder capable of delivering such reliability information is called a soft output decoder.

The reliability measure for a decoded symbol is the probability P_c that the symbol is correct or the probability of error $P_e = 1 - P_c$. Such quantities can be obtained by the symbol-by-symbol MAP (maximum *a posteriori* probability) algorithm. Unfortunately this algorithm is computationally inefficient. A soft output Viterbi algorithm (SOVA) [2] can provide an estimate of P_e which is accurate only for large SNR.

This paper proposes an efficient modified MAP algorithm for obtaining P_c for the outputs of convolutional inner decoders. The outer decoder uses P_c to perform SDD by choosing a codeword $\mathbf{y} = (y_0, y_1, \ldots, y_{l-1})$ which maximizes the maximum likelihood (ML) metric $\mu_p(\mathbf{y}) = \sum_{i=0}^{l-1} \ln \pi_i(y_i)$, where $\pi_i(y_i)$ is the probability P_c that symbol y_i is correct. Decoding based on this ML metric is referred to as *Generalized* SDD since it includes the Euclidean metric on AWGN channels and the one proposed in [2] for binary memoryless channels as special cases. The following theoretical and implementation issues are also investigated:

Convergence. Practical decoders have to have finite decoding delay (decoding depth) Γ which causes truncation errors for very long or infinite data streams. By a matrix formulation of the MAP algorithm, P_c can be seen to be a function of the products of infinite random matrices. The theory of products of random matrices (PRM) is then used to show that the estimated P_c over a finite Γ, $\hat{P}_c(\Gamma)$, converges to P_c exponentially fast with Γ. Thus the truncation error can be made arbitrarily small.

Complexity. The VA is the most efficient hard output convolutional decoder. A soft output decoder is expected to have more complexity because it extracts more information from the inputs. It is informative and of practical significance to use the Viterbi decoder as a complexity measure against other decoders. It is demonstrated that the MAP soft output decoder has a complexity of $\Gamma + 1$ times that of the Viterbi decoder.

Decoding delay. The Γ required for a fixed level of accuracy changes bit by bit, as well as with channel conditions such as SNR. A fixed delay would have to be very long to accommodate the worst case, which increases the complexity and is unnecessary most of the time. Using the PRM theory, the problem of obtaining P_c can be reformulated as a problem of best fit between two vectors. Solutions of the best fit problem provide a scheme which keeps Γ at a minimum for required precision. The scheme makes the modified MAP algorithm very efficient. Figure 1 shows that the average delay $\bar{\Gamma}$ versus SNR for a rate-1/2 code with constraint length 3. $\bar{\Gamma}$ is kept below 5 over the entire operating region of the code and rapidly drops to zero. The algorithm is as efficient as the VA for SNR as low as 3 dB.

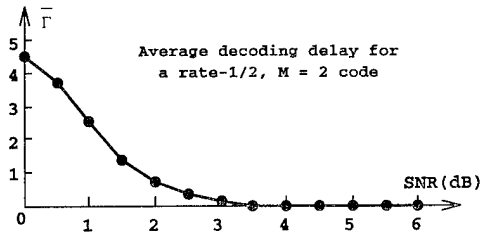


Fig. 1: Average decoding delay

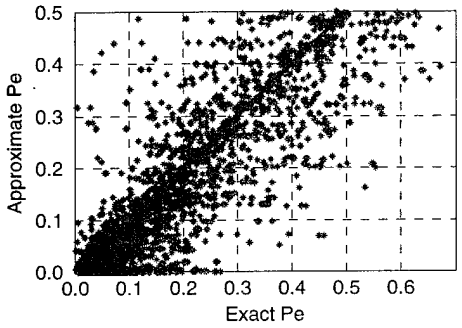


Fig. 2: Comparison of P_e and $\hat{P}_e$

Range overflow. This phenomena was shown to occur very easily during the decoding process. To solve this problem, it is shown that the relative amplitudes among the quantities in consideration are bounded, thus a very simple and effective scaling scheme is constructed.

Finally, a comparison is made between the exact P_e provided by the modified MAP algorithm and the approximation $\hat{P}_e$ in [1, 2]. It is shown that the approximation gives an optimistic estimate of P_e, especially for low SNR. The result is plotted in Figure 2 for 50,000 consecutive bits at SNR = 2 dB, using the same code mentioned above. The discrepancy becomes smaller for increasing SNR.

References

[1] J. Hagenauer, E. Offer and L. Papke, "Matching Viterbi decoders and Reed-Solomon decoders in a Concatenated system", in *Reed-Solomon Codes and Their Applications*, Edited by S. B. Wicker and V. K. Bhargava, IEEE press, 1994.

[2] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications", *Proceedings of IEEE Globecom Conference*, pp 47.1.1–47.1.7, Dallas, TX, USA, November 1989.

On the Impact of Laser's Relaxation Oscillation on Quadratically Detected Heterodyned Lightwave Signals

Yeheskel E. Dallal[1], Gunnar Jacobsen[2] and Shlomo Shamai (Shitz)[1]

[1]Dept. of Elect. Eng., Technion City, Haifa 32000, Israel

[2]Tele Denmark Research (TDR), Lyngsø Allé, DK 2970, Hørsholm, Denmark

We consider the performance of quadratically detected heterodyned lightwave signals in the presence of Laser's relaxation oscillations. Here the limiting form of the channel presents both additive white Gaussian noise (AWGN) of spectral density N_o and Laser's phase noise. The widely accepted model for the Laser's phase noise is a Brownian Motion giving rise to a Lorentzian line spectrum [1]. However, the actual line shape of semi-conductor lasers deviates from this simplified and idealized statistical characterization [2]. The analytical techniques provided here show that this deviation may have a signficiant impact on the communication-system performance.

The major difference stems from the Laser's relaxation oscillations, which induce periodic satellite peaks in the line spectrum. The resulting phase noise is characterized by a normalized zero-mean Gaussian process with autocovariance function

$$\mathbb{E}\phi_t\phi_s = \min(t,s) + \frac{1}{2}\,\mathrm{Real}\left\{Be^{-\xi_R|t-s|}\,e^{j\nu_R|t-s|}\right\}. \quad (1)$$

The term $\min(t,s)$ gives rise to a perfect Lorentzian spectrum. The second term presents a deviation from the Brownian Motion phase model. Here $\xi_R = (\pi B_l \tau_R)^{-1}$ where B_l is the underlying "Brownian" linewidth, τ_R is the decaying time-constant and $\nu_R = \Omega_R/\pi B_l$ where Ω_R is the resonance angular frequency. B is a complex constant depending on the laser's parameters.

The receiver in focus here comprises L-fold square-law detection of L filtered noisy phase frequency shift keying (FSK) signals observed in AWGN. The underlying decision statistics relies on Laser's phase noise via normalized exponential functionals of the form:

$$\varepsilon_t = \left|\int_o^t e^{j\sqrt{2}\phi_s}\,ds\right|^2, \quad \Gamma_t = \left|\int_o^t e^{j\sqrt{2}\phi_s}\,e^{j\delta\cdot s}\,ds\right|^2, \quad (2)$$

where ε_t crrepsonds to the inband received signal and Γ_t renders a crosstalk signal. Here δ is the normalized frequency spacing between the FSK signals.

The joint statistics inherited by the phase noise functionals (2), is unknown, even for the simplified Brownian Motion Model [1]. Assuming L-fold diversity reception (L statistically independent and identically distributed observed noisy phase signals) which can be achieved via interleaving techniques, power moment statistical characterization of ε_t and Γ_t is a useful approach. Indeed in the case of a Brownian Motion, tight Bit Error Rate (BER) bounds are achievable with the use of a few moments, for optimized system parameters [3].

Following [3], the application of the theory of limiting values of integrals, the Hölder-Inequality and the Chernoff bound yield upper bounds on the bit error probability (BER). The bounds are given in terms of the power moments induced by the phase noise functionals ε_t and Γ_t, the computation of which is required. Noting that the joint power moments featured by the Markovian Brownian Motion functionals are analytically tractable, [3] the considered power moments of ε_t and Γ_t are related to certain mutual moments governed by a Brownian Motion. For illustration, the first-order moment of ε_t at $t=\beta$ is given by

$$E\varepsilon_\beta = e^{-\mathrm{Real}\{B\}}\sum_{r=-\infty}^{\infty}\lambda_r\mathcal{F}_\beta\Big((1,r)(-1,r)\Big),\ 0\le\beta\le\overline{\beta} \quad (3)$$

where $\mathcal{F}_\beta(\cdot)$ is obtained via the inverse Laplace transform of $\overline{\mathbb{F}}_S(\cdot)$ stated below at $t=\beta$,

$$\overline{\mathbb{F}}_S\Big((I_1,W_1)(I_2,W_2)\Big) = 2\Big(S+1+j\tfrac{1}{2}I_1W_1+j\tfrac{1}{2}I_2W_2\Big)\Big(S+(I_1+I_2)^2\Big)^{-1}\Big(S+jI_1W_1+jI_2W_2\Big)^{-1}\Big(S+1+jI_2W_2\Big)^{-1}\Big(S+1+jI_1W_1\Big)^{-1}. \quad (4)$$

The $\{\lambda_r\}$ are Fourier-Series coefficients, computed on the interval $[-\overline{\beta},\beta]$, which are strictly determined by the relaxation oscillation parameters ξ_R and ν_R. Similar expressions are obtained in general for higher order moments.

The resulting upper bounds on the BER are determined by the various system parameters: laser linewidth-to-bit rate ratio B_l/R, the bit-energy-to-noise density ratio E_b/N_o, the diversity level L (or equivalently the IF bandwidth expansion relative to the bit rate), Laser's relaxation oscillation normalized decaying time $\frac{1}{\xi_R}$, and Laser's relaxation oscillation normalized resonance frequency ν_R. Orthogonal reception in the absence of phase noise is assumed namely $\frac{\Delta\Omega}{LR} = 2\pi\cdot\tilde{\delta}$, where $\Delta\Omega$ is the frequency spacing, R is the bit rate, and $\tilde{\delta}$ is a positive integer. The impact of Laser's relaxation oscillation on the obtained upper bounds is studied and it is shown that Laser's relaxation oscillation may result in a significant penalty relative to the simplified Brownian Motion model. For example, assume BER $= 10^{-9}$, $B_l/R = 1$, $\tilde{\delta} = 4$, $L = 25$, $\nu_R = 53.9$, $\xi_R = 2.3$. Then a relative penalty of nearly 9 dB is predicted. Increasing the IF bandwidth to $L = 30$ would result in a decreased penalty of 3.5 dB.

References

[1] G. J. Foschini and G. Vannuci, "Characterizing Filtered Lightwaves Corrupted by Phase Noise", *IEEE Trans. Inform. Theory*, Vol. 34, pp. 1437–1448, Nov. 1988.

[2] C. H. Henry, "Theory of the Phase Noise and Power Spectrum of a Single Mode Injection Laser", *IEEE J. of Quantum Elec.*, Vol. 19, No. 9, pp. 1391–1397, Sep. 1983.

[3] Y. E. Dallal and S. Shamai (Shitz), "An Upper Bound on the Error Probability of Quadratic Detection in Noisy Phase Channels", *IEEE Trans. Commun.*, Vol. 39, No. 11, pp. 1635–1650, Nov. 1991, and Vol. 40, No. 11, pp. 1781–1784, Nov. 1992.

On a new detection scheme of optical PPM signal.

Kouichi Yamazaki

Dept. of Inform. and Commun. Eng., Tamagawa Univ.
6-1-1, Tamagawa-gakuen, Machida, Tokyo, 194, Japan

Abstract — **A new detection scheme is proposed for optical pulse position modulation (PPM) communication system. Channel property of the proposed scheme is clarified theoretically.**

I. Introduction

This paper proposes a new detection scheme for detecting $M-ary$ optical PPM signal. It is shown that the proposed scheme performs almost optimum on error probability criterion. Channel capacity of the proposed scheme is compared with other detection schemes.

II. New detection scheme

The block diagram of the proposed receiver is shown in Figure 1. The receiver consists of a local laser, a highly transmissive beam splitter, a photodiode and a feedback control system of the local laser. Frequency of the local field is identical to the signal field, its phase is π-shifted with respect to the signal and its amplitude is set so that its reflected part is the same as the transmitted part of the signal. Then, if the local laser is *on*, it cancels out the signal field perfectly by the combination process through the beam splitter. At the beginning of each symbol, a local laser is *on*. A photon number of combined field is counted for each time-slot individually. If no photon is counted during a certain, say "ith", time-slot, the feedback control system switches the local laser off from the next time-slot. If no other photons are counted after that till the end of the symbol, a symbol having an optical pulse at the ith time-slot, m_i, is decided as the transmitted symbol. On the other hand, if some other photons are counted in the jth time-slot ($j > i$), a symbol m_j is selected. In the above operation, when a symbol m_j is transmitted, an error occurs if no photon is counted in a certain, say "ith" ($i < j$), time-slot, and if no photon is counted in the jth time-slot. The probability of this error depends on a symbol as follows:

$$Pe(m_j) = e^{-Ns}\{1-(1-e^{-Ns})^{j-1}\} \qquad (1)$$

Averaging these symbol-dependent error probabilities with respect to *a priori*-probabilities, we obtain average error probability. For equally probable signal ($=1/M$), an average error probability is given as follows:

$$Pe_{ave.}^{prop.} = \frac{M-1}{M}e^{-Ns} - \frac{1-e^{-Ns}}{M}\{1-(1-e^{-Ns})^{M-1}\} \qquad (2)$$

III. Numerical results

Error performance and channel capacity of the proposed scheme are shown as a function of signal energy N_s for symbol length M of 64 in Figures 2 and 3, respectively. Those of optimum-quantum receiver[1] and direct detection receiver are also shown. It is found in Figure 2 that the proposed scheme is superior to direct detection scheme on error performance, and it performs almost optimally. Fig.3 also shows superiority of the proposed scheme, especially for N_s over 6dB. It seems from these results that we can expect the proposed detection scheme to perform ultimately low-energy communication.

Acknowledgements

The author would greatly appreciate Prof. O. Hirota, Dr. M. Osaki of Tamagawa university and Dr.T.S.Usuda of Nagoya Inst. Tech. for their variable discussions and comments.

References

[1] O. Hirota, *Optical Communication Theory*, (Morikita Pub. Company: Tokyo, 1985).(*in Japanese*)

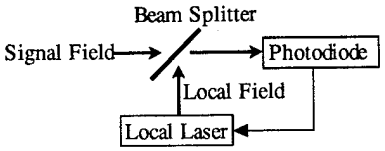


Figure 1: Block diagram of proposed detection scheme.

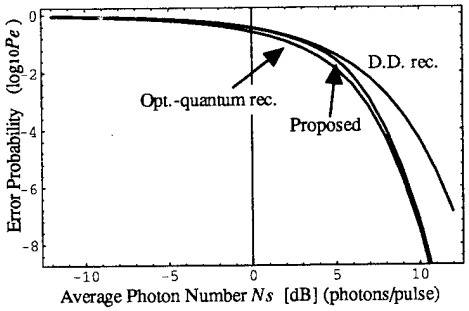


Figure 2: Error performance for symbol length $M = 64$.

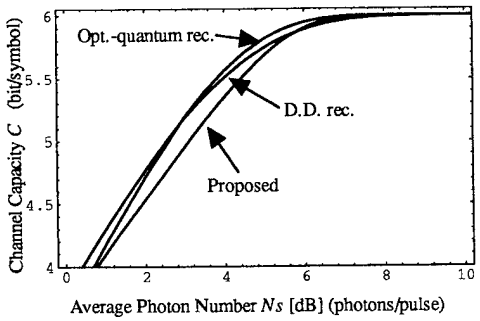


Figure 3: Channel capacity for symbol length $M = 64$.

Tight Lower Bounds on Capacity and Cutoff Rate of DOPPM in Optical Direct-Detection Channel

Tomoaki Ohtsuki†, Iwao Sasase‡, Shinsaku Mori‡

†Dept. of Electrical Engineering, Science University of Tokyo
2641 Yamazaki, Noda, Chiba, 278 Japan

‡Dept. of Electrical Engineering, Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama, 223 Japan

Differential overlapping pulse position modulation (DO PPM) can achieve higher capacity and cutoff rate than PPM, DPPM and OPPM when the pulsewidth and the average power of the channel are constrained [1]. In [1] erasure events of one pulsed chip that can be decoded correctly is defined as an erasure event. This results in loose lower bounds on the performance.

This summary analyzes the tighter lower bounds on capacity and cutoff rate of DOPPM in optical direct-detection channel. Considering what pulsed chips we have to detect to decode the words correctly, we derive the transition probability of DOPPM words and the tighter lower bounds in optical direct-detection channel.

We analyze the performance of DOPPM under the window scheme [1]. In a given window of length L (chips), we attempt to send W_d symbols of DOPPM with N chips consisting of Q-ary PPM: $L < W_d QN$. We specify the window scheme as follows: only if we detect photons at the both ends of each pulse for all the pulses, we can detect any sequence fitting in the window correctly. In particular for the pulses continuously generated from the left or right end of the sequence, we can decide the positions of the pulses only if we detect photons at the left or right end chip of each pulse, respectively. Unless we detect photons at the chips needed for correct decoding, then we consider the entire sequence to be garbled and define this sequence as an erasure sequence.

We denote the probability of using any one of the M symmetric inputs by $P(x_i) = \alpha$ and that of not sending a sequence by $P(x') = \beta$: $M\alpha + \beta = 1$. The mutual information can be derived as

$$I(x;y) = \sum_{P_{Lx_i}=0}^{W_d} \sum_{P_{Rx_i}=0}^{W_d - P_{Lx_i}} \alpha \cdot S(P_{Lx_i}, P_{Rx_i}) \cdot \left\{ \log\left(\frac{1 - P_c(P_{Lx_i}, P_{Rx_i})}{\varphi} \right) - P_c(P_{Lx_i}, P_{Rx_i}) \log\left(\alpha \frac{1 - P_c(P_{Lx_i}, P_{Rx_i})}{\varphi} \right) \right\} - \beta \log(\varphi) \quad (1)$$

where

$$\varphi = \beta + \sum_{P_L=0}^{W_d} \sum_{P_R=0}^{W_d - P_L} \alpha \cdot S(P_L, P_R) \{1 - P_c(P_L, P_R)\} \quad (2)$$

and $P_c(P_L, P_R)$ and $S(P_L, P_R)$ are the correct transition probability and the number of symbols having P_L and P_R pulses generated continuously from left and right ends of the block, respectively.

To calculate the cutoff rate of the channel, we use the formula [1]: $E_x[J]$ is derived as

$$E_x[J] = \beta^2 + \sum_{P_{Lx_i}=0}^{W_d} \sum_{P_{Rx_i}=0}^{W_d - P_{Lx_i}} \alpha \cdot S(P_{Lx_i}, P_{Rx_i}) \left[\sum_{P_{Lx_j}=0}^{W_d} \sum_{P_{Rx_j}=0}^{W_d - P_{Lx_j}} \alpha \cdot S(P_{Lx_j}, P_{Rx_j}) \sqrt{\{1 - P_c(P_{Lx_i}, P_{Rx_i})\}\{1 - P_c(P_{Lx_j}, P_{Rx_j})\}} + \alpha \cdot P_c(P_{Lx_i}, P_{Rx_i}) + \beta \sqrt{\{1 - P_c(P_{Lx_i}, P_{Rx_i})\}} \right] \quad (3)$$

Figure 1 shows the optimal capacity per slot of Q-ary PPM, DPPM, (Q, N) OPPM and DOPPM. It can be seen that DOPPM with new rule can achieve higher capacity than PPM, DPPM, OPPM and DOPPM with conventional rule [1]. This is because some erasure events in [1] that can be decoded correctly are not defined as an erasure event in this paper with new rule. Similar trends can be seen in the cut off rate performance.

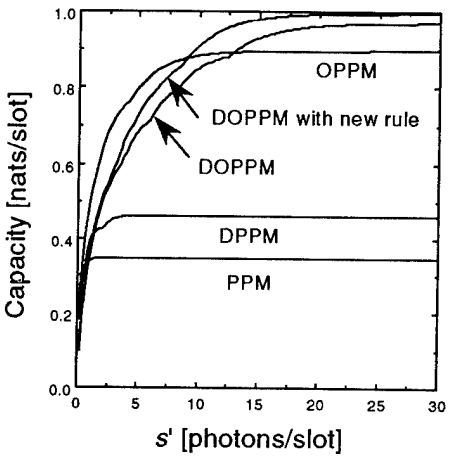


Fig. 1: Optimal capacity per slot [nats/slot] vs. average number of photons per slot s' [photons/slot].

Reference

[1] T. Ohtsuki, I. Sasase and S. Mori, "Differential overlapping pulse position modulation in optical direct-detection channel," *Conf. Rec. ICC'94*, pp. 680-684, New Orleans, Dec. 1994.

Bit-error-rate of an Optical two-users Communications Technique

Amer A. Hassan[1], Karl J. Molnar[1], and Hideki Imai[2]

[1] Ericsson, 1 Triangle Drive, Research Triangle Park, NC 27709 USA. [2] Institute of Industrial Science, University of Tokyo,22-1, Roppongi 7-chome, Minatoku, Tokyo 106, Japan

Abstract — **In this paper, we evaluate the performance of a two-user optical communications system over a noncoherent optical channel. The two users are separated in average received energy and are directly interfering with each other. We find expressions for uncoded bit error probability and codeword error probability for a particular scheme.**

SUMMARY

An information source outputs random binary data in $\{0,1\}$ with equal probability. During a time interval T, the laser of the j-th transmitter, $j = 1, 2$, is amplitude modulated by the data. Both transmitters use the same optical channel to communicate with a central receiver. At the receiver, the laser light is detected noncoherently with a photo-detector to count the photoelectrons.

We assume that photon arrival is due to the transmitting laser(s) only. The photon channel is a Poisson channel where if a positive average real number x is transmitted, the probability that the integer k is received is Poisson distributed according to

$$P(k;x) = e^{-x}\,\frac{x^k}{k!}.$$

Thus the discrete channel seen by a transmitter receiver pair is a Z-channel. Each user has available a distinct Z-channel.

The decoder outputs the information bits based on a maximum likelihood estimate of the transmitted bits given the output of the photo-detector. Let the average energy available to user i be E_i (photons), then the decoder finds

$$\arg \max_{b(1),b(2)} Pr\{m|b(1),b(2)\},$$

where m is the photon count at the output of the photo detector, and $b(j)$ is the bit transmitted by user j, $j = 1, 2$. The decision regions consist of $(0, L]$, $(L, U]$, and (U, ∞), where

$$L = \left\lfloor \frac{E_2 - E_1}{\log\left(\frac{E_2}{E_1}\right)} \right\rfloor \quad U = \left\lfloor \frac{E_1}{\log\left(1 + \frac{E_1}{E_2}\right)} \right\rfloor.$$

The probability of bit error for user j is evaluated for an uncoded system and for a coded system.

A close look at the error rate expressions and Fig. 1 will indicate that E_1 and E_2 should not be equal or close together as this will yield maximum interference. On the other hand if E_1 and E_2 are too far apart, the user with high energy will suffer from large variance, since for the Poisson process the mean and variance are equal (to the average energy). For this example, $E_1 + E_2 = 16$ dB, and the users have equal performance at $E_2 = 12.5$ dB.

Now consider the use of a t-error correcting (n, k) code. Each user pulses between the two Z-channels by alternating transmission. In this case the average transmitted energy E per codeword and per user is the same, and, therefore, both users will have the same probability of error.

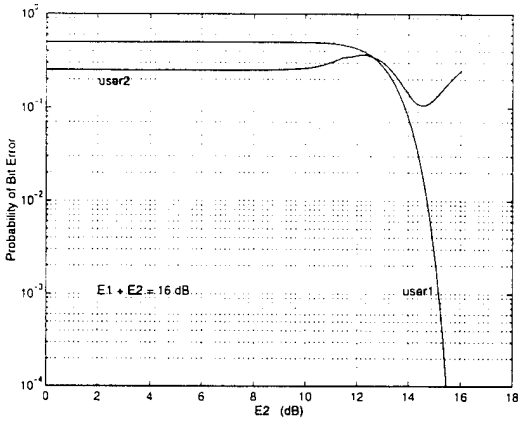


Figure 1: Uncoded performance for $E_1 + E_2 = 16$ dB.

With P_1 and P_2 as defined earlier, the codeword error probability $P_w(E_1, E_2)$ is evaluated. The optimal energy levels are given by

$$(E_1^*, E_2^*) = \arg\min P_w(E_1, E_2),$$

where the above minimization is over (E_1, E_2) with the following constraints

$$E_1 > 0, \quad E_2 > 0, \quad E_1 + E_2 = E.$$

It is not obvious to what values of energy levels result in minimum error rate. The above expression is evaluated numerically for each value of E. Fig. 2 shows typical behaviour of P_w as a function of separation between energy levels, and for a particular code of block length $n = 20$ and $t = 4$.

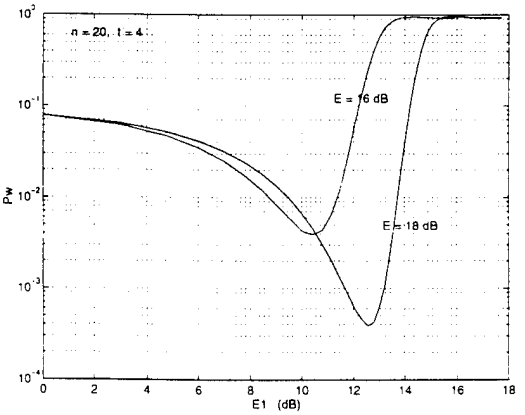


Figure 2: Codeword error rate as a function of separation between energy levels.

BER of Optical Communication System using Fiber Source

Lim Nguyen, Behnaam Aazhang, and James F. Young[1]
Elect. & Comp. Eng. Dept., Rice University,
Houston, TX 77251, USA

Abstract — **We analyze the bit-error rate (BER) of an optical communication system using the superfluorescent fiber source (SFS). The counting statistics of thermal light give improved performance relative to the Gaussian statistics that predict a BER floor.**

Summary

Consider a spectrum-sliced wavelength-division multiple-access (WDMA) system that employs the SFS [1]. In the SFS, spontaneous atomic emission is amplified through a rare-earth doped, single mode, optical fiber end-pumped by an external laser. The incoherent, broad bandwidth output is best modeled as thermal light. The output is spectrum-sliced then on-off modulated by a binary symbol stream, resulting in an intensity modulation waveform arriving at the photodetector. Neglecting the dark current and thermal noise, we obtain the BER as the Laplace transform of the integrated intensity W:

$$BER = \frac{1}{2}\int_0^\infty e^{-\alpha w} p_W(w)dw, \qquad (1)$$

where $p_W(w)$ models the stochastic fluctuation of the light, $\alpha = \eta/h\nu$ (η is the quantum efficiency and $h\nu$ the photon energy). Using the negative binomial statistics for the photoelectron count [2], the BER and the signal/noise ratio are:

$$BER = \frac{1}{2}\left[1 + \frac{\eta P}{P_c} + \left(\frac{\eta P}{2P_c}\right)^2 (1-\mathcal{P}^2)\right]^{-M}, \qquad (2)$$

$$SNR = \frac{M}{P_c/\eta P + .5(1+\mathcal{P}^2)}. \qquad (3)$$

The mode number, M, is the ratio of the symbol duration T to the coherence interval T_c of the incident light. That is, $M = B_o/2B_e$, where $B_o = 1/T_c$ and $Be = 1/2T$ are the optical and detection bandwidths, respectively. $P = E[W]/T$ is the received power, $P_c \triangleq h\nu/T_c$ is the coherence power of the photon and $\mathcal{P}$ the degree of polarization. The limiting SNR is $B_o/(1+\mathcal{P}^2)Be$ for high received power ($\eta P/P_c >> 1$).

The BER approaches the shot-noise limited value of $.5e^{-\alpha E[W]}$ if the count degeneracy parameter, $\eta P/P_c$, is much smaller than unity. Since the BER decreases monotonously with T_c, a Lorentzian spectral shape can have a lower BER at a higher symbol rate compared to a Gaussian shape with the same power and 3 dB linewidth. The ideal rectangular linewidth has the worst performance. This must be considered against the channel crosstalk since, not surprisingly, the tail of a Lorentzian has the slowest spectral decay. Without the linewidth constraint, we obtain a rather interesting theoretical result that the shot-noise performance is achieved with a spectral shape of infinite linewidth and zero spectral height. In comparision, this is also achieved with an ideal, coherent laser of zero linewidth and infinite spectral height.

[1] This work was supported by the Advanced Technology Program of the Texas Higher Education Coordinating Board, GTE, Inc., the U.S.A.F. Phillips Laboratory and its Palace Knight program.

In spectrum-sliced WDMA, the maximum SNR is inversely proportional to the number of channels. Equation (2) demonstrates that the BER is not determined solely by the SNR which reaches a limiting value; increasing the spectral intensity of the light reduces the BER. In fact, the number of channels can be increased by increasing the received power, while maintaining a desired BER for a given symbol rate. This has important implications for spectral amplitude encoded CDMA systems that require a large number of spectral chips [3]. Invoking the Gaussian assumption [1] would lead to incorrect conclusions. For example, the Gaussian predicts a BER floor due to the limiting SNR and therefore expects that it would be impossible to increase the number of channels by increasing the power, once the limiting SNR has been reached.

One can show that the BER is lower bounded by using the fact that $e^{-\alpha w}$ is convex over $[0,\infty)$ and applying Jensen's inequality to Eq. 1 to obtain $BER = .5E[e^{-\alpha W}] \geq .5e^{-\alpha E[W]}$. Accordingly, a light source that achieves this can be considered ideal, i.e. its intensity is deterministic: $p_W(w) = \delta(w - E[W])$, as would be expected. Figure 1 shows calculated BER values for the ideal and spectrum-sliced fiber sources, and the Gaussian assumption. The Gaussian predicts a BER floor and underestimates the true performance.

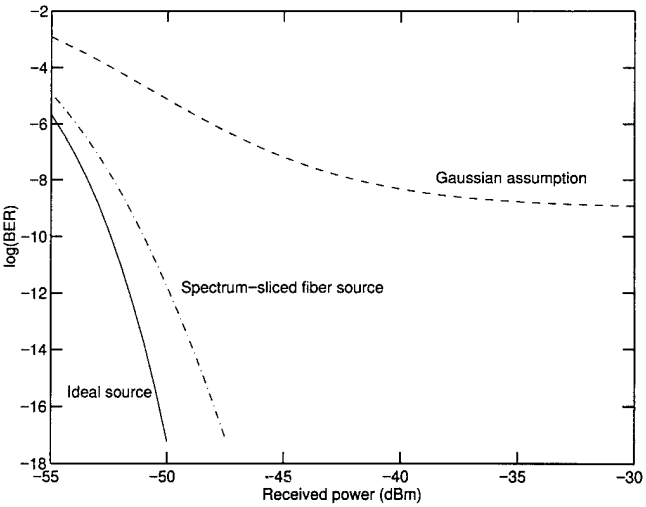


Fig. 1: BER comparisions at 1 Gbps. The spectrum-sliced SFS is polarized ($\mathcal{P} = 1$); $M = 36$, $\eta = 50\%$ at 1550 nm wavelength.

References

[1] J. S. Lee, Y. C. Chung, T. H. Wood, J. P. Meester, C. H. Joyner, C. A. Burrus, J. Stone, H. M. Presby, and D. J. Digiovanni. "Spectrum-Sliced Fiber Amplifier Light Source With a Polarization-Insensitive Electroabsorption Modulator," *IEEE Photonics Tech. Letters*, 6, (8), pp. 1035-1038, 1994.

[2] C. L. Mehta. "Theory of Photoelectron Counting," *Progress in Optics*, 8, edited by E. Wolf, pp. 375-440. North-Holland, Amsterdam, 1970.

[3] L. Nguyen, B. Aazhang and J. F. Young. "All-optical CDMA with Bipolar Codes," *Electron. Lett.* 31, (6), pp. 469-470, 1995.

The Channel Capacity of Hybrid Fiber/Coax (HFC) Networks

Kenneth J. Kerpez
Bellcore, 445 South Street, Morristown, NJ 07960

Hybrid fiber/coax (HFC) is emerging as an inexpensive architecture for providing broadband services to residences. It has optical fibers extending from the central office or head-end to remote fiber nodes. Extending from the fiber nodes to the residences is a coaxial cable distribution bus. Multiplexing allows 100 to 500 users to share the bandwidth of each coax distribution bus.[1] This architecture advantageously combines the long range of optical fiber with the high bandwidth and simple electrical interfaces of coaxial cable. HFC will initially provide telephony and cable TV, but it also has sufficient bandwidth for future interactive and multimedia services. Many regional telephone companies, and most cable TV companies, in the U.S. have committed to HFC. This architecture will be widely deployed well into the future, while the demand for residential bandwidth will increase. To meet this demand, there will be an increasing need for multi-user information theory and communication techniques to maximize the bandwidth and fully exploit the potential of this unique medium. This paper initializes exploration into maximizing the channel capacity of HFC.

Capacity calculations can safely ignore the optical link and instead focus on the coax distribution bus. The coax bus has a physical tree-and-branch architecture, but it is logically a shared bandwidth bus. Receivers are spatially distributed along the bus, with propagation distance l_i between the fiber node and receiver i. The magnitude response of a receiver i is $|H_i|=\Gamma^{l_i\sqrt{f}}$ where Γ is a constant and f is the frequency. The transmitted signal is $s(t)$, which is attenuated by H_i and delayed by l_i/v, where v is the propagation velocity. The signal received by the ith receiver, on carrier frequency f, is $r_i(t)=\Gamma^{l_i\sqrt{f}}s(t-l_i/v)$. The Fourier transform of $r_i(t)$ is $R_i(f)=\Gamma^{l_i\sqrt{f}}e^{-j2\pi f l_i/v}S(f)$, where $j=\sqrt{-1}$. Define $D=\Gamma e^{-j2\pi\sqrt{f}/v}$. Then $R_i(f)=D^{l_i\sqrt{f}}S(f)$. This response is fundamental to capacity calculations.

Gaussian noise is assumed. For a single user, the capacity is easily solved with the classic "water-filling" spectral density. However, there are multiple users, and the capacity is a multi-dimensional function with dimensionality equal to the number of users. Communication from the fiber node to the users is similar to the classic Gaussian broadcast channel, and from the users to the fiber node is similar to the classic Gaussian multiple access channel. Classic multi-user information theory assumes that the channel response is the same to each user. However, users on the coax bus are located at different distances from the fiber node, so although they each see the same superposition of signals, they each have a different channel response. Unlike classic multi-user capacity calculations, here the ensemble of propagation distances is a fundamental new variable. This problem is unsolved in its general form.

Assumptions are made to get the results here. Only two types of communications are considered: information that is broadcast to all users, and information that is specific to only a single user. Each user's specific information has the same bit-rate, and is carried in a single distinct interval on the frequency axis. The capacity of the user specific information transmitted to each user is calculated here. Closer users are assigned higher frequencies than more distant users are.

Two types of coax bus architectures are examined: a cable TV type of coax network with analog amplifiers and attenuating taps, and a passive coax network with ideal band-pass filter taps. A typical suburban tree-and-branch coax distribution bus [2] is modeled with 250 feet of 0.625 inch coax between the splitters and four-way taps. Users connect to the taps with 250 feet of coax drop. Each user's channel response is calculated, and with typical cable transmission parameters SNRs (about 40 dB) are found as a function of frequency. Downstream digital signals are restricted to the 450 to 1000 MHz band. Using Shannon theory, differential entropies and capacities are found, then frequency assignments are numerically calculated to maximize the capacity, which is shown in Fig.1. The sum total capacity is multiple Gbps.

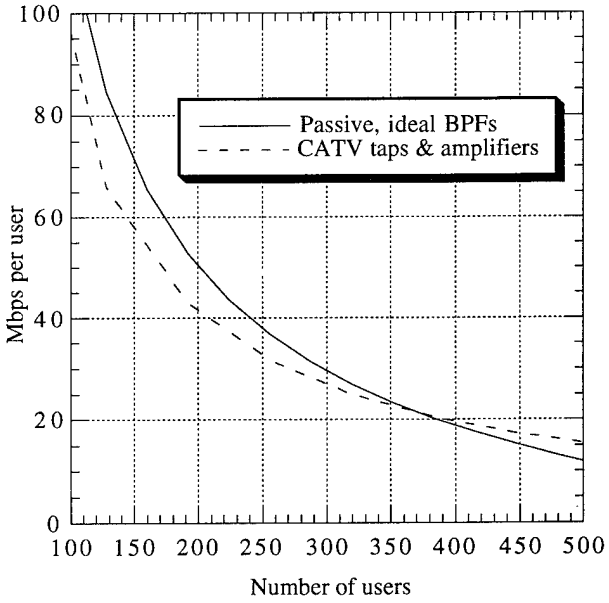


Fig.1. Shannon capacity dedicated downstream to each user.

There are simplifying assumptions here, and the multi-user channel capacity of the shared coax bus is in general an unsolved problem. The shared upstream channel from 5 to 42 MHz has much radio ingress, causing many noise spikes in the frequency domain, and making the upstream capacity calculation more difficult.

REFERENCES

[1] K. J. Kerpez "A Comparison of QAM and VSB for Hybrid Fiber/Coax Digital Transmission," *IEEE Trans. Broadcasting*, pp. 9-16, Vol. 41, March, 1995.
[2] E. R. Bartlett, *Cable Television Technology & Operations,* New York: McGraw-Hill, 1990.

Error Probability Evaluation of Optical Systems Disturbed by Phase Noise and Additive Noise

Göran Einarsson, Johan Strandberg and Idelfonso Tafur Monroy
Telecommunication Theory, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Abstract— **A direct and efficient method for evaluation of the error probability of optical heterodyne receivers in the presence of phase noise is presented.**

Closed form expressions for the statistics of the decision variable, including photodetector shot noise and thermal noise from electronic circuitry, are shown.

I. Introduction

The decision variable, in complex signal notation, of a heterodyne optical system with an envelope detector receiver has the form

$$|V|^2 = |\mathcal{Y} + X|^2 \tag{1}$$

where $\mathcal{Y}$ represents phase noise and X additive noise. The phase noise is produced by the transmitting and local oscillator lasers. The additive noise X is photodetector shot noise and thermal noise from the electric circuitry.

Further background and derivation of the formulas can be found in a forthcoming paper by Einarsson et al [1].

II. Amplitude-Shift Keying

To simplify the analysis let the prefilter be a bandpass integrator operating at the heterodyne frequency. During the data symbol interval the prefilter output is sampled L times at $t = kT'$, $k = 1, 2, \ldots, L$, generating a sequence of complex valued stochastic variables

$$V_k = A\mathcal{Y}_k + X_k \tag{2}$$

where

$$\mathcal{Y}_k = \frac{1}{T'} \int_{(k-1)T'}^{kT'} e^{j\theta(t)} dt \tag{3}$$

is filtered phase noise and X_k, filtered white noise, is a complex valued, zero mean Gaussian variable.

The phase noise is a continuous Brownian motion (Wiener-Lévy) process with Gaussian statistics. The primary statistical properties of $\theta(t)$ are easily specified but the probability distribution of $\mathcal{Y}_k$ is difficult to determine. Foschini and Vannuci [2] obtained a closed form approximate result by expanding the integrand $e^{j\theta(t)}$ in (3) in a Taylor series and keeping the first terms,

The decision variable U is the sum of $L = T/T'$ equally distributed independent variables $|V_k|^2$ and the approximate moment-generating function (mgf) of U is

$$\Psi_U(s) = \frac{1}{(1-s)^L} \exp\left(\frac{m_1 s}{1-s}\right) \left[\operatorname{sinch}\sqrt{\frac{2\beta m_1 s}{(1-s)L^2}}\right]^{-L/2} \tag{4}$$

where "sinch" denotes the hyperbolic sinc-function.

The parameter $\beta = 2\pi B_L T$ is equal to 2π times the product of the data symbol interval T and B_L, the sum of the 3-dB linewidths of the lasers at the transmitter and the local oscillator. The quantity $m_1 = A^2T/2 = A^2LT'/2$ is the expected number of photoelectrons in the received optical pulse.

III. Frequency-Shift Keying

Frequency-Shift Keying (FSK) is readily analyzed utilizing the results from ASK since an FSK receiver contains two branches, each identical to an ASK receiver.

IV. Differential Phase-Shift Keying

In Differential Phase-Shift Keying (DPSK) the phase of the transmitted optical field is modulated and the phase of the previous signal is used as a phase reference in the receiver.

We consider the case without predetector filtering where $T' = T$ and one sample per signal interval is generated.

An approximate mgf of the decision variable U is

$$\Psi_U(s) = \frac{\exp\left(\frac{2ms}{1-s}\right)}{\sqrt{1-(2m\beta/3+1)^2 s^2}\sqrt{1-s^2}} \tag{5}$$

V. Error Probability

The moment-generating function determines the statistical distribution of the decision variable. The transmission error probability is easy to calculate from $\Psi_U(s)$ using the saddlepoint approximation suggested by Helstrom [3]. The optimal value of the prefilter bandwidth parameter L is readily determined by this procedure.

The theory presented applies also to receivers with an optical preamplifier in the presence of phase noise. We refer to the text by Einarsson [4] for a discussion.

References

[1] G. Einarsson, J. Strandberg, and I. Tafur Monroy, "Error Probability Evaluation of Optical Systems Disturbed by Phase Noise and Additive Noise," To be published in *J. Lightwave Tech.*

[2] G. J. Foschini and G. Vannucci, "Characterizing Filtered Light Waves Corrupted by Phase Noise," *IEEE Trans. Info. Theory*, vol. 34, pp. 1437-1448, Nov 1988.

[3] C. W. Helstrom, Performance Analysis of Optical Receivers by the Saddlepoint Approximation," *IEEE Trans. Comm.*, vol. COM-27, pp. 186-191, Jan 1979.

[4] G. Einarsson "Principles of Lightwave Communications", Wiley, Chichester 1995.

This work was supported by the Swedish Research Council for Engineering Sciences

Applications of Coding and Design Theory to Constructing the Maximum Resilient Systems of Functions

Vladimir I. Levenshtein[1]

Institute for Applied Mathematics, RAS, Miusskaya Sq.4, 125047, Moscow, Russia

Abstract — **Some ideas, methods, and results of coding and design theory, especially a duality in bounding the optimal size of codes and designs (orthogonal arrays), are used to solve a new problem connected with randomized systems of functions.**

I. Introduction

A system of functions in n variables is called randomized if the functions preserve the property of their variables to be independent and uniformly distributed random variables. Such a system is called t-resilient if for any substitution of constants for any i variables, $0 \le i \le t$, the derived system of functions in $n-i$ variables is also randomized. A system of N Boolean functions in n variables of which any T form a t-resilient system is referred to as a (n,t,N,T)-system ($1 \le T \le N$, $0 \le t \le n$). We investigate the problem of finding the maximum number $N = N(n,t,T)$ of functions in a (n,t,N,T)-system. This problem is reduced to the minimization of the size of certain combinatorial designs, which we call split orthogonal arrays (SOA). A binary code C of length $n+N$ is called (n,t,N,T)-SOA if for any choice of binary word of length $t+T$ and any choice of $t+T$ places of which t belong to the first n places and T belong to the last N places there are exactly $|C|2^{-t-T}$ code words which contain this word in these places. Let $B(n,t+1,N,T+1)$ be the minimum size of a code C which is (n,t,N,T)-SOA.

II. Linear programming bounds

A (n,t,N,T)-system exists if and only if there exists a systematic (n,t,N,T)-SOA with the first n information symbols. This gives the following necessary condition for existence of a (n,t,N,T)-system:

$$2^n \ge B(n,t+1,N,T+1).$$

We extend the linear programming method of Delsarte [1] to obtain a lower bound on $B(n,t+1,N,T+1)$ and an upper bound on $N(n,t,T)$. Let $A(n,d)$ ($B(n,d)$) be the maximum (minimum) size of a code of length n with the minimal distance (respectively, with the dual distance) d. Let $K_k^n(z) = \sum_{j=0}^{k}(-1)^j\binom{z}{j}\binom{n-z}{k-j}$ be the Krawtchouk polynomial of degree k and suppose that for an arbitrary polynomial $f(z) = \sum_{i=0}^{n} f_i K_i^n(z)$, $\Omega(f) = f(0)/f_0$. If $A^*(n,d) = \min\Omega(f)$, where the minimum is taken over all polynomials $f(z)$ such that $f_0 > 0, f_i \ge 0$ for $i = 1,2,\dots,n$, and $f(0) > 0, f(i) \le 0$ for $i = d,\dots,n$; and $B^*(n,d) = \max\Omega(f)$, where the maximum is taken over all polynomials $f(z)$ such that $f_0 > 0, f_i \le 0$ for $i = d,\dots,n$, and $f(0) > 0, f(i) \ge 0$ for $i = 1,2,\dots,n$, then by the Delsarte inequalities,

$$A(n,d) \le A^*(n,d),\ B(n,d) \ge B^*(n,d).$$

Delsarte [1] found an $f(z)$ which gives the Rao bound

$$B^*(n,d) \ge R(n,d),$$

where $R(n,2l+1+\sigma) = 2^\sigma \sum_{i=0}^{l}\binom{n-\sigma}{i}$ when $\sigma \in \{0,1\}$. The author [2] found polynomials which imply

$$A^*(n,d) \le$$

$$L(n,d) = \begin{cases} L_k^n(d) & \text{if } d_k(n-1) \le d-1 \le d_{k-1}(n-2) \\ 2L_k^{n-1}(d) & \text{if } d_k(n-2) \le d-1 \le d_k(n-1), \end{cases}$$

where $d_k(n)$ is the smallest root of $K_k^n(z)$ and

$$L_k^n(z) = \sum_{i=0}^{k-1}\binom{n}{i} - \binom{n}{k}\frac{K_{k-1}^{n-1}(z-1)}{K_k^n(z)}.$$

Using the linear programming method for bounding $B(n,t+1,N,T+1)$ and the important relationship $A^*(n,d)B^*(n,d) = 2^n$ proved in [3], we obtain

Theorem 1. If there exists a (n,t,N,T)-system, then $2^n \ge R(n,t+1)R(N,T+1)$, $L(n,t+1) \ge R(N,T+1)$, and $2^n L(N,T+1) \ge 2^N R(n,t+1)$.

III. Sufficient condition

Let $l(n,d)$ be the minimum number of information symbols in a systematic binary code of length n with the dual distance d. In [4] it was shown that the condition $T \le n - l(n,t+1)$ is sufficient for the existence of a (n,t,T,T)-system. In the general case we have

Theorem 2. If $l(n,t+1) + l(N,T+1) \le n$, then there exists a (n,t,N,T)-system.

Theorems 1 and 2 give rise to good asymptotic bounds on $N(n,t,T)$ and imply complete results in some cases.

Examples. For any h=2,3,... ,

$$N(n,3,n/2-1) = n \quad \text{if} \quad n = 2^{h+1},$$
$$N(n,5,(n-\sqrt{n})/2-1) = n \quad \text{if} \quad n = 2^{2h},$$
$$N(n,3,5) = \sqrt{2^{n-1}/n} \quad \text{if} \quad n = 2^{4h-1}.$$

Indeed, the existence of the Hamming, Kerdock and Preparata codes implies that $l(n,4) = \log 2n$, $l(n,n/2) = n - \log 2n$ when $n = 2^{h+1}$, and $l(n,6) = 2\log n$, $l(n,(n-\sqrt{n})/2) = n - 2\log n$ when $n = 2^{2h}$. This gives the corresponding lower bounds by Theorem 2. On the other hand, $R(n,4) = 2n$, $L(n,n/2) = 2n$, $R(n,6) = n^2 - n + 2$, $L(n,(n-\sqrt{n})/2) = n^2 + \sqrt{n-1}(n-2)/2$ and the same upper bounds follow from inequalities of Theorem 1.

References

[1] Ph.Delsarte, "An algebraic approach to the association schemes of coding theory", *Philips Res. Reports, Suppl.*, vol. 10, 1973.

[2] V.I.Levenshtein, "On choosing polynomials to obtain bounds in packing problems", *in Proc. Seventh All-Union Conf. on Coding Theory and Inform. Transmission*, pt. 2, Moscow-Vilnius, 1978, pp.103-108 (in Russian).

[3] V.I.Levenshtein, "Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces". To appear in *IEEE Trans. Inform. Theory.*

[4] D.R.Stinson and J.Massey, "An infinite class of counterexamples to a conjecture concerning non-linear resilient functions". To appear in *Journal of Cryptology.*

[1]This work was partially supported by RFBR under grant 95-011-03 and by ISF under grant MEF300.

McEliece Public Key Cryptosystems Using Algebraic-Geometric Codes

H. Janwa
The Mehta Res. Inst. of Math. and Phys., Allahabad-211 002, India

O. Moreno
Dept. of Math., UPR, San Juan, PR 00931 USA

Abstract — **McEliece proposed a public-key cryptosystem based on binary linear codes, in particular binary classical Goppa codes. In this talk we will look at various aspects of McEliece's scheme in the general setting of q-ary codes. In particular, we consider schemes based on much larger class of q-ary algebraic-geometric (AG) Goppa codes, subfield subcodes of AG codes, and concatenated codes. We will give explicit constructions of several schemes which have very high work factor, excellent key-length/plain-text ratios, and relatively smaller size of the keys for given work factors. We will also present its modifications and generalizations following Krouk and others. Finally, we will discuss some open problems.**

I. Introduction

In 1978, McEliece [2] introduced a public key cryptosystem (PKS) based on binary linear codes and suggested the implementation of his scheme by randomly selecting the generator matrix of a $[1024, 524, 101]$ Goppa code and suitably modifying it. The security of this scheme is based on the well known NP-completeness of the decoding problem for general linear codes and the fact that there are a huge number of inequivalent Goppa codes with the given parameters.

For practical applications that need flexibility and complexity (e.g., computer communication), we will look at various aspects of McEliece's scheme using the newer and much larger class of q-ary AG Goppa codes. We will also present modifications and generalizations of this scheme using the ideas of Krouk and others. Furthermore, we also make some observations on the cryptanalytic attacks. We first discuss the McEliece PKS in the general setup applicable to q-ary codes. We show by analysis, by examples, and by heuristics that the complexity of breaking this scheme under one widely discussed attack is greater than previously believed.

II. The Proposed Schemes

Our constructions, modifications, and generalizations are based on the following coding schemes:

(A) AG codes defined over a finite field with q elements (immensely many choices are attained by varying various parameters of the corresponding curves);

(B) Subfield subcodes of q-ary codes in (A). This includes binary AG codes, some of which perform better than binary Goppa codes.

(C) Concatenation of q^m-ary AG code with good q-ary codes.

In each of the cases (A)–(C), we give explicit constructions of schemes where the work factor is quite substantial. They have excellent key-length/plain-text ratios and relatively smaller size of the key for the same work factor. The decrypting complexity from schemes based on plane curves, especially from maximal curves, is $\mathcal{O}(n^3)$ or better.

III. On the attack of Sidelnikov and Shestakov

Sidelnikov and Shestakov (S-K) [4] have shown that the Niederreiter PKS [3] scheme (known now to be equivalent to the McEliece PKS scheme) is insecure for the particular case of generalized Reed-Solomon codes.

The attack of S-K depends fundamentally on the Vandermonde structure of the generalized RS codes (and also on their MDS property), and is not applicable to systems based on other types of codes. For various considerations, our schemes are excellent alternatives.

IV. Implementing Krouk and Gabidulin Modifications

Krouk [1] strengthens the McEliece scheme by trying to remove symmetry from the coding scheme. We make some observations on his modification and show that AG codes are particularly suitable for it.

We will also present improvements, modifications, and implementations of some recent PKS schemes of Gabidulin.

V. Further Improvements and Open Problems

We show that many more constructions of PKS using AG codes are possible if certain curves that are *maximal* or *near maximal* exist.

Some of the results sumarize in this article will appear in details in Design Codes and Cryptography.

Acknowledgements

Work partially supported by NSF grants RII-9014056, component IV of the EPSCoR of Puerto Rico Grant and the ARO grant for Cornell MSI. The authors thank Prof. Harold F. Mattson Jr. for his helpful comments and John Ramirez for some help with computation. The first author is very thankful to the Gauss Laboratory of UPR for support and hospitality during October–November 1991 and November–December 1992 when some of this work was done.

References

[1] E. Krouk, "A new public key cryptosystem," *Proceedings of the Sixth Swedish-Russian International Workshop on Information Theory*, pp. 285–286, 1993.

[2] R.J. McEliece, "A Public-key cryptosystem based on algebraic coding theory," *DSN Progress Report*, Jet Propulsion Laboratory, Pasadena, CA, pp. 114–116, Jan./Feb. 1978.

[3] H. Niederreiter, "Knapsack-type cryptosystems and algebraic coding theory," *Problems of Control and Information Theory*, vol. 15, no. 2, pp. 159–166, 1986.

[4] V.M. Sidelnikov and S.O. Shestakov, "On insecurity of cryptosystems based on generalized Reed-Solomon codes," *Diskretnaya Matematika*, vol. 4, No. 3, 1992. Translated in, *Discrete Math. Appl.*, vol. 2, No. 4, pp. 439–444, 1992.

Binary Trinomials Divisible by a Fixed Primitive Polynomial

R. A. Games, E. L. Key, and J. J. Rushanan[1]
The MITRE Corporation, Bedford, MA 01730

I. Introduction

This paper examines the growth of the degrees of binary trinomials that are divisible by a fixed binary primitive polynomial $f(x)$ of degree n. Our goal is to find a heuristic distribution that depends only on n. Our motivation stems from some suggested correlation attacks on certain stream ciphers [1, 2, 3]. These attacks use binary relations—binary polynomials—as parity checks in order to recover information about the cipher key. Low weight relations perform best but require more sequence because of their large degrees.

II. Binary Trinomials

Let α be a primitive element of GF(2^n) with minimum polynomial $f(x)$. We consider the set of 3-term or trinomial relations $\alpha^b + \alpha^a + 1 = 0, b > a > 0$. That is, $f(x)$ divides x^b+x^a+1. If $b \in \{1,\ldots,2^n-2\} = \mathcal{I}$, then α^b+1 is some power of α with exponent also in $\mathcal{I}$. Thus the trinomials partition $\mathcal{I}$ into pairs. We denote the set of all trinomials as an ordered listing of ordered pairs:

$$\left\{(b_i, a_i) \mid b_i > a_i;\ b_i \text{ increasing};\ i = 1,\ldots,2^{n-1}-1\right\} \quad (1)$$

We call such an ordered listing of pairs a *pattern of trinomials.* As an example, the two trinomial patterns for $n = 3$ are

$$3\,1 \quad 5\,4 \quad 6\,2 \quad \text{and} \quad 3\,2 \quad 5\,1 \quad 6\,4.$$

Consider all partitions of $\mathcal{I}$ that are in the canonical form of (1); we call such partitions *patterns.* We take all patterns to be equally likely, and we model choosing a random trinomial pattern, i.e., a primitive polynomial, as choosing a random pattern. We model the distribution of trinomial degrees by the distribution of the size of b_i over all patterns:

$$R_i(k) = \text{Prob}(b_i = k)$$

In particular, $R_1(k)$ models the distribution of the lowest degree trinomial. We derive the distributions by considering patterns as in (1) defined for a general index set $\mathcal{I} = \{1,\ldots,N = 2M\}$. The distribution $R_i(k)$ is only nonzero for k between $2i$ and $M+i$.

III. Formula for $R_i(k)$

The calculation of the $R_i(k)$ is combinatorial and follows from calculating the total number of patterns and those patterns with $b_i = k$. (We also have an alternative derivation as a classical "birthday problem" in probability.)

Proposition 1 *For $i = 1,\ldots,M$ and $k = 2i,\ldots,M+i$,*

$$R_i(k) = 2^{k-2i+1}\,\frac{(k-1)!}{(k-2i)!(i-1)!}\,\frac{(N-k)!}{N!}\,\frac{M!}{(M-k+i)!}.$$

In Figure 1, we plot in ascending order (the 'x' curve) the degrees of the trinomials divisible by $f(x) = x^{16}+x^5+x^3+x^2+1$. Since the x-coordinates correspond to the i index in $R_i(k)$, we also plot for a given i the mean of the $R_i(k)$ distribution and ± 2 standard deviations from the mean. Though the actual trinomial curve levels off sooner than the model, the model captures the essence of the growth in the degrees.

[1]The authors were supported by the MITRE Sponsored Research program.

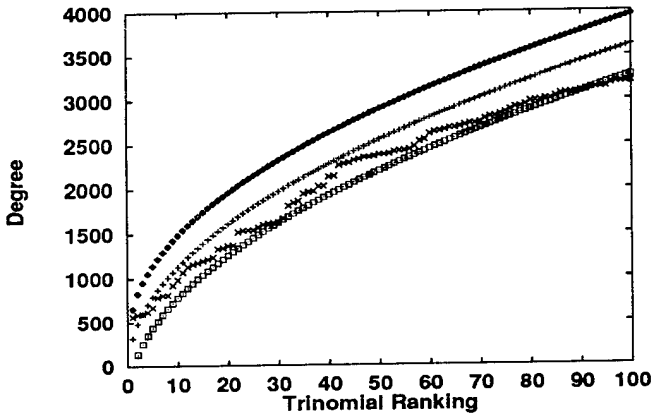


Figure 1: Growth of Trinomial Degrees ($n = 16$).

IV. Approximate Distributions

As N gets large, $R_1(k) \cong \frac{k}{N}e^{-k^2/N}$. This approximation yields an approximation for the mean: $\sqrt{N\pi/2}$. The general distribution can be approximated similarly:

$$R_i(k) \cong \frac{k^{2i-1}}{2^{i-1}N^i(i-1)!}e^{-k^2/N}.$$

If we replace k with a continuous variable, then the distribution is in fact a generalized Rayleigh distribution with parameters $2i$ and $\sqrt{N}$. In particular, a good approximation to the mean of $R_i(k)$ is

$$\frac{1\cdot 3\cdots(2i-1)}{2^{i-1}(i-1)!}\sqrt{N\pi/2}$$

Such formulas present an easy way to generate the model curves as in Figure 1 and offer an analytic method to measure the growth of the degrees of the binary trinomials.

Acknowledgements

We wish to thank Michael Sousa for useful insights and Brian Sroka for some help with the empirical results.

References

[1] V. Chepyzhov and B. Smeets, 1991, "On a Fast Correlation Attack on Certain Stream Ciphers," *Advances in Cryptology—EUROCRYPT'91*, Lecture Notes in Computer Science #547 (D. W. Davies, Ed.), Berlin: Springer-Verlag, pp. 176–185.

[2] W. Meier and O. Stafflebach, 1989, "Fast Correlation Attacks on Certain Stream Ciphers," *J. Cryptography*, Vol. 1, pp. 176–185.

[3] K. Zeng and M. Huang, 1990, "On the Linear Syndrome Method in Cryptanalysis," *Advances in Cryptology—CRYPTO'88*, Lecture Notes in Computer Science #403 (S. Goldwasser, Ed.), Berlin: Springer-Verlag, pp. 469–478.

New Digital Multisignature Scheme in Electronic Contract Systems

Chang Goo Kang
Electronics and Telecommunications Research Institute, P.O. BOX106, YUSONG, Taejon 305-606, KOREA

Abstract — **This paper analyzes risks and presents the requirements of digital multisignature scheme in electronic contract systems. A new digital multisignature scheme suitable for contract systems is proposed and the efficiency of the scheme is discussed.**

I. Introduction

The electronic contract system needs to replace hand written signatures with digital signatures, digital multisignature might also be needed in such environments where several persons must sign the same digital message.

There are the following potential risks: signature forgery, contract with the unauthorized party, denial of contract, misuse of contractor's signature, malicious contract destruction. It is desired that the digital multisignature scheme satisfy the following requirements in the electronic contract system: verifiability, viability, dishonesty - detectability, commonness (common procedures), generality, orderlessness.

This paper assumes that m users join the electronic contract system and sign the same contract message, and all signers are connected by a bridge node or MCU. Let M be the contract message to be signed. f and h denote public one-way functions which are easily computable and are hard to invert. Let ID_i and ID_{cm} denote the identification information of user (contractor) i and the concatenation of signers' IDs, i.e., $ID_{cm} = ID_1 \,||\, ID_2 \,||\, \cdots \,||\, ID_m$.

II. Key generation and publication

Signer i registers his identification information (ID_i) and the trusted center issues a smart card as follows :

1. The trusted center selects two large prime numbers p and q, and keeps them secret.
2. The trusted center publishes a modulus N which is the product of p and q.
3. The trusted center calculates integers S_{ij} for signer i :

$$I_{ij} = f(ID_i, j), \quad j = 1, \cdots, k \tag{1}$$

$$I_{ij}^{-1} = S_{ij}^2 \pmod N \tag{2}$$

4. The center issues a smart card to signer i after identifying his physical identity.

The smart card includes the set of $(N, f, h, S_{i1}, \cdots, S_{ik})$.

III. Multisignature generation

1. The signer n generates a random integer $R_n \in Z_N$ and calculates

$$X_n = R_n^2 X_{n-1} \pmod N \tag{3}$$

$$(e_{n1}, \cdots, e_{nk}) = h(M, ID_{cm}, X_n) \tag{4}$$

$$Y_n = Y_{n-1} R_n \prod_{e_{nj}=1} S_{nj} \pmod N \tag{5}$$

where $X_0 = 1, Y_0 = 1$ and $j = 1, \cdots, k$

2. The signer n broadcasts (X_n, Y_n) to all the other signers.

IV. Multisignature verification

When the multisignature generation procedures were completed, the verifier or signer gets the multisignature $(M, ID_{cm}, X_1, \cdots, X_m, Y_m)$, the verifier calculates

$$(e_{i1}, \cdots, e_{ik}) = h(M, ID_{cm}, X_i), \quad i = 1, \cdots, m \tag{6}$$

and stores only $(M, ID_{cm}, (e_{11}, \cdots, e_{1k}), \cdots, (e_{m1}, \cdots, e_{mk}), Y_m)$ for verification of the multisignature. When multisignature verification is requested, the verification procedures are as follows:

1. The verifier calculates I_{ij} with ID_{cm}.

$$I_{ij} = f(ID_i, j), \quad i = 1, \cdots, m, j = 1, \cdots, k \tag{7}$$

2. The verifier calculates Z_m.

$$Z_m = Y_m^2 \prod_{i=1}^{m} \prod_{e_{ij}=1} I_{ij} \pmod N, \quad j = 1, \cdots, k \tag{8}$$

3. The verifier calculates $h(M, ID_{cm}, Z_m)$ and checks whether the equation

$$(e_{m1}, \cdots, e_{mk}) = h(M, ID_{cm}, Z_m) \tag{9}$$

holds true.

If it does, the multisignature message is considered to be valid.

V. Efficiency and Conclusions

The proposed digital multisignature scheme satisfies with all the requirements of multisignature scheme in electronic contract systems. The proposed scheme requires $(k/2 + 3)t$ modular multiplications to generate a signature, m transmissions to complete the multisignature procedure and the information redunduncy of $(m|ID| + ktm + |N|)$ bits must be stored for multisignature verification where t is a security level parameter.

Since the new proposed multisignature scheme is based on the Fiat-Shamir scheme, the scheme is more efficient than other RSA based multisignature schemes and as secure as the Fiat-Shamir scheme. Owing to the high processing speed and the high degree of satisfaction to the requirements, the new proposed scheme is suitable for electronic contract systems.

References

[1] A.Fiat and A.Shamir, "How to prove yourself: Practical Solutions to Identification and Signature Problems," *Advances in Cryptology - Crypto'86.*, Lecture Notes in Computer Science 263, pp. 186–423, 1987.

[2] C.G. Kang, D.Y. Kim, D.H. Kim and D.K. Lee, "New Sequential and Simultaneous Multisignature Schemes," *Proceedings of ISITA'94.*, vol.1 , pp. 283–288, 1994.

The Binary Symmetric Broadcast Channel with Confidential Messages, with Tampering

Marten van Dijk

Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, Email address: marten@win.tue.nl

Abstract — **We introduce a new model, the broadcast channel with confidential messages, with tampering. Here, the enemy not only taps the wire but also actively tampers the signal communicated over the wire. We show that the legitimate users always have to take a certain worst case scenario into account.**

Csiszár and Körner [1] introduced the broadcast channel with confidential messages (BCC). It consists of three participants: two legitimate users of the main channel, Alice and Bob, and a wire-tapper, Eve, the enemy. Alice and Bob want to generate a secret key such that Eve can only obtain a negligible amount of information about it. It is assumed that all players know everything; the codes and protocol used by the legitimate users, and the noise characteristics of the main and wire-tap channel. The central question is to determine the secrecy capacity Cs, which is the maximal rate at which Alice and Bob can generate a secret key.

We introduce the BCC, with tampering (BCCT), in which in addition Eve actively tampers. Now, solely Eve is assumed to know everything. Alice and Bob can measure the noise characteristics of the main channel, and they have limited knowledge about the noise characteristics of the wire-tap channel.

If Alice wants to transmit the binary signal $\mathbf{a} \in \{0,1\}^n$ then she converts it into a polar analog signal $a(t)$ with signal power S_A, which she transmits to Bob over a distortionless channel with length $l_A + l_B$ and attenuation coefficient α. The first part of the main channel from Alice to the position where Eve taps the wire has length l_A, and, hence, transmission loss $(L_A)_{dB} = \alpha l_A$. The second part of the main channel has length l_B and transmission loss $(L_B)_{dB} = \alpha l_B$. Bob uses an amplifier with noise figure n_B and power gain g_B to obtain an analog signal $b(t)$, which he converts to a binary signal $\mathbf{b} \in \{0,1\}^n$. The wire-tap channel of Eve causes transmission loss L_E. Eve uses an amplifier with noise figure n_E and power gain g_E, to obtain an analog signal $e(t)$, which she converts to a binary signal $\mathbf{e} \in \{0,1\}^n$. The noise caused in both amplifiers is additive white Gaussian noise with zero mean, independent of the signals $b(t)$ and $e(t)$. We assume that the electrical noise of the channels is nihil compared to the noise caused in both amplifiers. Finally, Eve has inserted a tamper device which causes additional transmission loss $(L_T)_{dB} = \varepsilon_T(l_A + l_B)$ independent of her signal $e(t)$.

Alice and Bob know S_A, n_B (which they can measure as accurate as they like), l_A, l_B. They only know a probability distribution of the attenuation coefficient $Pr(\alpha)$, and they know $\bar{L_E}$ and $\bar{n_E}$ with $\bar{L_E} \leq L_E$ and $\bar{n_E} \leq n_E$. In the worst case for Alice and Bob $\bar{L_E} = L_E$ and $\bar{n_E} = n_E$. The tamper device introduces additional transmission loss L_T. Since, the exact value of the transmission loss over the main channel is unknown L_T is unknown. Alice and Bob can only obtain statistical information about L_T (as we shall see).

The signal power of $b(t)$ is $S_A g_B / L_A L_T L_B$, and the corresponding noise power is $n_B g_B$. Hence, the signal to noise ratio of the main channel equals $(S/N)_{AB} = S_A/L_A L_T L_B n_B = S_A/(n_B 10^{(\alpha+\varepsilon_T)(l_A+l_B)/10})$. Alice and Bob view this as a function of $\alpha + \varepsilon_T$. Thus the channel from Alice to Bob with input $\mathbf{a}$ and output $\mathbf{b}$ is a $BSC(p_{AB}(\alpha + \varepsilon_T))$ with $p_{AB}(\alpha + \varepsilon_T) = Q(\sqrt{(S/N)_{AB}})$, that is a binary symmetric channel with cross-over probability $Q(\sqrt{(S/N)_{AB}})$, where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\lambda^2/2} d\lambda$.

Suppose that prior to the secret key generation Alice transmits m zero's to Bob. Let the random variable $k(m)$ be the number of 1s Bob receives over the main channel. Let $P(x) = Pr(\alpha \geq x)$. Then we can derive

$$Pr\left(\left|p_{AB}(\alpha+\varepsilon_T) - \frac{k(m)}{m}\right| \leq \varepsilon, \alpha \geq x\right) \geq \left(1 - \frac{1}{4m\varepsilon^2}\right) P(x).$$

Hence, for m large enough Alice and Bob may approximate $p_{AB}(\alpha + \varepsilon_T)$ by $k(m)/m$, which leads to an approximation $\alpha(k(m)/m)$ of $\alpha + \varepsilon_T$ ($\alpha = p_{AB}^{-1}$). More precisely

$$Pr\left(\left|\alpha + \varepsilon_T - \alpha\left(\frac{k(m)}{m}\right)\right| \leq \varepsilon, \alpha \geq x\right) \geq \left(1 - \frac{1}{4m\delta(\varepsilon)^2}\right) P(x).$$

The signal power of $e(t)$ is $S_A g_E / L_A L_E$, and the corresponding noise power is $n_E g_E$. Hence, the signal to noise ratio of the channel from Alice to Eve equals $(S/N)_{AE} = S_A/L_A L_E n_E \leq S_A/10^{\alpha l_A/10} \bar{L_E} \bar{n_E}$. Thus the channel from Alice to Eve with input $\mathbf{a}$ and output $\mathbf{e}$ is a $BSC(p_{AE})$ with $p_{AE} = Q(\sqrt{(S/N)_{AB}}) \geq Q(\sqrt{S_A/10^{\alpha l_A/10} \bar{L_E} \bar{n_E}}) = \bar{p}_{AE}(\alpha)$.

We notice that all noise is generated in both amplifiers. Hence, the BCCT is equivalent to the binary symmetric BCC, in which the main channel is a $BSC(p_{AB}(\alpha + \varepsilon_T))$ and the wire-tap channel is a $BSC(p_{AE})$ from Alice to Eve. With probability at least $(1 - 1/4m\delta(\varepsilon)^2)P(x)$ the worst-case scenario for Alice and Bob is a $BSC(p_{AB}(\alpha(k(m)/m) + \varepsilon))$ as main channel and a $BSC(\bar{p}_{AE}(x))$ as wire-tap channel. The secrecy capacity of the worst-case scenario is equal to $Cs(x, k(m), m, \varepsilon) = h(\bar{p}_{AE}(x)) - h(p_{AB}(\alpha(k(m)/m) + \varepsilon)$ [1]. We notice that a secret key generated in the worst-case scenario is also secret (for Eve) in the real situation. We conclude that in the BCCT a secret key can be generated with rate at least

$$\sup_{x,\varepsilon,m} (1 - 1/4m\delta(\varepsilon)^2) P(x) \sum_{k \geq 0} Cs(x, k, m, \varepsilon) Pr(k(m) = k).$$

The final conclusion is: "Alice and Bob need to take tampering by Eve into account, which implies that they have to realize that $k(m)/m$ is an approximation of $\alpha + \varepsilon_T$, not of α".

References

[1] Imre Csiszár and János Körner. "Broadcast Channels with Confidential Messages". *IEEE Trans. Inform. Theory*, Vol. IT-24(3):339–348, May 1978.

Ideal Perfect Threshold Schemes and MDS Codes

G.R.Blakley and G.A.Kabatianski[1]
Department of Mathematics, Texas A&M University,
College Station, TX 77843-3368, USA
blakley@math.tamu.edu ; kaba@ippi.ac.msk.su

Abstract — **The two notions in the title coincide.**

I. Introduction

Secret sharing schemes (SSS) made their appearance (see [1], [2]) in the form of threshold (n,τ)-schemes in 1979. R.McEliece and D.Sarwate pointed out [3] a relationship between threshold schemes and MDS-codes in 1981. In 1983 E.Karnin,J.Greene and M.Hellman [4] gave an information-theoretic approach to SSS and proved some upper and lower bounds on the number of participants in an ideal perfect threshold SSS. The proof is based, in fact, on the observation that each ideal perfect threshold SSS determines a unique MDS code, and *vice versa*, when the secret and shadows belong to the same finite field. E.F.Brickell and D.M.Davenport [6] considered *combinatorial* ideal perfect SSS for the general access structure and established the relationship between such schemes and matroids. From their results the equivalence of combinatorial ideal perfect threshold SSS and MDS codes (*i.e. orthogonal arrays* $OA_1(\tau, n+1, q)$) follows almost immediately. In this paper we give an independent, self contained proof (following the ideas in [4]) for the (formally) more general information-theoretic definition of ideal SSS.

II. Definitions and a useful lemma

Let $\mathcal{S}_0;\mathcal{S}_1\ldots,\mathcal{S}_n$ be finite sets used by an SSS dealer as alphabets. $\mathcal{S}_0$ is for the secret, and other $\mathcal{S}_i$ for shares. We call a point $s=(s_0,s_1,\ldots,s_n)\in\mathcal{S}=\mathcal{S}_0\times\cdots\times\mathcal{S}_n$ a sharing rule. Any SSS can be defined as a probability distribution $P(s)$ on $\mathcal{S}$, which the dealer uses for generating sharing rules, *i.e.* for choosing a secret s_0 and giving a corresponding share s_i to the i-th participant.

Let Γ be an access structure , *i.e.* a set of subsets of $\{1,\ldots,n\}$ with the monotonic property ($A\in\Gamma, A\subset B$ imply $B\in\Gamma$). Consider $S_0,\ldots,S_n$ to be random variables with P as their mutual distribution. We call a pair (P,S) a perfect SSS, realizing an access structure Γ if (see [4], [5]) $H(S_0 \mid S_i, i\in A)=0$ or $H(S_0)$ according as $A\in\Gamma$ or not.

Denote by $\Gamma_{\min}$ the set of minimal subsets of Γ. The following lemma (see [5]) is very useful

Lemma 1 *$H(S_j \mid S_i, i\in A\backslash\{j\})\geq H(S_0)$ for any $A\in\Gamma_{\min}$ and any $j\in A$.*

Corollary 1 *$H(S_i, i\in A)\geq \mid A\mid \cdot H(S_0)$ for any $A\in\Gamma_{\min}$.*

III. An equivalence involving the combinatorial definition

We call $V=\{s\in S \mid P(s)>0\}$ the "code" of the SSS (P,S). Let $q=|S_0|$. Let us note that if the pair (P,S) perfectly realizes an SSS for the access structure Γ for some distribution $p(s_0)$ on secrets, then any distribution on $\mathcal{S}_0$ can be perfectly realized by the same S, the code V and an appropriate choice of P. From this remark and Corollary 1 one can show that the following are true.

Lemma 2 *$|V|\geq q^{|A|}$ for any perfect SSS and any $A\in\Gamma_{\min}$.*

Corollary 2 *For any perfect (n,τ)-threshold SSS the cardinality of its code satisfies the inequality $|V|\geq q^{\tau}$.*

We will distinguish between two definitions of ideal SSS. The *combinatorial* definition of an ideal perfect SSS is that $|\mathcal{S}_0|=|\mathcal{S}_i|$ for all i. A (formally) weaker information-theoretic (IT) definition is that $H(S_i)\leq H(S_0)$ for all i. The following corollary of Lemma 1 (see [4]) shows that the set V is a code with minimal Hamming distance $d(V)\geq n-\tau+2$.

Corollary 3 *$H(S_j \mid S_{i_1},\ldots,S_{i_\tau})=0$ for any IT-ideal perfect (n,τ)-threshold SSS and any distinct $j,i_1,\ldots,i_\tau\in\{0,1,\ldots,n\}$,* i.e. *$S_j$ is a function of $S_{i_1},\ldots,S_{i_\tau}$.*

Hence, for the *combinatorial* definition of an *ideal* perfect SSS, Corollaries 2 and 3 ensure that the code V of an ideal perfect (n,τ)-threshold SSS is a q-ary code of length $n+1$, distance $d(V)\geq n-\tau+2$ and cardinality $|V|\geq q^{\tau}$. Therefore V is an MDS code with $\mid V\mid = q^{\tau}$ and $d(V)=n-\tau+2$ (the converse, that any MDS code with the above parameters generates an ideal perfect (n,τ)-threshold SSS, is rather obvious).

IV. The main result - an equivalence involving the information-theoretic definition

Now we can prove that the same equivalence is true for IT-ideal SSS also. Denote by V_A the punctured code obtained from V by deleting coordinates "outside of A" (*i.e.* not belonging to A). Corollary 3 states that $|V|=|V_A|$ for any $A: |A|=\tau$. On the other hand, the random variables $S_{i_1}\ldots,S_{i_\tau}$ are mutually independent (see Lemma 1). Therefore, $|V_{i_1,\ldots,i_\tau}|=|\mathcal{S}_{i_1}|\cdots|\mathcal{S}_{i_\tau}|$. Hence, the cardinalities of all sets $\mathcal{S}_i$ are equal to $|\mathcal{S}_0|=q$ and we again have the case of *combinatorial* ideality.

References

[1] G.R.Blakley, "Safeguarding cryptographic keys," *Proceedings of AFIPS 1979 National Computer Conference*. vol. 48, pp. 313–317,1979.

[2] A.Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 1, pp. 612–613, 1979.

[3] R.J.McEliece and D.V.Sarwate, "On secret sharing and Reed-Solomon codes," *Communications of the ACM*, vol. 24, pp. 583–584, 1981.

[4] E.D.Karnin, J.W.Greene, and M.E.Hellman, "On secret sharing systems," *IEEE Transactions on Informatiom Theory* , vol. 29,no. 1, pp. 231–241, 1983.

[5] R.M.Capocelli, A.De Santis, L.Gargano, and U.Vaccaro, "On the size of shares for secret sharing schemes," *Jornal of Cryptology* , vol. 6, pp. 157–167,1993.

[6] E.F.Brickell and D.M.Davenport, "On the classification of ideal secret sharing schemes," *Journal of Cryptology*, vol. 4, pp. 123–134, 1991.

[1]on leave from IPPI,Moscow,Russia

PRODUCT CODES AND PRIVATE-KEY ENCRYPTION

J. Campello de Souza and R. M. Campello de Souza
Communications Research Group - CODEC
Departamento de Eletrônica e Sistemas - UFPE
E-mail ricardo@npd.ufpe.br CP 7800 50732-970, Recife PE, Brasil

Abstract - In this paper the use of product codes cryptographic purposfor es is discussed. The codes are used in a scheme that applies a special type of structured errors that, as far as we know, do not exist in any real communication channel. Although this fact seems of no importance, since the errors in any error-correcting code based cryptosystem are artificially generated at the transmitter, its use allows an improvement in the security level in comparison with similar schemes.

I. SUMMARY

The use of burst-correcting product codes for cryptographic purposes has been recently investigated [1],[2], where a private-key cryptosystem was proposed, based on the fact that the single burst-correcting capacity of a code is, in general, larger than its random error-correcting capacity. In this paper the scheme is revisited and a new class of product codes to implement the cryptosystem is proposed. The key idea of the original scheme has been to use a code which is capable of correcting a special kind of structured errors and then disguise it as a code that is only linear, which makes it unable of correcting the errors as well as their permuted versions. Specifically, in the search for such a structure, the choice for bursts and burst-correcting codes was a natural one in the context of error control codes. However, we observe that the errors structure to be used does not have to necessarily exist in a real communication channel, once that, for cryptographic purposes, they are artificially generated at the transmitter. With that in mind we introduce the following concepts :

Definition 1-The direct mapping of parameters l and s, denoted DM_{ls} (.), is the one that maps the vector $v = (v_1 , ... v_{ls})$ into the matrix V (v_{ij}), of elements $v_{ij} = v_{is+j+1}$, i = 0, 1, ..., l-1, j = 0, 1, ..., s-1.

Definition 2-The vector $e = (e_1 , e_2 , ... e_{ls})$, $e_i \in GF(q)$, is said to be a biseparable error over GF(q), denoted BSE (l,s), if (i) Its components are nonzero distinct elements of GF(q) and (ii) each row and each column of $DM_{ls}(e)$ contains, at most, one nonzero component.

From the above definitions, it can be seen that the maximum weight of a BSE(l,s) over GF(q) is $\omega_{max} = \min (q-1, \min(l,s))$ and the number of BSE's with a given weight ω is (ω=1,2... ω_{max})

$$N_{BSE(l,s)} (\omega) = \binom{q-1}{\omega} \prod_{i=1}^{\omega} (l+1-i)(s+1-i)$$

Proposition- A product code PC (n, k, d) over GF(q), whose constituent row and column codes are, respectively, single parity-check codes C_1 ($N_1 = s+1$, $K_1 = s$, $D_1 = 2$) and C ($N_2 = l+1$, $K_2 = l$, $D_2 = 2$), can correct BSE(l+1,s+1)'s of weights up to ω_{max}.

Denoting by G the generator matrix of PC(n, k, d), the encryption procedure consists of calculating the ciphertext C from the plaintext M, using $C = (MSG + E_{ls,\omega})P$, where $E_{ls,\omega}$ is a BSE (l+1,s+1) of weight ω, P is an n x n permutation matrix and S is a k x k scrambling matrix, used to hide the structure of the matrix GP. The working factors for breaking the system by some of the attacks that are typically applied against cryptosystems based on error control codes, are related with the number of codes in the class defined above, which is $Nc = (l+1)(s+1)(ls)!$. Using G' = SGP, the cryptanalyst must find, among all Nc matrices, one of the (l+1)! (s+1)! matrices that can be used to decode the corrupted received vector (ciphertext). That means a working factor of (l s) ! / l ! s! , which compairs favourably with the results obtained by the previous scheme.

ACKNOWLEDGEMENTS

This work received support from Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, under Grant 304248/84.5-EE.

REFERENCES

1. F.M. Alencar, A.M. Léo and R.M. Campello de Souza, Private-Key Burst Correcting Code Encryption, Proceedings of the IEEE Int. Symp. on Info. Theory, pp. 227, Jan. 1993.
2. R.M. Campello de Souza and J. Campello de Souza, Array Codes for Private-Key Encryption, Electronics Letters, Vol. 30, No. 17, pp. 1394-1396, Aug. 1994.

On Interval Linear Complexity of Binary Sequences

V.B.Balakirsky

Data Security Association "Confident", 193060 St.-Petersburg, Russia

Abstract — **We consider the problem of partial approximation of binary sequences by the outputs of linear feedback shift registers. A generalization of the linear complexity profiles of binary sequences leads to a sequence that is regarded as the profile of interval linear complexity. Some properties of this sequence are examined.**

I. Introduction

A widely used criterion of the linear complexity of binary sequences was introduced by R.Rueppel [1]. In accordance with this criterion, we construct an integer-valued sequence, called the 'linear complexity profile' (LCP), whose j-th component, L_j, is the shortest length of a linear feedback shift register (LFSR) generating the first j bits of our binary sequence. The component L_j can be found using Berlekamp-Massey algorithm [2], and pseudo-random sequences possess an LCP with $L_j \approx j/2$ for all j. However, there are many examples when the deviations of the LCP from the sequence $j/2$, $j = 1, 2, ...$ do not characterize the 'randomness' of binary sequences. For example, let us suppose that we are given a sequence $\mathbf{u}^*01^\infty$, where $\mathbf{u}^*$ is a sequence of length n having a 'good' LCP. Then $\mathbf{u}^*$ is generated by an LFSR of length $\approx n/2$. Nevertheless, the final result of Berlekamp-Massey algorithm, applied to the whole sequence, is the LFSR of length $n+2$, and the LCP is as good as before up to the $\approx 2n$-th component. It is easy to see that if we construct the LCP of the sequence starting at the $n+1$-st bit then the conclusion would be different. Thus, to extend the Rueppel's approach we need to construct the LCPs for all subsequences of the input sequence and select the worst one, whose deviations from the line $j/2$ should be used as a measure of complexity. Such a procedure seems to be rather complicated.

In this paper, we introduce a new measure of complexity, called an interval linear complexity. For all $L < m$, where m is a given positive integer, we find all the fragments of the binary sequence that have length $L+m$ and can be generated by an LFSR of length L. The number of such fragments and their lengths contain information on LCPs for all starting positions, and the results of analysis can be useful for different methods based on linear approximations.

II. Some properties of the m-interval linear complexity

Let $\mathbf{u} = u_1, u_2, ...$ be a binary sequence. We assume that $u_1 = 1$ and $u_i = 0$ for $i = 0, -1, ...$ For all $k > 0$, we set $\mathbf{u}_j^{(k)} = (u_{j-k+1}, ..., u_j)$ and write $\mathbf{u}_j^{(k)} \prec \mathcal{F}(L)$ iff there exists a binary vector $(a_1, ..., a_L)$ such that $u_t = a_1 \cdot u_{t-1} + ... + a_L \cdot u_{t-L}$ for all $t \in \{j-k+1, ..., j\}$. Furthermore, we write $\mathbf{u}_j^{(k)} \not\prec \mathcal{F}(L')$ iff $u_t \neq b_1 \cdot u_{t-1} + ... + b_{L'} \cdot u_{t-L'}$ for at least one $t \in \{j-k+1, ..., j\}$, where $(b_1, ..., b_{L'})$ is any binary vector.

Let us fix $j > m$ and define $L_j^{(m)}$ as the shortest length of an LFSR, generating the fragment $\mathbf{u}_j^{(m)}$ provided that the subsequence $\mathbf{u}_{j-m}^{(L)}$, where $L = L_j^{(m)}$, forms the initial content of the shift register, i.e., (a) $\mathbf{u}_j^{(m)} \prec \mathcal{F}(L_j^{(m)})$; (b) if $L' < L_j^{(m)}$, then $\mathbf{u}_j^{(m)} \not\prec \mathcal{F}(L')$. Using conventional notations [2], we claim that $L_j^{(m)} = L$ iff L is the shortest length of an LFSR generating the subsequence $u_{j-m-L+1}, ..., u_j$, and all the subsequences $u_{j-m-L'+1}, ..., u_j$, where $L' < L$, cannot be generated by an LFSR of length L'. The parameter $L_j^{(m)}$ will be referred to as the *m-Interval Linear Complexity* (m-ILC) of $\mathbf{u}$ at position j, and the sequence $L_{m+1}^{(m)}, L_{m+2}^{(m)}, ...$ will be regarded as the profile of the m-ILC of $\mathbf{u}$. Some properties of the m-ILC are detailed below.

Theorem.

1. Let L_{ij} be the shortest length of an LFSR generating $u_i, ..., u_j$. Then
$$L_j^{(m)} = \min_{i:\, i+L_{ij} \leq j-m+1} L_{ij}.$$
2. If $L_j^{(m)} = L \leq m$, then there is exactly one LFSR of length L generating $u_{j-m-L+1}, ..., u_j$.
3. If
$$\begin{cases} L_{j-1}^{(m)} \neq L_j^{(m)}, \\ L_j^{(m)} = ... = L_{j+l-1}^{(m)} = L < m, \\ L_{j+l}^{(m)} \neq L_{j+l-1}^{(m)} \end{cases}$$
then
$$\begin{cases} L_{j-1}^{(m)} \geq m, \\ L_{j+l+\Delta l}^{(m)} = m + l \text{ for all } \Delta l = 0, ..., m-1-L, \\ L_{j+l+m-L}^{(m)} \leq m. \end{cases}$$

The theorem claims that the profiles of the m-ILC have very regular structure. If the current element of the profile, $L_{j-1}^{(m)}$, is greater than m then the next element, $L_j^{(m)}$, can be less than m, i.e., the profile 'falls into the pit'. In this case, the profile can stay in the pit for l times or jumps at the level $m+l$ and stays at this level for $m - L_j^{(m)}$ times. The parameters l and $m - L_j^{(m)}$ can be interpreted as the 'length' and the 'depth' of the pit, and the duality between them takes place.

Such a behaviour gives an opportunity to realize an interval atack on the stream cipher when the cipher is constructed using some complex scheme, but an eavesdropper approximates its fragments by LFSRs of length $< m$. Suppose that the eavesdropper has some set of the key words and assumes that they are written in the plain text. If he is right and the position of one of these words corresponds to a pit in the profile of the m-ILC, then he reconstructs the LFSR and reads all the other words of the plain text while the corresponding elements of the profile belong to this pit.

References

[1] R. A. Rueppel, "New approaches to stream ciphers," Ph.D.dissertation, Swiss Federal Institute of Technology, 1984.

[2] J.L.Massey, "Shift register synthesis and BCH decoding," *IEEE Trans.Inform.Theory*, vol.15, pp.122-127, Jan. 1969.

The permutation group of affine-invariant codes

Thierry P. BERGER and Pascale CHARPIN
INRIA-Rocquencourt, France

Abstract — **Affine-invariant codes are primitive cyclic codes whose extension is invariant under the affine-group. We present the formal expression of the permutation group of these codes. We after give several tools in order to determine effectively the permutation groups. Our main application is the permutation group of primitive narrow-sense BCH-codes defined on any prime field.**

I. THE FORMAL EXPRESSION

The reader can refer to [2, 4] for the definition of affine-invariant codes and their description by antichains. The permutations of coordinate places which send a code C into itself form the *permutation group* of C, denoted by $Per(C)$; when the code is binary, this group is actually the *automorphism group* of C, usually denoted by $Aut(C)$.

Let G be the finite field of order p^m and let $\mathbf{k}$ be a subfield of G. We denote by $AGL(m,p)$ the affine group of G over $GF(p)$ and, for any divisor e of m, by $AGL(m/e,p^e)$ the affine group of G over $GF(p^e)$. The corresponding semi-affine group is denoted by $A\Gamma L(m/e,p^e)$. We consider cyclic codes C of length p^m-1 over $\mathbf{k}$. The extended code $\widehat{C}$ is said to be an affine-invariant code if and only if its permutation group contains $AGL(1,p^m)$. Affine-invaraint codes form a class including codes of great interest as BCH-codes or Reed-Muller (RM) codes (and generalized RM-codes). BERGER has recently proved that the permutation group of an affine-invariant code is contained in $AGL(m,p)$ [1]. Then a formal expression of the permutation group of any affine-invariant code can be deduced:

Theorem 1 *Denote by θ_k the kth-power of the Frobenius mapping on G. Let $\widehat{C}$ be a non trivial affine-invariant code; let ℓ be the smallest integer dividing m such that θ_ℓ leaves $\widehat{C}$ invariant. Then there is a divisor e of m such that $Per(\widehat{C})$ is generated by $AGL(m/e,p^e)$ and θ_ℓ - respectively $Per(C)$ is generated by $GL(m/e,p^e)$ and θ_ℓ.*

II. TO DETERMINE THE PERMUTATION GROUPS

For a large part of affine-invariant codes, mainly when m is a prime, the permutation group is completely determined by applying Theorem 1. The problems appear when m has no trivial divisors.

Let $S=[0,p^m-1]$ and let α be a primitive root of G. We call *defining set* of $\widehat{C}$ the subset T of S consisting of 0 and of the s such that α^s is a zero of C. Let e be a divisor of m and $v=p^e$. We identify any $s\in S$ with its v-ary expansion $(s_0,\dots s_{m/e-1})$. The v-weight of s is $\omega_v(s)=\sum_{i=0}^{m/e-1}s_i$. Then we can define the poset $(S,\ll_e)$: s and t in S satisfy $s\ll_e t$ iff $\omega_v(p^k s)\le\omega_v(p^k t)$, for all k in $[0,e-1]$. In terms of partial order the condition of DELSARTE, given in [3], becomes:

Theorem 2 *Assume that $\widehat{C}$ is affine-invariant. Then $\widehat{C}$ is invariant under $AGL(m/e,p^e)$ iff its defining set T satisfies:*

$$t\in T \text{ and } s\ll_e t \;\Rightarrow\; s\in T\ .$$

We give two conditions equivalent to DELSARTE's condition, providing new tools for the study of infinite classes of codes. The first one is derived from the result of DELSARTE, by using the description of affine-invariant codes by antichains. The second one comes from the polynomial representation of permutations:

Theorem 3 *The code $\widehat{C}$ is invariant under $AGL(m/e,p^e)$ iff its defining set T satisfies:*

$$t\in T \text{ and } j\ll_m t \;\Rightarrow\; t+j(p^e-1)\in T\ .$$

III. THE p-ARY BCH-CODES

Theorem 4 *Denote by $B(d)$, the BCH-code of designed distance d and length p^m-1 over $GF(p)$, and by $\widehat{B}(d)$ the extended code. Suppose that $\widehat{B}(d)$ is not trivial, i.e. $d\notin\{1,p^m-1\}$ (in the trivial case $Per(\widehat{B}(d))$ is the symmetric group). Then the permutation group of $\widehat{B}(d)$ is the semi-affine group $A\Gamma L(1,p^m)$, except for the following cases.*

- *When $p=2$, we have three kinds of exception:*
  1. *If $d\in\{3,2^{m-1}-1\}$, for any m, or $d=7$ for $m=5$, then $Aut(\widehat{B}(d))$ is $AGL(m,2)$; whence $\widehat{B}(d)$ is a Reed-Muller code.*
  2. *If $d=2^{m-1}-2^{(m-2)/2}-1$, for m even, then $Aut(\widehat{B}(d))=A\Gamma L(2,2^{m/2})$.*
  3. *If $m=6$, then $Aut(\widehat{B}(7))=A\Gamma L(2,2^3)$ and $Aut(\widehat{B}(15))=A\Gamma L(3,2^2)$.*
- *For p odd, the only exceptions are whenever $\widehat{B}(d)$ is a p-ary Reed-Muller code. That is : $d\in\{2,p^{m-1}(p-1)-1\}$, for any m; $d=p^2-2p-1$, for $m=2$ and $p>3$; $d=5$ for $m=3$ and $p=3$. In these cases $Per(\widehat{B}(d))=AGL(m,p)$.*

Note that $Per(B(d))$ is the linear group $GL()$ (or the semi-linear group $\Gamma L()$), when $Per(\widehat{B}(d))$ is $AGL()$ (or $A\Gamma L()$).

REFERENCES

[1] T. BERGER, *On the Automorphism Groups of Affine-Invariant codes*, Designs, Codes and Cryptography, to appear.

[2] P. CHARPIN, *Codes cycliques étendus affines-invariants et antichaines d'un ensemble partiellement ordonné*, Discrete Mathematics 80 (1990), 229-247.

[3] P. DELSARTE *On cyclic codes that are invariant under the general linear group*, IEEE Trans. on Info. Theory, vol. IT-16, n.6, 1970.

[4] T. KASAMI, S. LIN & W.W. PETERSON *Some results on cyclic codes which are invariant under the affine group and their applications*, Info. and Control, vol. 11, pp. 475-496 (1967).

[5] C.C. LU & L.R. WELCH, *On automorphism groups of binary primitive BCH codes*, Proceedings 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, p.51.

Mixed-Rate Multiuser Codes for the T-User Binary Adder Channel

A. Brinton Cooper III
Information Science and Technology Directorate, Army Research Laboratory, APG, MD 21005-5067 USA
Brian L. Hughes[1]
Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA

Abstract — **Coding schemes for the T-user binary adder channel are investigated. Recursive constructions are given for two families of mixed-rate, multiuser codes. These basic codes can be combined by time-sharing to yield codes approaching most rates in the T-user capacity region. The best codes constructed herein achieve a rate sum, $R_1 + \cdots + R_T$, which is higher than all previously reported codes for $T \geq 4$ and is within 0.519 bits/channel use of the information-theoretic limit.**

Summary

One of the most extensively investigated multiple-access channels is the *binary adder channel,* described as follows. T users communicate with a single receiver through a common discrete-time channel. At each time epoch, user i selects an input $X_i \in \{0, 1\}$ for transmission. The channel output is

$$Y \triangleq \sum_{i=1}^{T} X_i \tag{1}$$

where summation is over the real numbers. We assume that there is no feedback and all users are synchronized.

Chang and Weldon [1] showed that the capacity region of the T-user binary adder channel is the set of all nonnegative rates $(R_1, \ldots, R_T)$ satisfying

$$\begin{aligned} 0 &\leq R_i \leq H_1 , \\ 0 &\leq R_i + R_j \leq H_2 , \\ &\vdots \quad\quad \vdots \quad\quad \vdots \\ 0 &\leq R_1 + \cdots + R_T \leq H_T , \end{aligned} \tag{2}$$

where

$$H_m \triangleq -\sum_{i=0}^{m} \binom{m}{i} 2^{-m} \log_2 \binom{m}{i} 2^{-m} . \tag{3}$$

In particular, observe that the largest achievable sum-rate, $R_{sum}(T) \triangleq R_1 + \cdots + R_T$, is $C_{sum}(T) \triangleq H_T$, which is called the *sum-capacity.*

Chang and Weldon [1] also presented a family of multiuser codes which are asymptotically optimal in the sense that $R_{sum}(T)/C_{sum}(T) \to 1$ as $T \to +\infty$. In their construction, each user's code consists of only two codewords which are defined recursively (so $R_1 = R_2 = \cdots = R_T$). This basic construction has been generalized in several ways [2, 3, 5, 7], and alternate constructions have been proposed based on coin weighing designs [6] and additive number theory [4].

Chang and Weldon's construction shows how to approach one point on the boundary of the T-user capacity region. Similarly, all subsequent work for $T > 2$ has focused on the symmetric rate case, except for [5] where $R_1 = R_2 = \cdots = R_{T-1}$ but $R_T > R_1$. It is natural to ask, however, whether other points in the capacity region can be approached by a similar construction.

This talk will present two mixed-rate, multiuser code constructions for the binary adder channel. The codewords contained in these codes are equivalent, up to an affine transformation, to those in [1] and [6]; however, the recursions are adapted in order to distribute these codewords among *as few users as possible.* As a result, we obtain codes with a wide range of information rates. In particular, we show that these basic codes can be combined to approach all rates in the polytope

$$\begin{aligned} 0 &\leq R_i \leq H_1 - \epsilon_1 , \\ 0 &\leq R_i + R_j \leq H_2 - \epsilon_2 , \\ &\vdots \quad\quad \vdots \quad\quad \vdots \\ 0 &\leq R_1 + \cdots + R_T \leq H_T - \epsilon_T , \end{aligned} \tag{4}$$

where $0 \leq \epsilon_m < 1.090$ bits/channel use, $1 \leq m \leq T$. Moreover, we construct a family of T-user codes with $R_{sum}(T) \geq C_{sum}(T) - 0.519$ bits/channel use, which exceeds the sum-rate of all codes previously reported in [1, 2, 3, 4, 5, 6, 7] for $T \geq 4$. Extensions to a T-user, Q-frequency adder channel are also discussed.

[1]B. L. Hughes was supported by the National Science Foundation under grant NCR-9217457, and by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant DAAL03-89-K-0130.

References

[1] S. C. Chang and E. J. Weldon, Jr, "Coding for T-user multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 25, pp. 684-691, Nov. 1979.

[2] S. C. Chang, "Further results on coding for T-user multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 30, pp. 411-415, Mar. 1984.

[3] T. J. Ferguson, "Generalized T-user codes for multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 28, pp. 775-778, Sept. 1982.

[4] D. B. Jevtić, "Disjoint uniquely decodable codebooks for noiseless synchronized multiple-access adder channels generated by integer sets," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1142-1146, May 1992.

[5] G. K. Khachatrian and S. S. Martirossian, "Codes for T-user noiseless adder channel," *Prob. Contr. Inform. Theory*, vol. 16, pp. 187-192, Mar. 1987.

[6] S. S. Martirossian and G. K. Khachatrian, "Construction of signature codes and the coin weighing problem," *Prob. Inform. Trans.*, vol. 25, pp. 334-335, Oct.-Dec. 1989.

[7] J. H. Wilson, "Error-correcting codes for a T-user binary adder channel," *IEEE Trans. Inform. Theory*, vol. 34, pp. 888-890, July 1988.

New Quaternary Linear Codes of Dimension 5[1]

T. Aaron Gulliver

Dept. of Systems & Computer Eng., Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6

Abstract — **In this paper, new (pm,m) and (pm,m-1) quaternary linear codes of dimension 5 are presented. These codes belong to the class of quasi-twisted codes.**

I. Introduction

A fundamental and challenging problem in coding theory is to find a linear (n,k) code over GF(q) achieving the maximum possible minimum Hamming distance. This value is denoted as $d_q(n,k)$, and linear codes which have a minimum distance equal to $d_q(n,k)$ are called *optimal*. For $q=4$, $d_q(n,k)$ has been determined for $k \leq 3$ and all but 10 values of d for $k=4$[1]. Many values of $d_4(n,5)$ have been established, and Brouwer [2] maintains an up to date table of upper and lower bounds for $k \leq n \leq 132$. In this paper several values of $d_4(n,5)$ are determined.

II. Quasi-Twisted Codes

The class of quasi-twisted (QT) codes is a generalization of the class of quasi-cyclic (QC) codes over GF(q), $q>2$[4]. A code is called quasi-twisted if a negacyclic[2] shift of a codeword by p positions results in another codeword. The blocklength, n, of a QT code is a multiple of p, so that $n=mp$. Many QT codes codes can be constructed from $m \times m$ twistulant matrices (with a suitable permutation of coordinates). In this case, the generator matrix, G, can be represented as,

$$G = [B_1,\ B_2,\ \ldots,\ B_p] \tag{1}$$

where the B_i are $m \times m$ *twistulant* matrices of the form

$$B = \begin{bmatrix} b_0 & b_1 & b_2 & \cdots & b_{m-2} & b_{m-1} \\ \alpha b_{m-1} & b_0 & b_1 & \cdots & b_{m-3} & b_{m-2} \\ \alpha b_{m-2} & \alpha b_{m-1} & b_0 & b_{m-4} & \cdots & b_{m-3} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \alpha b_1 & \alpha b_2 & \alpha b_3 & \cdots & \alpha b_{m-1} & b_0 \end{bmatrix} \tag{2}$$

and $\alpha \in$GF(q)$\backslash\{0\}$. The algebra of $m \times m$ twistulant matrices over GF(q) is isomorphic to the algebra of polynomials in the ring GF(q)$[x]/x^m-\alpha$ if B_i is mapped onto the polynomial $b_i(x)$ formed from the entries in the first row of B_i. The $b_i(x)$ are called *defining polynomials*.

The 1-generator QC codes[6] can be generalized to 1-generator QT codes. The *order* of a 1-generator QT code, V, is defined as

$$h(x) = \frac{x^m-\alpha}{\gcd\{x^m-1, c_0(x), c_1(x), \cdots, c_{p-1}(x)\}}, \tag{3}$$

where $\alpha \in$GF(q)$\backslash\{0\}$, and k, the dimension of V, is equal to the degree of $h(x)$. If $h(x)$ has degree m, (1) is a generator matrix for V. If $\deg(h(x))=k<m$, a generator matrix for V can be constructed by deleting $m-k$ rows of (1). Codes with $k=5$ and $m=5$ and 6 are considered here.

A search for good QT codes requires a representative set of defining polynomials[4] which can be enumerated using Burnside's Lemma[5]. For $q=4$ and $m=5$, there are 70 for all values of α. However, since $4 \not| m$, the quaternary QT codes with $\alpha \neq 1$ are not equivalent to QC codes[4]. The results of a greedy local search are given in the next section.

III. Construction Results

In addition to establishing many lower bounds on $d_4(n,5)$, the following new optimal codes (based on the bounds in [2] and the Griesmer bound) were found. A (50,5) code with $d=35$, A (105,5) code with $d=76$, a (110,5) code with weight distribution

Weight	0	80	84	88	92
Count	1	618	225	105	75

a (115,5) code with $d=80$, a (120,5) code with distribution

Weight	0	88	96	120
Count	1	765	255	3

a (126,5) code with $d=92$, a (132,5) code with $d=96$, a (205,5) code with weight distribution

Weight	0	152	156	160	168
Count	1	810	120	4	90

and a (216,5) code with weight distribution

Weight	0	160	168	176	192
Count	1	792	192	36	3

IV. Summary

The construction of quasi-twisted (QT) codes over GF(4) has been presented. Many of the codes constructed have a minimum distance which establishes a lower bound on the maximum minimum distance. The new codes include several optimal codes which determine $d_4(n,5)$ for $n=50, 105, 110, 115, 120, 126, 132, 205$ and 216.

References

[1] P.P. Greenough and R. Hill, "Optimal linear codes over GF(4)," *Disc. Math.*, vol. 125, pp. 187-199, 1994.

[2] A.E. Brouwer, Table of minimum-distance bounds for linear codes over GF(4), lincodbd server, aeb@cwi.nl, Eindhoven University of Technology, Eindhoven, the Netherlands.

[3] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, New York: North-Holland, 1977.

[4] R. Hill and P.P. Greenough, "Optimal quasi-twisted codes," Proc. *Int. Workshop Algebraic and Comb. Coding Theory*, Voneshta Voda, Bulgaria, pp. 92-97, June 1992.

[5] T.A. Gulliver, "New optimal ternary linear codes," *IEEE Trans. Inf. Theory*, July 1994.

[6] G.E. Séguin and G. Drolet, "The theory of 1-generator quasi-cyclic codes," preprint, Royal Military College of Canada, Kingston, ON, 1991.

[1] This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.

[2] A negacyclic shift of an m-tuple $(x_0, x_1, \ldots, x_{m-1})$ is the m-tuple $(\alpha x_{m-1}, x_0, \ldots, x_{m-2}), \alpha \in$GF($q$)$\backslash\{0\}$.

Hecke Modules as Linear Block Codes and Block m-PSK Modulation Codes

Karl-Heinz Zimmermann
Mathematical Institute, University of Bayreuth
95440 Bayreuth, Germany

Abstract — **We discuss the error correction capabilities of a class of Hecke modules as linear codes and free linear block m-PSK modulation codes.**

We provide an introduction to the study of modules for a Hecke algebra (of type A) as linear codes for the Hamming and the Euclidean metric. These modules are called Hecke modules and play an important role in another branch of mathematics, representation theory of groups and algebras [1].

We first introduce a class of Hecke modules in a purely combinatorial manner. In particular, we provide a basis for each of these modules which can be easily calculated by a computer [2].

These Hecke modules are very interesting from the point of view of coding theory. For this, note that these Hecke modules can be defined as vector spaces over any field and so may be considered as linear block codes [2],[3]. In particular, the primitive generalized Reed-Muller codes over the primes as well as shortened versions of them and the Simplex codes emerge as subclasses of our Hecke modules in a very natural way. We review Hecke modules whose coding parameters are known as the Specht modules and several one-step majority logic decodable codes [4]. Then we consider so-called characteristic Hecke modules. A characteristic Hecke module is a free $\mathbb{Z}$-module yielding a linear code over $\mathrm{GF}(p)$ by reducing the coefficients of all linear combinations of its generating elements modulo p so that the parameters n, k and d of the code are independent of the choice of p. For instance, binary Reed-Muller codes and Simplex codes emerge in this way but not generalized Reed-Muller codes.

Moreover, these Hecke modules can be considered as free modules over the ring $\mathbb{Z}_m$ or one of its extension rings and therefore represent free linear block m-PSK modulation codes [5]. We have calculated the minimum squared Euclidean distance of the Hecke modules over $\mathbb{Z}_m$, m a prime, resulting from the previously discussed characteristic Hecke modules. Furthermore, we give a list of Hecke modules of length $n = 6, \ldots, 16$ over $\mathbb{Z}_8$ with a good minimum squared Euclidean distance calculated by an exhaustive computer search. Finally we compare the resulting codes with further classes of block m-PSK modulation codes such as cyclic codes over $\mathbb{Z}_m$ and multilevel codes. It will turn out that at least for short length, the minimum squared Euclidean distance of our codes is as good as the one of the best unrestricted modulation codes.

References

[1] C.W. Curtis and I. Reiner, *Methods of Representation Theory I.* New York: Wiley & Sons, 1981.

[2] K.-H. Zimmermann, "On weight spaces of polynomial representations of the general linear group as linear codes", *J. Comb. Theory (A)*, vol. 67, pp. 1-22, 1994.

[3] K.-H. Zimmermann, *Beiträge zur algebraischen Codierungstheorie mittels modularer Darstellungstheorie.* Habilitationsschrift, Bayreuther Mathematische Schriften, vol. 48, 1994.

[4] R.A. Liebler, K.-H. Zimmermann, "Combinatorial S_n-modules as codes", *J. Alg. Comb.*, vol. 4, pp. 47-68, 1995.

[5] K.-H. Zimmermann, "Hecke modules as linear codes for the hamming and euclidean metric", submitted to *IEEE Trans. Inform. Theory.*

REED-SOLOMON GROUP CODES

A. A. Zain, Student member, IEEE and B. Sundar Rajan, Member, IEEE

I. Introduction Reed-Solomon codes over $GF(p^m)$, p a prime and m a positive integer, are cyclic, Maximum Distance Seperable (MDS) and of length p^m-1. The additive group of $GF(p^m)$ is elementary abelian of type $(1,1,\ldots,1)$, isomorphic to a direct product of m cyclic groups of order p, denoted by C_p^m. This paper deals with MDS codes over C_p^m of length p^m-1 which is cyclic and MDS is called a Reed-Solomon group code. In general, a group code over C_p^m need not be a linear code over $GF(p^m)$ as shown in the following example.

Example 1: *Consider length 4, code over* $C_2^2=\{1,x,y,xy\}$ *consisting of the following* 16 *codewords.*

$(1,1,1,1)$	$(1,x,xy,y)$	$(1,y,y,x)$	$(1,xy,x,xy)$
$(x,1,xy,xy)$	$(x,x,1,x)$	(x,y,x,y)	$(x,xy,y,1)$
$(y,1,y,y)$	$(y,x,x,1)$	$(y,y,1,xy)$	(y,xy,xy,x)
$(xy,1,x,x)$	(xy,x,y,xy)	$(xy,y,xy,1)$	$(xy,xy,1,y)$

The Hamming distance of this code is 3 and hence this is a MDS group code.

In [1], it is shown that if C is an $(n,k,n-k+1)$ group code over an abelian group G that is not elementary abelian, then there exists an $(n,k,n-k+1)$ group code over a smaller elementary group G'. In view of these results a natural question that arises is "Are all MDS group codes over C_p^m linear over $GF(p^m)$ as well?" Example 1 shows that is not true, in general. But, if one considers only cyclic and length p^m-1 group codes then it is true. In other words, all Reed-Solomon group codes over C_p^m are conventional linear codes over $GF(p^m)$. This can be shown by extending the well known transform approach for cyclic codes over finite fields [2] to group codes over elementary abelian groups.

II. Transform approach to cyclic codes over elementary abelian groups: Let C_p^m, denote the elementary abelian group isomorphic to direct sum of m cyclic groups of order p each. The ring of endomorphisms of C_p^m, is denoted by $End(C_p^m)$. The set of automorphisms of C_p^m, denoted by $Aut(C_p^m)$, form a group whose order is $p^{\frac{m(m-1)}{2}}\Pi_{i=1}^{m}(p^i-1)$. Among the cyclic subgroups of $Aut(C_p^m)$,the maximal order subgroups have order (p^m-1). The ring $End(C_p^m)$ is isomorphic to $M_m(p)$, the ring of $m\times m$ matrices over $GF(p)$ [3]. This isomorphism gives matrix representation for elements of $End(C_p^m)$. It can be easily seen that, when this matrix representation is used, the groups of nonzero elements of $GF(p^m)$ when represented by their companion matrices [4] corresponding to each irreducible polynomial of degree m coincides with a maximal order cyclic subgroup of $Aut(C_p^m)$. There are other cyclic subgroups of maximal order and one can use them to define transforms which are counterparts of transforms over finite fields.

Definition 1: *For any* C_p^m, *let* Ψ *denote a maximal order cyclic subgroup of* $Aut(C_p^m)$. Ψ *with all zero matrix consitute an elementary abelian group isomorphic to* C_p^m *and considered along with matrix multiplication, form a ring calles a cononical ring of* C_p^m.

For example, the representation of a finite field with a canonical matrix and its powers along with all zero matrix, clearly gives a canonical ring of C_p^m.

Definition 2: *Generalized Discrete Fourier Transform (GDFT)): Let* $\underline{a}=(a_0,a_1,\ldots,a_{n-1})$, *where* $a_i\in C_p^m, i=0,1,\ldots,n-1$, *and* $n=p^m-1$. *The transform vector of* $\underline{a}$, *denoted by* $\underline{A}$, *is defined by*

$$A_j=\otimes_{i=0}^{n-1}\alpha^{ij}(a_i), j=0,1,\ldots,n-1,$$

where α *is a generator of a cyclic subgroup of* $Aut(C_p^m)$ *of order n, and* $\otimes$ *denotes group operation in* C_p^m.

When C_p^m is made $GF(p^m)$ by imposing a multiplication structure with an irreducible polynomial $g(x)$ then all non zero elements of $GF(p^m)$ can be represented by the companion matrix of $g(x)$ and its powers and α in Definition 2 can be replaced by the companion matrix of $g(x)$. Then, Definition 2 coincides on the conventional DFT over $GF(p^m)$, of length p^m-1.

Using the GDFT given in Definition 2 and the properties of $Aut(C_p^m)$ and its matrix representation the following can be proved.

Theorem 1: *Every cyclic and length* p^m-1 *MDS group code is a conventional linear code over* $GF(p^m)$. *In other words, all Reed-Solomon group codes over* C_p^m *are conventional linear codes over* $GF(p^m)$.

The authors are with the Department of Electrical Engineering, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi 110 016, India.
email: bsrajan@ee.iitd.ernet.in

REFERENCES

[1] G. D. Forney, Jr., "On the Hamming distance property of group codes", IEEE Trans. Information Theory, Vol. IT-38,No. 6, pp 1797-1801, Nov. 1992.
[2] R. E. Blahut, Theory and Practice of Error Control Codes, Addison-Wesley, 1979.
[3] B. R. McDonald, Finite Rings with Identity, Marcel-Dekker, 1974.
[4] MacWilliams and Sloane, The Theory of Error-Correcting Codes, North-Holland Pub. Company, 1977.

A New Construction of Nonlinear Unequal Error Protection Codes

Mao-Ching Chiu and Chi-chao Chao

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan 30043, R.O.C.

Abstract — **We propose a new construction of nonlinear unequal error protection (UEP) block codes whose encoding complexity is approximately equivalent to the decoding complexity of a linear block code. Some classes of codes that are better than any linear UEP codes with the same parameters are presented.**

I. Introduction

In the literature studies of UEP block codes were mainly concentrated on linear codes because of easy implementation of encoding and decoding. However, there are nonlinear UEP block codes that are better than any linear ones. In [1][2] a construction of such codes were presented along with examples, which is based on the idea of superimposing codeword clouds originally introduced by Cover. But the drawback of the construction in [1][2] is that there do not appear to be easily implementation methods of encoding. We propose a new construction of nonlinear UEP block codes whose encoding complexity is approximately equivalent to the decoding complexity of a linear block code.

II. Description of Construction

Here for simplicity we only consider two-level UEP codes. Let C be an $(n, k_1 + k_2)$ UEP code for the message space $M_1 \times M_2$, where $M_i = GF(q)^{k_i}$, for $i = 1, 2$. Each message $\mathbf{m}$ can be written as $(\mathbf{m}_1, \mathbf{m}_2)$, where $\mathbf{m}_i \in M_i$, for $i = 1, 2$. Let $\mathbf{c}(\mathbf{m})$ denote the corresponding codeword in C for the message $\mathbf{m}$. The error-correcting capability of a UEP block code is described by its separation vector $\mathbf{s} = (s_1, s_2)$ defined by $s_i = \min\{d(\mathbf{c}(\mathbf{m}), \mathbf{c}(\mathbf{m}')) : \mathbf{m}_i \neq \mathbf{m}_i'\}$, for $i = 1, 2$, where $d(\mathbf{a}, \mathbf{b})$ denotes the Hamming distance between $\mathbf{a}$ and $\mathbf{b}$. Let C_1, C_2, and C_3 be linear codes of block length n and generator matrix G_1, G_2, and G_3, respectively. Define C_{23} to be the code with generator matrix $[G_2^T, G_3^T]^T$. The important message $\mathbf{m}_1$ is encoded to a codeword $\mathbf{c}_1$ in C_1. The less important message $\mathbf{m}_2$ is first encoded to a codeword $\mathbf{c}_2$ in C_2. The codeword $\mathbf{c}_2$ is then decoded by using a complete nearest-neighbor decoder of C_3 and the output codeword denoted by $\mathbf{c}_3(\mathbf{c}_2) \in C_3$ is produced. The codeword $\mathbf{b}$ which carries the less important message $\mathbf{m}_2$ is obtained by $\mathbf{b} = \mathbf{c}_2 - \mathbf{c}_3(\mathbf{c}_2)$. The final transmitted codeword $\mathbf{c} = \mathbf{c}_1 + \mathbf{b}$. Clearly, the overall two-level UEP code $C = C_1 + B$, where B is the set of all $\mathbf{b}$.

Property 1: If all the rows of $[G_2^T, G_3^T]^T$ are linear independent, the encoding mapping from the less important message space M_2 to B is one-to-one.

Let w represent the maximum weight of codewords $\mathbf{b} \in B$. Since all $\mathbf{b}$ are minimum-weight coset leaders of C_3, we have $w \leq \rho$, where ρ is the covering radius of C_3 defined by $\rho = \max\{\min\{|\mathbf{y} - \mathbf{c}| : \mathbf{c} \in C_3\} : \mathbf{y} \in \mathrm{GF}(q)^n\}$. Let d_1 denote the minimum distance of C_1 and d_{23} be the minimum distance of the code C_{23}.

Property 2: $s_1 \geq d_1 - 2w \geq d_1 - 2\rho$.

Property 3: If $d_1 \geq d_{23} + 2w$, $s_2 \geq d_{23}$.

Consider two lower bounds on block length for linear UEP codes. The first bound is a generalization of the well-known Singleton bound: $n \geq s_1 + k_1 + k_2 - 1$. The second one is a generalization of the Griesmer bound: $n \geq \sum_{i=1}^{k_1} \left\lceil \frac{s_1}{q^{i-1}} \right\rceil + \sum_{i=k_1+1}^{k_1+k_2} \left\lceil \frac{s_2}{q^{i-1}} \right\rceil$. Notations n_S and n_G will be used to represent these two lower bounds.

With this new construction, there exist codes which are better than any linear ones. For example, let C_1 be a repetition code of length 24 and C_{23} be a $(24, 23)$ parity check code. We can choose C_2 to be a $(24, 12)$ extended Golay code because the $(24, 12)$ extended Golay code is a subcode of the $(24, 23)$ parity check code. The covering radius of the $(24, 12)$ extended Golay code is 4. Hence this construction gives $k_1 = 1$, $k_2 = 11$, $s_1 \geq 24 - 2 \cdot 4 = 16$, and $s_2 \geq 2$. The bounds give $n_S = n_G \geq 27$. However, our construction only has $n = 24$.

Other new UEP codes can be constructed from BCH codes and Reed-Muller codes. Examples of these codes which are better than any linear ones with the same parameters are given in Tab. 1 and Tab. 2.

Tab. 1: Examples of UEP codes constructed from BCH codes of length $2^m - 1$ which are are better than any linear codes. (NC: no coding, SEC, DEC, TEC: SEC, DEC, TEC BCH code.)

$m \geq$	C_1	C_{23}	C_3	k_1	k_2	$s_1 \geq$	$s_2 \geq$
3	Repetition	NC	SEC	1	m	$2^m - 3$	1
4	Repetition	NC	DEC	1	$2m$	$2^m - 7$	1
5	Repetition	NC	TEC	1	$3m$	$2^m - 11$	1
4	Simplex	NC	SEC	m	m	$2^{m-1} - 2$	1
6	Simplex	NC	DEC	m	$2m$	$2^{m-1} - 6$	1
7	Simplex	NC	TEC	m	$3m$	$2^{m-1} - 10$	1
6	Repetition	SEC	DEC	1	m	$2^m - 7$	3
5	Repetition	SEC	TEC	1	$2m$	$2^m - 11$	3
8	Repetition	DEC	TEC	1	m	$2^m - 11$	5
11	Simplex	SEC	DEC	m	m	$2^{m-1} - 6$	3
10	Simplex	SEC	TEC	m	$2m$	$2^{m-1} - 10$	3
19	Simplex	DEC	TEC	m	m	$2^{m-1} - 10$	5

Tab. 2: Examples of UEP codes constructed from Reed-Muller codes of length 2^m which are better than any linear codes. (The entries for C_1, C_2, and C_3 give the orders of the Reed-Muller codes.)

$m \geq$	C_1	C_{23}	C_3	k_1	k_2	$s_1 \geq$	$s_2 \geq$
5	0	m	m-3	1	$\sum_{i=0}^{2} \binom{m}{i}$	2^{m}-2(m+2)	1
5	0	m-1	m-3	1	$\sum_{i=1}^{2} \binom{m}{i}$	2^{m}-2(m+2)	2
7	0	m-2	m-3	1	$\binom{m}{m-2}$	2^{m}-2(m+2)	4
9	1	m	m-3	m+1	$\sum_{i=0}^{2} \binom{m}{i}$	2^{m-1}-2(m+2)	1
9	1	m-1	m-3	m+1	$\sum_{i=1}^{2} \binom{m}{i}$	2^{m-1}-2(m+2)	2
11	1	m-2	m-3	m+1	$\binom{m}{m-2}$	2^{m-1}-2(m+2)	4
7	0	m	m-4	1	$\sum_{i=0}^{3} \binom{m}{i}$	2^{m}-m^2-3m+12	1
7	0	m-1	m-4	1	$\sum_{i=1}^{3} \binom{m}{i}$	2^{m}-m^2-3m+12	2
8	0	m-2	m-4	1	$\sum_{i=2}^{3} \binom{m}{i}$	2^{m}-m^2-3m+12	4
10	0	m-3	m-4	1	$\binom{m}{3}$	2^{m}-m^2-3m+12	8
14	1	m	m-4	m+1	$\sum_{i=0}^{3} \binom{m}{i}$	2^{m-1}-m^2-3m+12	1
14	1	m-1	m-4	m+1	$\sum_{i=1}^{3} \binom{m}{i}$	2^{m-1}-m^2-3m+12	2
14	1	m-2	m-4	m+1	$\sum_{i=2}^{3} \binom{m}{i}$	2^{m-1}-m^2-3m+12	4

References

[1] E. K. Englund, "Nonlinear unequal error-protection codes are sometimes better than linear ones," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1418–1420, Sep. 1991.

[2] E. K. Englund, "Nonlinear unequal error-protection codes exceeding Katsman bound," *Proc. 1994 IEEE Int. Symp. Inform. Theory*, Trondheim, Norway, p. 502, June 1994.

Linear Code Construction for the 2-User Binary Adder Channel

Hermano A. Cabral and Valdemar C. da Rocha Jr.

Comms. Research Group - CODEC, Dept. of Electronics and Systems - UFPE, 50741-540 Recife PE BRASIL

Abstract — **This paper deals with the construction of a class of binary uniquely decodable code pairs (C_1, C_2) for the two-user binary adder channel (2-BAC), where C_1 is a linear code. The generator matrix G for code C_1 has the property that any of its columns has at most a single 1 among its k elements. These codes are called *strongly orthogonal codes* in the sense that the Hadamard product of any two rows of G is the all-zero n-tuple. The proposed 2-BAC codes achieve the upper bound for the sum rate when the rate of C_1 is greater than or equal to 1/2. Block and bit synchronization is assumed between the users and the receiver.**

I. INTRODUCTION

A code pair (C_1, C_2) is called a linear code for the 2-BAC if either C_1 or C_2 is a linear code. Without loss of essential generality we shall assume henceforth that C_1 is a linear code. Our goal is to start from C_1 and to construct the largest code C_2 such that (C_1, C_2) is uniquely decodable in the 2-BAC. Due to the linearity of code C_1 we can conveniently make use of the standard array decomposition of the set of binary n-tuples into cosets of C_1. The codewords of code C_2 will be chosen from the cosets of C_1. We have shown [6] that the search of codewords for code C_2 in one coset of C_1, say $\boldsymbol{v} \oplus C_1$, can be performed without interfering with future choices of potential, i.e., not yet chosen, codewords for C_2 contained in other cosets. We denote by $A_{\boldsymbol{v}\oplus C_1}$ the set of vectors in the coset $\boldsymbol{v} \oplus C_1$ which are codewords of C_2. We have also shown in [6] that it is possible to simplify the search for codewords for C_2, within a given coset, by decomposing it into disjoint subsets of n-tuples. The decomposition of a coset is neatly done with the use of a subspace of C_1. In order to specify $A_{\boldsymbol{v}\oplus C_1}$ it is convenient to partition $\boldsymbol{v} \oplus C_1$ into disjoint subsets and we thus define the set

$$\mathcal{S}_{\boldsymbol{v}\oplus C_1} = \{\boldsymbol{x_3} \in C_1;\ \boldsymbol{x_3} \cdot (\boldsymbol{x_1} \oplus \boldsymbol{y_2}) = \boldsymbol{0}, \text{ for some } \boldsymbol{x_1} \in C_1\} \subseteq C_1 \tag{1}$$

where $\boldsymbol{y_2} \in \boldsymbol{v} \oplus C_1$. We do not need here to go further with this theory but remark that the objective of our specific code construction in this paper is to guarantee that $\mathcal{S}_{\boldsymbol{v}\oplus C_1}$ is always a subspace (of dimension $l \leq k$) and notice that in general this is not case. We therefore introduce next a class of linear codes for the 2-BAC for which all cosets $\boldsymbol{v} \oplus C_1$ give rise to sets $\mathcal{S}_{\boldsymbol{v}\oplus C_1}$ which are subspaces easily derivable from code C_1.

By a *strongly orthogonal code* we mean a binary linear code C_1 of blocklength n and dimension k, with generator matrix G whose rows $\boldsymbol{c}_i$, $i = 1, 2, ..., k$, have the property that

$$\boldsymbol{c}_i \cdot \boldsymbol{c}_j = (0, 0, ..., 0),$$

$\forall i, j = 1, 2, \ldots, k$, with $i \neq j$.

We define code the pairs (C_1, C_2) for the 2-BAC as strongly orthogonal codes whenever C_1 is a strongly orthogonal code. A strongly orthogonal code is characterized as follows.

II. CODE CONSTRUCTION

Proposition 1: Let C_1 be a binary linear code of blocklength n and dimension k, with generator matrix G. Code C_1 is strongly orthogonal if and only if each column of its generator matrix has at most a single 1 among its k elements.

Without loss of essential generality in the sequel we consider a combinatorially equivalent form of $G = [I_k : g]$, where I_k is the $k \times k$ identity matrix and g is a $k \times (n-k)$ matrix whose i^{th} row $g_i, 0 \leq i \leq k-1$, has a string of l_i consecutive 1's and the remaining coordinates are filled with 0's. If we denote by l_{k+1} the number of all-zero columns of g it follows that $\sum_{i=0}^{k+1} l_i = n - k$. The following theorem establishes the maximum rate $R_{2,\max}$ achievable for code C_2 under the constraint that C_1 is strongly orthogonal.

Proposition 2: Let C_1 be a strongly orthogonal code. The maximum rate $R_{2,\max}$ for a code C_2 such that the pair (C_1, C_2) is uniquely decodable in the 2-BAC is given by

$$R_{2,\max} = \frac{\log\left(\sum_{m=0}^{k}(2^m \times N_m)\right) + l_{k+1}}{k + \sum_{i=1}^{k} l_m + l_{k+1}} \tag{2}$$

where N_m is the number of distinct cosets whose leaders $\boldsymbol{v}$ have exactly m non-zero blocks out of the k blocks $\{\boldsymbol{v}_i\}_1^k$. This number N_m is given by

$$N_m = \sum_{i1=1}^{k} \sum_{i2=i1+1}^{k} \cdots \sum_{i_m=i_{m-1}+1}^{k} (2^{l_{i_1}}-1)\times(2^{l_{i_2}}-1)\times\cdots\times(2^{l_{i_m}}-1), \tag{3}$$

where l_i is the blocklength of $\boldsymbol{v}_i$, $1 \leq i \leq k+1$.

ACKNOWLEDGEMENTS

This work was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under the grant No.304214/77-9.

REFERENCES

[1] E. J. Weldon, Jr., "Coding for a multiple-access channel", *Information and Control*, vol.36, pp.256-274, 1978.

[2] G.H. Khachatrian,"On the construction of codes for noiseless synchronized 2-user channel", *Problems of Control and Information Theory*, vol.11, No.4, pp.319-324, 1982.

[3] J.L. Massey, "On codes for the two-user binary adder channel", Information Theory Meeting, Oberwolfach, Germany, 8^{th} April, 1992.

[4] V.C. da Rocha, Jr. and J.L. Massey, "A new approach to the design of codes for the binary adder channel", in *Cryptography and Coding III* (Ed. M.J. Ganley), IMA Conf. Series, New Series No. 45. Oxford: Clarendon Press, 1993, pp.179-185.

[5] I.F. Blake, "Coding for adder channels", in *Communications and Cryptography* (Eds. R.E. Blahut, D.J. Costello, U. Maurer and T. Mittelholzer), Kluwer Academic Publishers, 1994, pp.49-58.

[6] H. A. Cabral, *Coding for Synchronous Multiple Access Channels*, M.Sc. Thesis, Dept. of Electronics and Systems, Federal University of Pernambuco, Recife, Brasil, 1994. (in Portuguese)

Extending Reed-Solomon codes to modules

Yvo Desmedt*
Dept. EE & CS, Univ. of Wisconsin - Milwaukee, WI 53201, U.S.A.

Abstract — **Desmedt-Frankel (June 1991) presented an erasure code in which the entries of the codewords belong to any Abelian group. We extend this work to error-correction.**

I. Introduction

In Reed-Solomon codes the entries of the generator and the parity check matrix, the message and the codeword vector all belong to a finite field. We discuss a generalization of Reed-Solomon in which the entries of the message tuple and the codeword tuple belong to any Abelian group K. The entries of the generator matrix $\mathbf{G}$ and the parity check matrix $\mathbf{H}$ are similar as in alternant codes but belong to a ring R such that K is an R-module. Clearly R is not necessarily a field.

II. Background

Let $\mathbf{H}$ be the Vandermonde matrix $[v_{h,i}]$, where $v_{h,i} = \alpha_i^h$, $h = 0 \ldots, n-k-1$, $i = 0, \ldots, n-1$ and $\alpha_i \in R$. To guarantee a similar distance as for alternant codes each $(n-k)\times(n-k)$ submatrix should be invertable, which if R is commutative implies that for all i, i' $(i \neq i')$: $\alpha_i - \alpha_{i'}$ are units in R. The following R has, for example, been chosen [1][1]: $R = Z[u] \cong Z[x]/((x^q-1)/(x-1))$, where q is a prime larger than $n-1$. Choosing $\alpha_0 = 0$ and the other $\alpha_i = \sum_{j=0}^{i-1} u^j$ satisfies the requirements. Now, K needs to be replaced by an expanded Abelian group $K' = Z[u] \otimes_Z K$, where $\otimes$ indicates the tensor product of modules (no knowledge of tensor products of modules is required to understand the essence of this text). K' is a $Z[u]$-module. So the entries of c and u belong to K'. Clearly any $k \in K$ maps easily into a $k' \in K'$. This code (to be more precise its dual) was studied in [1] (see also [3]) as an erasure code. The purpose of this paper is to study this code as an error-correcting code.

III. Decoding

Let K' be the R-module where R is a commutative ring. As for extended BCH codes, there exist the following equations between the syndromes:

$$\beta_{j,v} = \sum_{l=0}^{v} \Lambda_l S_{j+v-l} = 0, \quad \text{where} \qquad (1)$$

$$\Lambda(x) = \Lambda_0 + \Lambda_1 x + \ldots + \Lambda_v x^v = \prod_{l=1}^{v} (1 - x\alpha_{i_l}) \qquad (2)$$

and $j = 0, \ldots, n-1-k-v$, i_l is an error location $(0 \leq i_l \leq n-1)$, for all i and i' $(i \neq i')$: $\alpha_i - \alpha_{i'}$ is a unit and the syndrome $S_j \in K'$. Since the syndromes no longer belong to a ring, the Peterson-Gorenstein-Zierler decoder cannot be used. Indeed on K' only an addition is defined and no internal multiplication. This implies that the standard technique to prove that if (1) is satisfied, then there are at maximum v errors in the received word, can no longer be used. Fortunately, one can still prove (details skipped) that if v errors have occurred then for all $v' < v$ some $\beta_{j,v'} \neq 0$ for $0 \leq j \leq v - v' - 1$. Let us discuss decoding of this code in more details.

The obvious decoder for alternant codes is the Berlekamp-Massey algorithm. However, the syndromes are no longer in a finite field, but in a module. So it seems that we need to extend this algorithm. Extensions have been presented, *e.g.* [4]. Unfortunately it is not too difficult to prove that if one could extend Berlekamp-Massey's algorithm to our scenario, then discrete logarithm modulo p and factoring integers would be easy. (Both problems are assumed to be hard.) Let us explain this. Given *any* sequence $(s_0, s_1, \ldots, s_{n-1-k})$ of elements of a finite field, Berlekamp-Massey finds the smallest v and Λ_i such that (1) is satisfied. Now we allow any R-module, and s_i belong to the R-module and $\Lambda_i \in R$. Now take the Z_{p-1}-module $Z_p^0(*)$, p a prime, and define the scalar operation $a \cdot x$ as $x^a \bmod p$, where $a \in Z_{p-1}(+,*)$ and $x \in Z_p^0(*)$. Take $v = n-1-k = 1$, then if Berlekamp-Massey could be extended to any module, it would find Λ_1, where $-\Lambda_1$ is the discrete log of s_1 in base s_0 (if it exists), which is believed to be hard. Worse, replacing the ring Z_{p-1} by $Z_{\phi(m)}$ and the Abelian group $Z_p^0(*)$ by $Z_m^*(*)$ implies that if Berlekamp-Massey could be extended for this module, factoring also would be easy. This discussion easily extends to the expanded K'. So under the assumption that discrete log is hard, Berlekamp-Massey cannot be extended to be used to decode this code. However, when v is small an exhaustive search will allow one to *easily* find the error locations! So far, we have not been able to develop a decoder when v is large.

We conclude by saying that Berlekamp and Massey were lucky that BCH codes were studied over finite fields.

*This research has been partially supported by NSF Grant NCR-9106327.

[1] A similar ring was used later on in [2], but they worked modulo a prime p, while no limitation on K is set here.

References

[1] Y. Desmedt and Y. Frankel, "Perfect zero-knowledge sharing schemes over any finite Abelian group," in *Sequences II (Methods in Communication, Security, and Computer Science)* (R. Capocelli, A. De Santis, and U. Vaccaro, eds.), pp. 369–378, Springer-Verlag, 1993. Positano, Italy, June 17–21, 1991.

[2] M. Blaum and R. M. Roth, "New array codes for multiple phased burst correction," *IEEE Transaction on Information Theory*, vol. IT-39, pp. 66–77, January 1993.

[3] Y. G. Desmedt and Y. Frankel, "Homomorphic zero-knowledge threshold schemes over any finite abelian group," *SIAM Journal on Discrete Mathematics*, vol. 7, pp. 667–679, November 1994.

[4] J. A. Reeds and N. J. A. Sloane, "Shift-register synthesis (modulo m)," *Siam J. Comput.*, vol. 14, no. 3, pp. 505–513, 1985.

Quasi-Cyclic Goppa Codes

Bezzateev S.V. and Shekhunova N.A.
Academia of Aerospace Instrumentation, St. Petersburg, Russia
e-mail: bsv@compromise.spb.su

Abstract — **We describe here the one subclass of quasi-cyclic Goppa codes with Gopa polinomial** $G(x) = x^t - 1$.

I. Introduction

It is well known that Goppa codes include as cyclic codes only BCH codes (with Goppa polinomial $G(x) = x^t$)[1] and double error-correcting cyclic codes(extended double error-correcting Goppa codes)[2]. Here we will discuss a subclass of binary Goppa codes with Goppa polinomial $G(x) = x^t - 1$ and location set $L = \{\alpha_j \cdot \alpha^{i \cdot l}\}$, $j = 1..p$, $i = 0..t-1$, $p \leq l, \alpha-$ is a primitive element of $GF(2^m)$, $l \cdot t = 2^m - 1$ and $G(\alpha_j) \neq 0$. It is easy to show that such Goppa codes are quasi-cyclic.

II. Quasi-cyclic Goppa codes

In this paper,as an example, we would like to discuss some codes from special subclass of quasi-cyclic Goppa codes with following type of generator matrix:

$$G = \begin{bmatrix} |c_{1,1}| & |c_{1,2}| & \cdots & |c_{1,p-1}| & |0| \\ |c_{2,1}| & |c_{2,2}| & \cdots & |0| & |c_{2,p}| \\ |v_{1,1}| & |v_{1,2}| & \cdots & |v_{1,p-1}| & |v_{1,p}| \\ |v_{2,1}| & |v_{2,2}| & \cdots & |v_{2,p-1}| & |v_{2,p}| \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ |v_{q,1}| & |v_{q,2}| & \cdots & |v_{q,p-1}| & |v_{q,p}| \end{bmatrix}$$

where $|c_{ij}|$- generator submatrix of cyclic code with lenght m and generator polinomial $c_{ij}(x)$, $|0|$- zero submatrix, $|v_{ij}|$ - all-zero or all-one vector.

1. (55,16,19)-Goppa code [3,4]. $G(x) = x^9 - \alpha^{54}$ and location set $L = \{\alpha_j \cdot \alpha^{i \cdot 7}\}$, $j = 1..6$, $i = 0..8$,
$\alpha_1 = 1, \alpha_2 = \alpha, \alpha_3 = \alpha^2, \alpha_4 = \alpha^3, \alpha_5 = \alpha^4, \alpha_6 = \alpha^5, \alpha_7 = 0$, $\alpha-$ is a primitive element of $GF(2^6)$.

$$G = \begin{bmatrix} |451| & |231| & |066| & |563| & |440| & |000| & |0| \\ |462| & |707| & |275| & |743| & |000| & |440| & |0| \\ |000| & |777| & |000| & |777| & |000| & |777| & |0| \\ |000| & |777| & |777| & |777| & |777| & |000| & |0| \\ |777| & |000| & |000| & |777| & |777| & |000| & |0| \\ |777| & |777| & |000| & |777| & |000| & |000| & |1| \end{bmatrix}$$

where, for example, $|451_8|$ corresponds to [100101001] - generator submatrix of (9,6,2)-cyclic code with generator polinomial $g(x) = (x^3 + 1) \cdot (x^5 + 1)$. ¿From this code it is easy to construct two different(with different weight distribution) (46,9,19)[5] quasi-cyclic codes ,(46,11,17)[5] and (28,9,10) quasi-cyclic code.

2. (103,20,35)-Goppa code. $G(x) = x^{17} - 1$ and location set $L = \{\alpha_j \cdot \alpha^{i \cdot 15}\}$, $j = 1..7$, $i = 0..17$,
$\alpha_1 = \alpha^3, \alpha_2 = \alpha^5, \alpha_6 = \alpha^2\, \alpha_4 = \alpha^9, \alpha_5 = \alpha^{10}, \alpha_6 = \alpha^{12}, \alpha_7 = 0$, $\alpha-$ is a primitive element of $GF(2^8)$.

$$\begin{bmatrix} |342021| & |011251| & |331364| & |074202| & |143377| & |000000| & |0| \\ |332533| & |364213| & |016316| & |264774| & |000000| & |143377| & |0| \\ |377777| & |377777| & |000000| & |377777| & |000000| & |000000| & |0| \\ |377777| & |000000| & |377777| & |000000| & |377777| & |000000| & |0| \\ |000000| & |377777| & |377777| & |000000| & |000000| & |377777| & |0| \\ |377777| & |377777| & |377777| & |000000| & |000000| & |000000| & |1| \end{bmatrix}$$

The generator polinomials are given here in octal, i.e. $|011251_8|$ coresponds to |00001001010101001| - generator submatrix of (17,8,6)-cyclic code with generator polinomial $g(x) = (x^8 + x^4) \cdot (x^8 + x^5 + x^4 + x^3 + 1)$.
¿From this code it is easy to construct best known (102,16,40) quasi-cyclic code with generator matrix

$$\begin{bmatrix} |342021| & |011251| & |331364| & |074202| & |143377| & |000000| \\ |332533| & |364213| & |016316| & |264774| & |000000| & |143377| \end{bmatrix}$$

This code improves the lower bounds on the maximum minimum distance for (102,16), (101,16), (100,16) and (99,16) binary linear codes[6].

3. (136,20,52)-Goppa code. $G(x) = x^{17} - 1$ and location set $L = \{\alpha_j \cdot \alpha^{i \cdot 15}\}$, $j = 1..8$, $i = 0..17$,
$\alpha_1 = \alpha^1, \alpha_2 = \alpha^2, \alpha_3 = \alpha^4\, \alpha_4 = \alpha^7, \alpha_5 = \alpha^8, \alpha_6 = \alpha^{11}, \alpha_7 = \alpha^{13}, \alpha_8 = \alpha^{14}$, $\alpha-$ is a primitive element of $GF(2^8)$. From this code it is easy to construct (119,11,52) quasi-cyclic code.

References

[1] V.D.Goppa, "A new class of linear error correcting codes", Probl. Pered. Inform., vol.6,pp.24-30,Sept. 1970

[2] F.J.MacWilliams and N.J.A.Sloane, The Theory of Error Correcting Codes, Amsterdam:North Holland, 1977.

[3] M. Loeloeian and J.Conan, " A [55,16,19] binary Goppa code", IEEE Trans. Inform. Theory, vol. IT-30, pp.773, Sept. 1984.

[4] S.V.Bezzateev, E.T. Mironchikov, N.A. Shekhunova, "Subclass of binary Goppa codes", Probl. Pered. Inform., vol.25, pp.98 -102,Oct. 1989.

[5] B.Groneick and S.Grosse, "New Binary Codes", IEEE Trans. Inform. Theory, vol. IT-40, pp.510-512, March 1994.

[6] A.E.Brouwer and T.Verhoeff, "An update table of minimum distance bounds for binary linear codes", IEEE Trans. Inform. Theory, vol.39, pp.662-667,Mar.1993.

New ternary linear codes

Iliya G. Boukliev
Institute of Mathematics, Bulgarian Academy of Sciences, P.O.Box 323, 5000 V.Tarnovo, Bulgaria

Abstract - **Linear codes with parameters [47, 5, 30; 3], [44, 6, 27; 3], [90, 6, 57; 3] and [94, 6, 60; 3] have been found.**

I. Introduction

Let $GF(q)$ denote the Galois field of q elements. A linear code of length n, dimension k, and minimum Hamming distance d, over $GF(q)$ is called an $[n, k, d; q]$-code. Let $n_q(k, d)$ denote the minimum n for which an $[n, k, d; q]$ code exists.

For linear codes over $GF(q)$ with $q > 2$, there is a natural generalization of the class of constacyclic codes to the class of cyclic codes [2]. A constacyclic (α - twisted) code has the following property: For some fixed element α of $GF(q)$, if $(a_0, a_1, \ldots, a_{n-1})$ is a codeword then $(\alpha a_{n-1}, a_0, a_1, \ldots, a_{n-2})$ is a codeword too. The theory of constacyclic codes is very similar to that of cyclic codes.

The algebra of twistulant $m \times m$ matrices over $GF(q)$ is isomorphic to the algebra of polynomials in the ring $F[x]/(x^m - \alpha)$. The $[pm, k]$- codes C with generator matrices of type: $[B_1, B_2, \ldots, B_p]$ where each B_i is a twistulant matrix are called quasi-twisted [4].

Let $c_1(x), c_2(x), \ldots, c_p(x)$ be the polynomials corresponding to twistulant $m \times m$ matrices $B_1, B_2, \ldots, B_p$ and $h(x) = (x^m - \alpha)/gcd(x^m - \alpha, c_1(x), c_2(x), \ldots, c_p(x))$. Then the dimension k of C is equal to the degree of $h(x)$. Two polynomials, $c_j(x)$ and $c_i(x)$, belong to the same class if $c_j(x) = \alpha x^l c_i(x)$ mod $(x^m - \alpha)$, for some integer, $l \geq 0$. Two twistulant matrices, B_i and B_j, are called conjugates if $c_i(x)$ and $c_j(x)$ belong to the same class.

Good quasi-twisted codes are obtained if there are no conjugates in the generator matrix.

II. Results

LEMMA 1.[3] $47 \leq n_3(5, 30) \leq 48$, $44 \leq n_3(6, 27) \leq 45$, $89 \leq n_3(6, 57) \leq 91$, $93 \leq n_3(6, 60) \leq 96$.

THEOREM 1.

(i) $89 \leq n_3(6, 57) \leq 90$;

(ii) $n_3(5, 30) = 47$, $n_3(6, 27) = 44$, $93 \leq n_3(6, 60) \leq 94$.

Proof:

(i) The $[90, 5, 57; 3]$ codes were constructed as quasi-twisted with a rate $1/p$ and $(m - 4)/pm$. The generator polynomials are:

110000, 000121, 000122, 001002, 001022, 001101, 001211, 010122, 010212, 011011, 011021, 011112, 011122, 011212, 111111;

1021210000, 1210111000, 2111211000, 2021101000, 1202001100, 1022111100, 1112221100, 1211012100, 1210201110.

(ii) Codes with parameters $[47, 5, 29; 3]$, $[44, 6, 27; 3]$ are constructed by the method from [1]. The generator matrices are:

$$G_1 = \begin{pmatrix} 00000000000000001111111111111111111111111111111 \\ 00000011111111110000000001111111112222222222 \\ 00111100000111200111122220000111222000011122 \\ 11000111122002211100120012011201200111200100 2 \\ 01012010121201010010122122110102001210120020122 \end{pmatrix}$$

$$G_2 = \begin{pmatrix} 000000000000001111111111111111111111111111111 \\ 000001111111100000000011111111111222222222 \\ 00111001122222011112222001111122222011112222 \\ 00012120101122201220112120112201122101220112 \\ 01002202110102020021022121020120102012022011 \\ 10010100011220000212102201021221001112100212 \end{pmatrix}$$

The weight distributions are:

$[47, 5, 30; 3]$ - $A_0 = 1$, $A_{30} = 166$, $A_{33} = 46$, $A_{36} = 20$, $A_{39} = 8$, $A_{42} = 2$,

$[44, 6, 27; 3]$ - $A_0 = 1, A_{27} = 352, A_{30} = 264, A_{33} = 24, A_{36} = 88$.

A $[94, 6, 60; 3]$-code was also constructed by the method from [1], and has a weght distribution $A_0 = 1, A_{60} = 456, A_{63} = 76, A_{69} = 192, A_{72} = 4$.

References

[1] I. Boukliev, " A method for construction of good linear codes and its application to ternary and quaternary codes," International Workshop on Optimal Codes, Sozopol, Bulgaria, May-June 1995.

[2] F. J. MacWilliams, and N. J. A. Sloane, The theory of error-correcting codes, Amsterdam: North-Holand 1977.

[3] N. Hamada, "A survey of recent work on characterization of minihypers in $PG(t, q)$ and nonbinary linear codes meeting the Griesmer bound," *J. Combin. Inform. Syst. Sci.* vol. 18, pp. 161-191, 1993.

[4] R. Hill, P.P. Greenough, "Optimal Quasi-Twisted Codes," in Proc. International Workshop on Algebraic and Combinatorial Coding Theory, Voneshta voda, Bulgaria, June 22-28,pp. 92-97, 1992.

AUTHOR INDEX

A

Aakvaag, N.D. 116
Aazhang, B. 316, 480
Abdat, M. 18
Abdel-Ghaffar, K.A.S. 143, 341
Abdi, A. 425
Abrahams, J. 326
Acheroy, M. 365
Achour, B. 115
Aguado-Bayón, L.E. 347
Agus, S. 362
Ahlswede, R. 19, 69
Al-Bassam, S. 144
Al-Semari, S.A. 216
Alajaji, F. 286, 360
Alencar, M.S. 117
Amari, S.-I. 447
Amirmehrabi, H. 295
Amrani, O. 337
Andersen, J.D. 36
Anderson, J.B. 66, 400
Andersson, H. 309
Arani, F. 289
Arikan, E. 322
Arimoto, S. 56, 82
Arpasi, J.P. 307
Augot, D. 349
Avetissian, A.E. 135

B

Baccarelli, E. 68, 179, 333
Baggen, C.P.M.J. 243
Bajić, D. 273
Balachandran, K. 66
Balakirsky, V.B. 490
Baraniuk, R.G. 426
Barbulescu, A.S. 37
Barnes, C.F. 185
Basseville, M. 330
Baum, C.W. 398
Be'ery, Y. 129
Beaulieu, N.C. 340
Bejjani, E. 212
Belfiore, J.-C. 212
Bell, M.R. 290
Benedetto, S. 32
Berger, T. 74, 79, 192, 260, 263, 265, 266
Berger, T.P. 491
Berger, Y. 129
Berrou, C. 34
Beth, T. 132, 279, 429
Betz, J.W. 298
Bezzateev, S.V. 499
Bhargava, V.K. 285
Biglieri, E. 114, 182, 208, 306, 472
Bist, A. 374
Bitmead, R. 121
Bitzer, D.L. 164, 227
Blachman, N.M. 188
Blackburn, S.R. 409
Blahut, R.E. 101
Blake, I.F. 102, 125
Blakley, G.R. 488
Blaum, M. 246, 412
Blinovsky, V. 57
Bloemen, A.H.A. 445
Blostein, S.D. 155
Blum, R.S. 214
Bobrowski, R. 67
Börjesson, P.O. 331, 332
Bose, B. 144, 236
Boukliev, I.G. 500
Boutros, J. 157
Brandestini, M. 206
Brandt-Pearce, M. 31
Brassard, G. 4
Bratt, G. 288
Braun, V. 203
Broeg, R. 236
Bruck, J. 246
Bucklew, J.A. 418
Buisán Gómez del Moral, J. 114
Burlina, P. 360
Burnashev, M.V. 167
Burr, A.G. 213
Buz, R. 151
Buzzi, S. 297

C

Cabral, H.A. 497
Cai, Z.Q. 368
Caire, G. 208, 472
Calderbank, A.R. 149
Campello de Souza, J. 489
Campello de Souza, R.M. 489
Cardoso, J.-F. 330
Carlet, C. 241
Casadei, F. 323
Castoldi, P. 301
Chan, F. 61
Chan, M.Y. 355
Chandran, G. 93
Chang, C.S. 103
Chang, Y.-W. 108
Chao, C.-C. 496
Charon, I. 242
Charpin, P. 491
Chellappa, R. 360
Chen, C.J. 411
Chen, P.-N. 118
Chen, Q. 435
Chen, X. 94, 97
Cheng, J.-F. 33, 325
Cheng, R.S. 137
Cheng, V.W. 215
Cherubini, G. 401
Cheung, K.-M. 376
Chiu, M.-C. 496
Chou, P.A. 371
Choy, W.W. 340
Chugg, K.M. 402
Cioffi, J.M. 335, 399
Clarke, W.A. 204
Coelho, P.H.G. 172
Coffey, J.T. 53, 410, 455
Cohen, G.D. 234
Collins, O.M. 269, 352, 410
Comaniciu, D. 441
Cong, L. 174
Conner, K.F. 398
Conte, E. 297
Cooper III, A.B. 492
Cordier, M. 338
Corrada, C.J. 464
Costello, D.J. 160, 220
Cover, T.M. 9, 73
Csibi, S. 385
Csiszár, I. 6
Cusani, R. 68, 179, 333

D

D'yachkov, A.G. 75
da Rocha, Jr., V.C. 497
Dabak, A.G. 268
Dabiri, D. 102
Dallal, Y.E. 476
Dam, W.C. 152
Daneshgaran, F. 30, 65

Darnell, M. 460, 461, 462
Daubechies, I. 5
de Alencar, C.D. 153
DeLeone, J. 361
Desmedt, Y. 498
Dholakia, A. 164, 227
Di Blasio, G. 333
Divsalar, D. 35
Dolinar, S. 325
Drajić, D. 273
Duman, T.M. 378, 440
Duverdier, A. 116, 419

E

Effros, M. 325, 371
Einarsson, G. 482
Encheva, S.B. 130
Ephremides, A. 105
Ericson, T. 311
Eriksson, H.B. 331, 332
Erkip, E. 9, 73
Esener, S.C. 141
Esmaelli, M. 348
Evans, W.S. 456

E

Fahimi, H. 369
Fair, I.J. 285
Fan, P.Z. 460, 461, 462
Fang, S.C. 170
Farrell, P.G. 219, 224, 347
Feder, M. 16, 71, 133, 233
Feng, D.-G. 358
Feng, G.L. 95
Ferreira, H.C. 119, 147, 204
Fessler, J.A. 176
Fine, T.L. 168
Fischer, T.R. 435, 437
Fitz, M.P. 291, 336
Flandrin, P. 426
Fonollosa, J.R. 384
Fonollosa, J.A.R. 384
Forchhammer, S. 249
Forney Jr., G.D. 1
Fossorier, M.P.C. 55, 415
Franaszek, P. 15
Francos, J.M. 364
Freund, Y. 233
Friedlander, B. 364
Fu, F.-W. 238, 424
Fuja, T.E. 216, 286
Fujiwara, A. 138
Fujiwara, E. 148
Fujiwara, T. 470

G

Gabidulin, E.M. 460, 467
Gallager, R.G. 139
Galli, S. 179
Games, R.A. 485
Gass, J.H. Jr. 24
Gehrmann, C. 350
Geischläger, F. 439
Gelblum, E.A. 149
Gelfand, S.B. 291, 336
Georghiades, C.N. 244
Georgiadis, L. 109
Geraniotis, E. 108, 338
Gersho, A. 257, 432
Gibson, J.D. 197, 420
Glavieux, A. 34
Goel, M. 413
Goh, S.-C 48.
Golomb, S.W. 458, 464
Gonçalves, V. 284
Grant, A.J. 383, 448
Gray, P.K. 64
Gray, R.M. 371
Grenander, U. 392
Grishin, Y.P. 302
Gubner, J.A. 421
Guey, J.-C. 290
Gulliver, T.A. 348, 493
Gusmão, A. 284
Gustafsson, F. 121

H

Haccoun, D. 61
Haché, G. 100
Hajek, B. 104
Halford, K.W. 31
Hamkins, J. 184
Han, S.-J. 353
Han, T.S. 19, 324
Hanly, S.V. 446
Hansen, C.J. 387
Hardin, R.H. 181
Haroutunian, E.A. 135
Haroutunian, M.E. 135
Hartmann, C.R.P. 414
Hasan, M.A. 49
Hashimoto, T. 231, 343
Hassan, A.A. 479
Hau, K.P. 262
Hayat, M.M. 421
He, Z.Y. 292
Hedelin, P. 436
Heegard, C. 126, 187
Helleseth, T. 88, 92, 94, 274, 281, 283, 408
Hero, A.O. 176, 180
Herzberg, H. 270
Hill, R. 345
Hiltgen, A.P. 206
Hirasawa, S. 17, 50, 388
Hole, K.J. 146
Hollmann, H.D.L. 327
Holubowicz, W. 34, 63, 67
Homer, J. 121
Honary, B. 142, 289
Honig, M.L. 314, 381
Hsieh, M.-H. 393
Hsuan, Y. 410
Huang, H.C. 380
Huber, J. 62
Hudry, O. 242
Hughes, B.L. 106, 442, 492
Hui, D. 372

I

Ihara, S. 191, 300
Iltis, R.A. 320
Imai, H. 145, 159, 479
Interlando, J.C. 277
Itoh, S. 231
Iwata, K. 21, 22
Izzo, L. 339

J

Jacobsen, G. 476
Jacquet, P. 14
Jaffe, J.S. 93
Jagerman, D.L. 42
Janwa, H. 484
Jayaraman, S. 263
Jensen, J.M. 280
Jensen, O.R. 469
Ji, C. 175
Jin, F. 461
Johannesson, R. 163, 288
Johansson, T. 354
Joiner, L.L. 416
Judd, J.S. 169
Justesen, J. 249

K

Kabatianski, G.A. 488
Kafedziski, V.G. 156
Kahrizi, M. 299
Kaleh, G.K. 189, 201, 473
Kalouti, H. 279
Kanaya, F. 83
Kang, C.G. 486
Kang, D.-S. 367
Kanlis, A. 264
Kasami, T. 154, 470, 474
Kasner, J.H. 433
Katić, O. 273
Kato, A. 324
Kawabata, T. 391
Kempf, P. 321
Kerpez, K.J. 481
Keuning, J. 444
Key, E.L. 485
Khandani, A.K. 207
Khansari, M. 140
Khayrallah, A.S. 363
Khayrallah, A.S. 199
Kieffer, J. 267
Kiely, A.B. 394
Kim, K.-J. 48, 356
Kim, M.-G. 202
Kim, S.W. 27, 444
Kim, Y. 356
Kitakami, M. 148
Kittel, L. 211
Klappenecker, A, 429
Kløve, T. 237, 342
Kobayashi, H. 107
Kobayashi, K. 19
Koga, H, 82
Kogan, J.A. 178
Kohno, R. 319
Kolodziejski, K.R. 298
Komo, J.J. 416
Komori, S. 362
Kondo, H. 362
Kong, H. 210
Koorapaty, H. 164, 227
Koplowitz, J. 361
Koshelev, V.N. 77
Koski, T. 331, 332
Kot, A.D. 128
Kötter, R. 468
Koumoto, T. 470
Krasikov, I. 344
Krichevskii, R.E. 431
Krzyzak, A. 258
Kschischang, F.R. 122
Kulkarni, S.R. 251, 255
Kumar, P.V. 88, 92, 274, 283, 408
Kurihara, M. 99
Kurtas, E. 245
Kwon, J.M. 27

L

Lacaze, B. 116, 419
Lafourcade-Jumenbo, A. 123, 124
Lahtonen, J. 85
Lapidoth, A. 193
Larrea-Arrieta, J. 218
Larsson, T. 51
Lazic, D.E. 132, 279
Leclair, P. 212
Lee, D.-K. 48
Lee, H. 411
Lee, J.H. 202
Lee, J.S. 28
Lee, L.H.C. 219
Lee, L.L. 171
Lee, S. 356
Lee, S.-J. 48
Leung, C. 128
Leung, N.K.N. 455
Leung, P.S.C. 219
Levenshtein, V.I. 483
Levitin, L.B. 205
Levy-dit-Vehel, F. 278
Li, T.-H 420
Lin, S. 55, 127, 154, 209, 415, 470, 474
Lin, W. 131
Linder, T. 258, 370
Litsyn, S.N. 234, 278, 344
Liu, B. 358
Liu, Y. 155
Liu, Y.-S. 442
Liu, Z.-J. 45
Lizak, P. 345
Ljungberg, P. 225
Lobstein, A. 234, 242
Loeliger, H.-A. 304, 309, 468
Loher, U. 44
Löhnert, R. 52
Lops, M. 297
Lorenzelli, F. 120
Luczak, T. 80
Lugosi, G. 229, 254, 258
Luo, J. 200
Luo, Z.-Q. 152

M

Ma, S. 175
Ma, X. 113
Madhow, U. 313
Mallik, R.K. 272
Mandayam, N.B. 26
Mansour, Y. 233
Marcellin, M.W. 433
Marcus, B. 2
Mareels, I. 121
Marić, S.V. 87
Mark, B.L. 42
Mark, K.E. 392
Markarian, G. 142
Masry, E. 259, 427
Massey, J.L. 11
Massey, P. 303
Mathys, P. 303
Matsushima, T. 17, 388
Mattson, Jr., H.F. 234
Maurer, U.M. 12
McClellan, S.A. 197
McEliece, R.J. 33, 131, 325, 329
McLaughlin, S.W. 200
Medard, M. 139
Meeuwissen, H.B. 445
Méhes, A. 377
Menyennett, C. 147
Merhav, N. 16
Michel, G. 41
Michel, O. 426
Miller, D. 257, 432
Miller, J.W. 454
Miller, L.E. 28
Miller, M.I. 392
Milstein, L.B. 23
Mitra, U. 312
Mittelholzer, T. 305
Modha, D.S. 259
Mohan, C.K. 414
Molnar, K.J. 479
Mondin, M. 30, 65
Monroe, L. 235
Monroy, I.T. 482
Montorsi, G. 32
Moon, T.K. 250
Moorthy, H.T. 127, 474
Morelos-Zaragoza, R.H. 154
Moreno, O. 87, 283, 464, 484
Mori, S. 478
Mori, T. 145
Morita, H. 465, 466

Moulin, P. 252
Mow, W.H. 90, 450, 459
Müller, F. 194, 439
Murali, R. 106
Muramatsu, J, 83

N

Nader-Esfahani, S. 425
Nagaoka, H. 138, 324
Narayan, P. 264
Nassar, C.R. 403
Natarajan, P. 115
Nelson, G. 267
Neuhoff, D.L. 199, 372, 438
Neyt, X. 365
Nguyen, L. 480
Nikias, C.L. 113
Niles, L.T. 434
Nilsson, J.E.M. 407
Nishijima, T. 50
No, J.-S. 86
Nobel, A. 254
Noneaker, D.L. 24
Noonan, J.P. 115
Noubir, G. 43
Nurmela, K.J. 346

O

O'Sullivan, J.A. 177, 248
Ödling, P. 331, 332
Ogiwara, H. 47
Oh, H.-S. 353
Ohtsuki, T. 478
Okamoto, E. 21, 22, 134
Ölçer, S. 401
Olson, B.H. 141
Oohama, Y. 261
Oppermann, I. 110
Ordentlich, E. 443
Orlitsky, A. 451
Östergard, P.R.J. 346
Ottosson, T. 315
Ouaissa, K. 18

P

Paar, C. 58
Paaske, E. 469
Palazzo, Jr., R. 277, 307
Papamarcou, A. 118
Papasakellariou, A. 316
Papproth, E. 473
Paris, B.-P. 405
Park, S. 356
Paterson, K.G. 206
Pawlak, M. 253
Pearlman, W.A. 373
Peng, X.-H. 224
Perez, L.C. 160
Persson, J. 59
Phamdo, N. 286
Piazzo, L. 68
Pietrobon, S.S. 37, 471
Pilipchouk, N.I. 379
Pinsker, M.S. 10
Pless, V. 235, 282
Plume, P. 18
Podemski, R. 34
Pokam, M.R. 41
Pollara, F. 35
Polydoros, A. 402
Poor, H.V. 312, 428
Porter, D.G. 248
Portugheis, J. 153
Posner, S.E. 251
Potapov, V.N. 431
Pottie, G.J. 217, 387
Prelov, V.V. 10, 70
Proakis, J.G. 245
Pursley, M.B. 24, 60

Q

Qian, Z. 282
Qing, G. 46

R

Raheli, R. 301
Raja, P. 43
Rajagopalan, S. 453
Rajan, B.S. 413, 495
Ramamurthy, G. 42
Ran, M. 275
Rao, A. 257, 432
Rao, T.R.N. 95
Rapajic, P.B. 110
Rasmussen, L.K. 64, 317
Ray-Chaudhuri, S. 240
Redinbo, R. 221
Reed, I.S. 94, 97
Regazzoni, C.S. 296
Ren, Q. 107
Retter, C.T. 276
Reznikova, Z. 78
Rhee, D.J. 209
Riedel, S. 39
Rimoldi, B. 136, 448
Ritthongpitak, T. 148
Roche, J.R. 262, 451
Rose, K. 257, 432
Rosenthal, J. 162, 165
Rossin, E.J. 126
Roth, R.M. 239
Rowe, D.J. 126
Rowitch, D.N. 23
Roy, S. 381
Rozenbaum, Y. 31
Ruprecht, J. 89
Rushanan, J.J. 485
Ruszinkó, M. 76
Ryabko, B. 78, 395

S

Saadat, A. 369
Sadeh, I. 84, 196
Sadowsky, J.S. 156, 423
Said, A. 226, 373, 400
Saifuddin, A, 319
Sakai, T. 47
Sakata, S. 96, 99
Sakuma, Y. 300
Salazar-Anaya, G. 422
Salehi, M. 198, 245, 378, 440
Sallent, S. 430
Saltzberg, B.R. 270
Sarkar, S. 428
Sarvis, J.P. 308
Sasase, I. 478
Savari, S.A. 328
Schalkwijk, J.P.M. 445
Scharcanski, J. 117
Schlegel, C.B. 318, 383
Scholtz, R.A. 272
Schotten, H.D. 29, 89
Schulman, L.J. 452, 453, 456
Schwarte, H. 423
Schwartz, S.C. 380
Sechrest, S. 455
Secord, N.P. 348
Seroussi, G. 390
Seyfe, B. 299
Shah, A.A. 112
Shah, A.R. 405
Shamai (Shitz), S. 7, 13, 476
Shanbhag, A.G. 88, 92, 274, 283
Sharma, V. 40
Shea, J.M. 60
Shekhunova, N.A 499.

Shen, B.-Z. 98, 186
Shen, S.-Y. 238, 424
Shih, C.-C. 414
Shilou, J. 46
Shimokawa, H. 447
Shiu, J. 158
Shulman, N. 133
Shwedyk, E. 210
Siala, M. 189, 201, 473
Sidelnikov, V.M. 75
Skoglund, M. 315, 436
Sloane, N.J.A. 181
Smietana, R. 289
Snapp, R.R. 256
Snyder, D.L. 177
Snyders, J. 275, 415
Sokolov, A.I. 302
Sola, M.Á. 430
Soleymani, M.R. 403
Soljanin, E. 244
Sosa, B.R.M. 171
Stadtmüller, U. 253
Stark, W.E. 215, 386
Stasevich, S.I. 77
Stojanovic, M. 382
Strandberg, J. 482
Subrahmanya, P. 266
Suehiro, N. 91
Sun, F.-W. 161
Sung, W. 53
Suzuki, H. 56
Suzuki, J. 232, 389
Svirid, Y.V. 38, 39, 237
Swaszek, P.F. 54
Szpankowski, W. 14, 80, 109

T

Taieb, K.H. 473
Tait, D.J. 218, 219
Tajima, M. 222
Takata, T. 470
Takeuchi, J.-I. 228, 391
Tallini, L. 144
Tamm, U. 72
Tanaka, H. 357
Tanda, M. 339
Taricco, G. 472
Tarköy, F. 334
Tarokh, V. 125
Tassiulas, L. 109
Tavares, S.E. 351
Taylor, D.P. 152
Telatar, I.E. 8
Tesei, A. 296
Thomas, J.A. 15, 103
Thomson, D.J. 111
Tolhuizen, L.M.G.M. 243
Traganitis, A. 105
Tran, V.N. 368
Trott, M.D. 308
Tsuchiya, H. 231
Tufts, D.W. 112
Turmon, M.J. 168
Tyczka, P. 63
Tzeng, K.K. 98, 186, 411

U

Uçan, O.N. 406
Uehara, G. 127
Ungerboeck, G. 401
Urbanke, R. 136, 448
Urías, J. 422
Usman, M. 180
Uyematsu, T. 21, 22, 134

V

Vaccaro, R.J. 112
Vainstein, F.S. 205
van der Meulen, E.C. 70
van Dijk, M. 487
van Tilborg, H.C.A. 412
van Wijngaarden, A.J. 463, 465, 466
Vanroose, P. 327
Varanasi, M.K. 25, 404
Vardy, A. 123, 124, 183, 246, 337
Varshney, P.K. 293
Varvarigos, E.A. 40
Veeravalli, V. 294
Venkatesh, S.S. 169, 170, 256
Ventura-Traveset, J. 208
Verdú, S. 7, 10, 13, 380
Véron, P. 359
Vetterli, M. 140
Veugen, T. 457
Vidal, J. 384
Vijayananda, K. 43
Villalobos, I.R.T. 366
Vinck, A.J.H. 161, 444, 466
Viswanathan, H. 260
Viswanathan, R. 295
Viterbo, E. 157, 182
Volf, P.A.J. 20
Von York, E. 162, 165
Vouk, M.A. 164, 227
Vucetic, B.S. 110

W

Wachsmann, U. 62
Wahlberg, B. 121
Wan, Z.-X. 166
Wang, C. 169, 186
Wang, F.-Q. 220
Wang, Q. 271, 285
Wang, X.-A. 173, 475
Wang, X.-M. 417
Wang, Y.-P. 386
Weber, J.H. 143
Wei, L. 317
Wei, V.K. 190
Wei. C.-H. 393
Weinberger, M. 390
Welch, L.R. 396
Wesel, R.D. 399
Whiting, P.A. 446, 448
Wiberg, N. 468
Wicker, S.B. 173, 185, 475
Wilcox, L.D. 434
Wilhelmsson, L. 310
Willems, F.M.J. 20, 323
Willink, T.J. 397
Wilson, S.K 335
Winjum, E. 281
Wittke, P.H. 397
Wong, P.W. 195
Wornell, G.W. 150
Wu, J.-L. 158
Wulff, C.R. 414

X

Xiang, Z. 318
Xiao, G.-Z. 358
Xie, Q. 200
Xuemai, G. 46
Xydeas, C.S. 174

Y

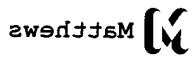
Yamaguchi, K. 159
Yamamoto, H. 449
Yamazaki, K. 477
Yang, E.-H. 69, 79, 81, 190
Yang, L. 437
Yang, Y. 396
Yao, K. 120
Yao, S. 108, 292
Ye, Z. 192, 247
Yellin, D. 337

Yeung, R.W. 74, 262, 355
Young, J.F. 480
Youssef, A.M. 351
Ytrehus, Ø. 146
Yu, B. 230, 375
Yu, C.-T. 293
Yu, J.-P. 45, 417
Yu, Z. 329
Yurchenko, Y.S. 302

Z

Zaidan, M.Y. 185
Zain, A.A. 495
Zamir, R. 71, 265
Zeger, K. 184, 229, 370, 377
Zeitouni, O. 255
Zeng, M. 271
Zepernick, H.-J. 287
Zerai, A.A. 418
Zhang, Z. 69, 79, 81, 190, 247
Zhao, Y.-B. 45, 417
Zhong, L. 46
Zhuang, Z. 269
Zigangirov, K.S. 163, 223, 288
Zimmermann, K.-H. 494
Zinoviev, V. 311
Ziv, J. 3
Zvonar, Z. 382, 384

ISIT'95 CASEBOUND
ISBN 0-7803-2454-4



EAN, Mag 90%
BWR 1 mils