Information theory : proceedings of the 1990 IEEE international workshop, Eindhoven, June 10-15, 1990

PROCEEDINGS OF THE 1990 IEEE INTERNATIONAL WORKSHOP ON INFORMATION THEORY

INTERNATIONAL CONFERENCE CENTER KONINGSHOF,
EINDHOVEN, THE NETHERLANDS

June 10-15, 1990

Sponsored by:
The Institute of Electrical and Electronics Engineers,
Information Theory Society,
Eindhoven University of Technology.

Workshop chairmen: A.J. Vinck, K.A. Schouhamer Immink, and S. Verdu

Session Chairmen:

New Trends in Shannon Theory, organizers: *F. Willems and E. van der Meulen,* Eindhoven University of Technology and Katholic University of Leuven;

Covering Radius Problems in Algebraic Coding, organizers: *H. van Tilborg and J. van Lint,* Eindhoven University of Technology;

Convolutional Codes, organizer: *R. Johannesson,* University of Lund;

Artificial Neural Networks, organizer: *J. Bruck,* IBM Almaden Research Center;

Shannon Theory, organizer: *S. Verdu,* Princeton University.

Recording Systems, organizer: *K. Schouhamer Immink,* Philips Research Laboratories, Eindhoven;

Cryptology, organizer: *J. Massey,* ETH Zuerich;

Source Coding, organizer: *T.S. Han,* Senshu University Tokyo;

Local Arrangements: P.G.M. de Bot
Finance: A.J. Vinck
Registration: Tj.J. Tjalkens
Publicity: W.M.C.J. van Overveld
Proceedings Editor: K.A. Schouhamer Immink

Acknowledgement

The following organizations have provided financial support for the 1990 IEEE IT Workshop:

AT&T, Dutch Radio and Electronics Engineering Society (NERG), IBM Nederland N.V., IEEE Benelux Section, IEEE Region 8, OCE-van der Grinten N.V., Philips International B.V., PTT Research, PTT Telecom, Scientific University Fund Eindhoven, SIEMENS, and others pending.

CIP-gegevens Koninklijke Bibliotheek, Den Haag
Proceedings
Proceedings of the 1990 IEEE International Workshop on Information Theory / edited by K.A. Schouhamer Immink. Met lit. opg.
ISBN 90-9003360-2
SISO xxx.x UDC xxx.xx
Trefw.: Informatietheorie

1.0 Contents

Preface

Information Theory is an active field of research in the Benelux (BElgiumNEtherlandsLUxembourg). At present, we have the Benelux Information and Communication Theory Society with approximately 100 members and the IEEE Chapter on Information Theory. This year we organize the 11-th Symposium on Information Theory in the Benelux. In Eindhoven, we are proud of having two associate editors: Henk van Tilborg and Frans Willems. Their contribution to the Information Theory Community is greatly acknowledged. Going back into history we can mention the first IEEE International Symposium on Information Theory in 1962 in Brussels (Belgium) and the 1970 Symposium in Noordwijk (the Netherlands) organized by Prof. L. Stumpers. These symposia positively influenced research and education in the Benelux. We expect that the 1990 Workshop on Information Theory again stimulates researchers and young scientists to further information and communication theory.

We hope that you enjoy your stay and that it is worthwhile both from a scientific as well as from a cultural point of view.

Han Vinck, Workshop Chairman.

Session 1

New Trends in Shannon Theory

On Identification via Multi-Way Channels with Feedback

R. AHLSWEDE AND B. VERBOVEN, University of Bielefeld, Germany, and University of Leuven, Belgium

"Identification for Multi-way channels" was mentioned by Ahlswede and Dueck /1/ as a challenging direction of research. Here we present in case of complete feedback a rather unified theory of identification. (For the classical transmission problem the dream of such a theory did not get fulfilled for more than twenty years.) Its guiding principle is the discovery of /2/, that communicators (sender and receiver) must set up a common random experiment with maximal entropy and use it as randomization for a suitable (see /2/) identification technique. Here we show that this can be done in a constructive way. The proof of optimality (weak converse) is based on a new entropy bound, which can be viewed as a substitute for Fano's Lemma in the present context. The "single-letter" characterization of (second order) capacity regions rests now on an "entropy characterization problem", which often can be solved. For many channels such as the multiple-access, broadcast, interference and two-way channel this is an exercise.

References

/1/ R. Ahlswede and G. Dueck, 'Identification via channels', IEEE Trans. Inform. Theory, vol. 35, pp. 15-29, Jan. 1989.

/2/ R. Ahlswede and G. Dueck, 'Identification in the presence of feedback - a discovery of new capacity formulas', IEEE Trans. Inform. Theory, vol. 35, pp. 30-39, Jan. 1989.

Arbitrarily Varying Channels with Continuous Alphabets

I. CSISZAR, Mathematical Institute of the Hungarian Academy of Sciences, Hungary

The capacity of discrete memoryless arbitrarily varying channels (AVC's) with input and/or state constraints, for deterministic codes and the average probability of error criterion has been determined by Csiszar and Narayan (IEEE Trans. IT-34, pp. 181-193, March 1988). A complete extension of their result is obtained for AVC's with finite input alphabet but general state set and output alphabet, via approximation arguments. When also the input alphabet is infinite a general capacity formula remains elusive. Still, the familiar representation of random coding capacity as a mutual information minmax is proved and a sufficient condition for capacity equal random coding capacity is obtained.

For AVC's with scalar or vector inputs and additive interference consisting of a deterministic part and noise, both arbitrarily varying subject to average power constraints, the above mentioned sufficient condition is satisfied if the input power exceeds the power of the deterministic interference while otherwise the capacity is zero. In particular, for scalar inputs, capacity equals that of an ordinary memoryless channel with additive Gaussian noise whose power is the sum of the powers of the deterministic interference and the noise, provided that the former is smaller than the input power, while otherwise the capacity is zero. A similar though somewhat more complex result is obtained also for the case of vector inputs when a possibly different power constraint is imposed on the noise in each dimension.

Histogram-Based Distribution Estimation and Hypothesis Testing

L. GYORFI, Hungarian Academy of Sciences, Technical University of Budapest, Hungary

There is no distribution estimate which is consistent in total variation for all distributions. A histogram-based distribution estimate is presented which is consistent in total variation under mild conditions on the underlying distribution. This estimate generates a fairly good test for a simple versus composite hypothesis for which there is no most powerful test.

Statistical Inference with Multiterminal Data Compression

T.S. HAN AND S. AMARI, Department of Information Systems, Senshu University, Higashimita 2-1-1, Tama-ku, Kawasaki 214, Japan, and Department of Mathematical Engineering and Informational Instruments, University of Tokyo, Hongo 7-3-1, Tokyo 113, Japan

Suppose we have two hypotheses H: p(x,y) (null hypothesis) and K: q(x,y) (alternative hypothesis) on joint distributions, and let us consider the multiterminal framework such that a pair of i.i.d. data sequences x(n) and y(n) of length n are generated at two separate sites remote from one another, according to the joint distribution p(x,y) or q(x,y). Statisticians usually make a decision about which one of p(x,y) or q(x,y) is actually operating, under the assumption that these data are fully available to them. In many practical situations, however, this is not necessarily the case, and rather there may be imposed some limitations to the capacity for observing the generated data. Therefore, we are reasonably led to the data compression problem of how to encode x(n) and y(n) into the compressed forms f(x(n)) and g(x(n)), which are to be transmitted to a common information-processing center (decoder) . The center is required to make a decision for selecting a better reasonable one among two hypotheses H and K. Here, the sizes of the ranges of the encoder functions f and g are upperbounded by exp(nR) and exp(nS), respectively (R, S are called the rates). The performance of this system is measured by the value of the second kind of error probability b(n) = exp(-sn) given the first kind of error probability a(n) = exp(-rn).

The same situation with such a multiterminal framework can be considered also for the problem of parameter estimation, where, instead of two hypotheses p(x,y) and q(x,y), we are given a parametric family of joint distributions p(x,y|t) indexed by t. The data sequences x(n) and y(n) are separately encoded in the same way as above. The encoder is required to make a good estimate of t with a variance as small as possible, which may be regarded as establishing a multiterminal version of the Cramer-Rao bound for the given family p(x,y|t).

These two problems, that is, hypothesis testing and parameter estimation, are deeply related at the structural level. The basic questions here are how to construct effective encoders f and g and the related optimal decision (or optimal estimate). What is the minimum value of the second kind of error probability (or the minimum value of the variance) which is attainable under the rate constraints R, S. We shall show several basic results concerning these problems with special reference to structural duality.

Qualitatively Independent Partitions and the Information Theory of Weird Differences

J. KOERNER, Dip. Informatica, Universita di Salerno, Italy. The author is on leave from the Mathematics Institute of the Hungarian Academy of Sciences, 1364 Budapest, POB 127, Hungary

Qualitatively independent partitions have been introduced by A. Renyi (Foundations of Probability, Wiley, NY, 1971). Two partitions of a set are said to be QI (qualitatively independent) if they can be generated by two independent random variables, i. e., if every class of one partition intersects all the classes of the other. Given a set of n elements, we denote by $N(k, n)$ the maximum number of pairwise QI k-partitions of an n-set. S. Poljak and Z. Tuza have proved

$$\frac{\log n}{n(n-1)} \leq \frac{1}{n} \log N(n,k) \leq \frac{2}{n}.$$

In particular, to prove the existence bound, they used projective geometries. On the other hand, they noticed that for 3-partitions their lower bound can be improved to 1/3 using Sperner's lemma on incomparable subsets of an n-set. J. Koerner and G. Simonyi (A Sperner-type theorem and qualitative independence, J. Comb. Th., Ser. A, to appear) have improved the latter bound to 0.41. They pointed out that the problem of QI partitions is a special case of a "symmetric version" of the zero-error capacity of the compound channel. Although these symmetrized capacities have no information-theoretic interpretation, the more severe restrictions they impose do not seem to reduce the value of "capacity" with respect to the Shannon theory problem they can be associated with. Were the Poljak-Tuza upper bound tight, it would give another example of this strange phenomenon.

Recently, L. Gargano, U. Vaccaro and the author have found new classes of compound channels for which "symmetrization does not reduce capacity". This leads to improvements in the Poljak-Tuza lower bounds for k up to 13. The constructions suggest a new proof technique in asymptotic combinatorics based on "trimming" variable-length codes and use Shannon's capacity theorem for the noiseless channel.

Applications of Recent Results in Ergodic Theory to Old and New Problems in Information Theory

P.C. SHIELDS, Deptartment of Mathematics, University of Toledo, Toledo, OH43606, USA

A new covering technique, introduced by Ornstein and Weiss in their efforts to extend the asymptotic equipartition property (AEP) to random fields, has turned out to be quite useful in many settings. I will discuss the technique in general and show how it gives fresh insight into the connections between entropy and coding.

Session 2

Covering Radius Problems in Algebraic Coding

On the Covering Radius of Linear Codes

S.M. DODUNEKOV, Bulgarian Academy of Sciences, Bulgary

A survey and recent results about the covering radius of linear codes will be presented. Special attention will be paid to the covering radius of cyclic codes and to some combinatorial functions with the covering radius.

Covering Radius of Shortened Codes and Applications

G. COHEN, ENST, Rue Barrault 75634, Paris Cedex 13, France

Let C be a given $[i,k,d]$-code. We denote by $k(i,d,C)$ the maximal dimension of a code of length i and distance d containing C. We set $k(i,d) = k(i,d,\{\underline{0}\})$.

Proposition 1: Any code $C[n,k,d]$ has a covering radius ρ satisfying: $k(\rho, d) + k \leq k(n,d,C)$.

Now we use proposition 1 to upperbound $\rho(C(S))$, where $C(S)$ is a t-error correcting BCH code shortened up to length $|S|$.

Proposition 2: For $t = 2$

i) if $|S| > (n/2) + n^{1/2}$, then $\rho(C(S)) \leq 9$

ii) if, moreover, $|S| > (\sqrt{2}/2)n$, then $\rho(C(S)) \leq 7$.

Proposition 3: For $t = 3$

i) if $|S| > n/2 + n^{1/2}$, then $\rho(C(S)) \leq 13$

ii) if $|S| > 2^{1/3}n/2$, then $\rho(C(S)) \leq 12$

iii) if $|S| > 2^{2/3}n/2$, then $\rho(C(S)) \leq 10$.

As an application, we construct easily decodable error-correcting WOM-codes. This is joint work with G. Zemor.

Bounds for Non-Linear Covering Codes

I. HONKALA, Department of Mathematics, University of Turku, SF-20500 Turku 50, Finland

Denote by $K(n,R)$ the minimum number of codewords in a binary code of length n and covering radius R. We discuss some lower bounds on $K(n,R)$ which use results about the classical combinatorial problem of covering pairs by k-tuples. We also mention results about normality and subnormality of binary codes with covering radius one, and give some lower bounds on the minimum cardinality of normal binary codes with covering radius one. Finally, we briefly review the forthcoming paper by Hekki Hamalainen and Seppo Rankinen about upper bounds for binary/ternary mixed covering codes.

Decoding of the Nordstrom-Robinson Code up to the Covering Radius

S.N. LITSYN, Perm Agricultural Institute, Dept of Computer Science

For the maximum likelihood (ML) decoding of codes in binary symmetric channels, it is necessary to correct error patterns of the Hamming weight up to the covering radius of the code. Usual ML decoding algorithm for the first-order Reed-Muller codes is based on the fast Hadamard transform and has complexity $n \log_2 n$ additions, n comparisons and n absolute value calculations (n is the length of the code). Further improvement of this algorithm is described. It requires $n \log_2 n - 1.25n + 2$ additions, $1.25n + 1$ comparisons and $n/2 + 1$ absolute value calculations.

If we want to decode a code consisting of the union of cosets of the first-order Reed-Muller code, we can use decoding in each coset up to the covering radius of the initial code that is not necessary to be ML decoding in the coset. This idea combined with the decoding algorithms of the first-order RM code with linear complexity and ideas of mentioned above algorithm give us the ML decoding algorithm requiring only 88 additions.

Covering Radius and Perfect (Mixed) Codes

G.J.M. VAN WEE, Eindhoven University of Technology, Dept. Mathematics and Comp. Science, PO Box 513, 5600MB Eindhoven, The Netherlands

Let C be a (linear or nonlinear) binary code of length n. The number $R := \max\{d(\mathbf{x},C) \mid \mathbf{x} \in \mathbf{F}_2^n\}$ is called the *covering radius* of C. In /1/, a method was given to improve (generally) on the trivial bound

$$|C| \geq \frac{2^n}{1 + \binom{n}{1} + \cdots + \binom{n}{R}}.$$

The idea was to consider the spheres with radius 1, centered at words at maximal distance from the code, and count properly the number of intersections over there of spheres with radius R centered at codewords. For instance, it followed that $|C| \geq 2^n/n$ if $R = 1$. This idea, or variations on it, also works for *nonbinary* codes, even for *mixed* codes, see /3,4/.

So we get various lower bounds on covering codes. In /3/, it was also pointed out that analogously one derives upper bounds on R-error-correcting codes. Another application is on perfect mixed codes, see /2/. Herzog and Schoenheim (1972) gave examples of nontrivial perfect mixed codes (with $R = 1$). In /2/, a general nonexistence theorem for perfect mixed codes was proved.

References

/1/ G.J.M. van Wee, 'Improved sphere bounds on the covering radius of codes', IEEE Trans. Inform. Theory, vol 34, pp. 237-245, 1988.

/2/ G.J.M. van Wee, 'On the non-existence of certain perfect mixed codes', Discr. Math., to appear.

/3/ G.J.M. van Wee, 'Bounds on packings and coverings by spheres in q-ary and mixed Hamming spaces', submitted to J. Comb. Theory.

/4/ J.H. van Lint Jr and G.J.M. van Wee, 'Generalized bounds on binary/ternary mixed packing and covering codes', submitted to J. Comb. Theory.

This research was supported by the Netherlands organization for scientific research (NWO).

New Bounds for Codes and Lattices

N.J.A. SLOANE, Mathematical Sciences Research Center, AT&T Bell Laboratories, Murray Hill NJ 07974, USA

Many of the best codes and lattices are 'self-dual' (the Hamming, Golay, and quadratic-residue codes; the simple cubic, E_8 and Leech lattices, etc.). This talk, which is based on joint work with John Conway, will describe new bounds for such codes and lattices. For example, the minimal squared length in self-dual lattice cannot exceed about one-tenth of the dimension.

I shall also briefly describe the work that R.H. Hardin, W.D. Smith, and I have been doing over the past couple of years on placing points 'uniformly' on the surface of an n-dimensional sphere. More precisely, (i) how to place M points so as to minimize the distance from any point to one of the code-points; (ii) how to place M points so as to maximize the minimal distance between them? (the packing problem).

Session 3

Convolutional Codes

From a Proof to a Product

A.J. VITERBI, Qualcom, San Diego, USA

Twenty-four years ago in March, I proposed a simple algorithm for decoding convolutional codes, more as a didactic tool to reveal and exploit the natural state-machine nature of the encoder than as a practical implementation. Since then, the industrial (even more than the academic) telecommunications community has adopted it as an integral block in most wireless digital communication systems.

This talk will trace the evolution of the decoder from early deep space applications, through government and eventually commercial satellite systems, to present-day mobile cellular radio. The gradual enhancement of the theory as well as of the hardware over two decades will be reviewed, with emphasis on the speaker's predilections.

Multi-Level Trellis Coded Modulation and Multi-Stage Decoding

D.J. COSTELLO, J. WU, AND S. LIN Dept Electrical and Computer Engineering, University of Notre Dame, IN 46556, USA, and University of Hawai, USA

New multi-level trellis coded modulation schemes using generalized set partitioning are developed for QAM and PSK signal sets. Several constructions are presented and many multi-level codes with better performance than previously known codes are found. These codes provide a flexible trade-off between coding gain, decoding complexity, and the decoding delay.

Necessary and sufficient conditions for a code to be rotationally invariant are derived. It is shown that most previously known multi-level codes are not rotationally invariant. Moreover, even rotationally invariant multi-level codes cannot always be combined with differential encoding to resolve phase ambiguity, and

additional information bits are needed as phase reference. In this paper, we present some new rotationally invariant multi-level codes which can be combined with differential encoding to resolve phase ambiguity. (This work was supported by NSF Grant NCR80-03429 and NASA Grant NAG5-557).

Linear Algebraic Formulation of the Viterbi Algorithm

H. MEYR AND G. FETTWEIS, Technical University Aachen, W.-Germany

The main part of the Viterbi algorithm is a nonlinear data dependent recursion, in which add-compare-select operations are performed. A new algebra will be described which allows to rewrite the recursion as a linear recursion. This result opens new interesting viewpoints and possibilities, both for theory and practice. The Viterbi algorithm can now be examined as a linear system, allowing linear algebraic transformations as well as superpositions etc. The impact of this leads to a deeper insight and allows the derivation of new algorithmic modifications. Since from a conventional point of view the recursion was nonlinear and data dependent, it was a bottleneck for high-speed parallel VLSI-implementations. The linear algebraic formulation now allows us to employ powerful techniques of parallel processing and of pipelining, known for conventional linear systems, to achieve high throughput rates. VLSI design examples show that high speed Viterbi decoders for more than 100 Mbit/s can be implemented on VLSI integrated circuits to date.

Algebraic-Sequential Decoding of Convolutional Codes

K.SH. ZIGANGIROV AND D.J. COSTELLO, Institute for Problems of Information Transmission, USSR Academy of Sciences, Moscow, USSR, and Dept Electrical and Computer Engineering, University of Notre Dame, IN 46556, USA

A new decoding algorithm for convolutional codes is proposed. It is a sequential decoding algorithm which uses the algebraic properties of a code over $GF(q)$. The algorithm is particularly useful for large q. To detect the error positions, it uses sequential decoding, and to find the error values, it solves an algebraic system of linear equations. We show that the computational cut-off rate of the new algorithm is larger than the computational cut-off rate of conventional decoding. The algorithm can be applied to the decoding of trellis codes over q-ary signaling alphabets, and to the decoding of concatenated codes using a q-ary outer convolu-

tional code. (This work was supported by NSF Grant NCR80-03429 and NASA Grant NAG5-557).

The Matching of Modulation Types with Ring Convolutional Codes

J.L. MASSEY AND H.-A. LOELIGER, Institute for Signal and Information Processing, ETH, CH 8092 Zuerich, Switzerland

The metric structure of M-ary phase modulation is matched to the algebraic structure of the ring of integers modulo M, Z_M, in the sense that when the M signals of the phase-modulation signal set are associated with the elements of Z_M in a natural way, the squared Euclidean distance between signals is a function only of the differences of the corresponding elements. This fact motivated the recent investigations of ring convolutional codes at the Swiss Federal Institute of Technology and the subsequent discovery of some ring convolutional codes for phase modulation that are rotationally invariant and possess large free Euclidean distance. It will be shown that there are other modulation signal sets, some quite strange, that are similarly matched to the algebraic structure of Z_M. The notion of metric/algebraic matching will be extended to modulation types with memory, such as the various continuous-phase modulations. It will be shown that for every M, there is an M-ary continuous-phase frequency-shift-keying (CPFSK) modulation matched to Z_M. For $M = 2$, this CPFSK modulation is the familiar minimum-shift-keying (MSK) modulation, but the other CPFSK modulations in this family appear to be new.

A brief review of the pertinent aspects of the theory of convolutional codes over rings will be given. Some newly constructed ring convolutional codes for the modulation signal sets considered previously will be presented. Codes will be presented for the new CPFSK modulations that exploit the state of the modulator by making it part of the encoder state, thereby minimizing the number of states in the corresponding Viterbi demodulator/decoder.

Session 4

Artificial Neural Networks

The Complexity of Classification Problems

GEORGE CYBENKO, Department of Electrical and Computer Engineering and Center for Supercomputing Research and Development University of Illinois at Urbana Urbana, IL 61801, USA, E-mail: gc%sp12@csrd.uiuc.edu

Much current research on artificial neural networks is devoted to techniques such as multilayered feedforward networks and backpropagation and other algorithms. At the same time, there has been little work on understanding the complexity of classification problems independently of the technique used to solve them. To study such complexity issues, we believe that it is necessary to study the difficulty of representing functions when the allowable operations are constrained in some way.

We start by surveying some notions introduced years ago by Kolmogorov, Vitushkin and others to quantify the complexity of functions. It will be shown how these notions can be used to derive a new notion of function complexity according to the difficulty of generating a sigma algebra whose measurable functions approximate the given function well. We present some properties of this notion of function complexity and relate it to neural networks and a given architecture's ability to solve a problem.

Finally, we will present some preliminary experiments that correlate these complexity notions to classical hypotheses of recognition difficulty in speech problems as measured by articulatory features.

A Markovian Generalization of Valiant's Learning Model

UMESH VAZIRANI, University of California at Berkeley, Department of Computer Science, Berkeley, California 94720, USA, E-mail: vazirani@ernie.berkeley.edu

We give a generalization of Valiant's distribution free model of learning: in our model the positive and negative examples of the concept to be learnt are placed on

the vertices of a large (arbitrary) graph. The learner performs a random walk on this graph (a reversible Markov chain in general), and tries to learn the concept based on the data encountered. Valiant's model is the case when the walk is performed on the complete graph. The advantage of the generalization is that the model can express short term correlations in the observed data (e.g. two successively observed examples of chairs are likely to look similar).

A significant advance in our understanding of induction followed from Valiant's distribution free learning model, combined with the work on Occam algorithms by Blummer et. al. Justifying the process of induction is not only a important philosophical problem, but also an issue that must lie at the core of any theory of learning. We study the question of whether induction can be justified when the data is picked by a random walk on a graph. This question can be posed as a natural problem about random walks on graphs; whereas we conjecture that the answer is always affirmative, we are able to prove this for a special case. Our result also establishes the following theorem which is of independent interest: Consider a random walk on an arbitrary n-vertex graph G whose vertices are labeled with (arbitrary) integers. Then the expected length of the lexicographically first increasing subsequence of integers encountered in a random walk of length t is at most $\sqrt{t} \log n$. This is joint work with David Aldous.

Folklore and Mathematics: What is the Capacity of a Neural Network to Learn?

SANTOSH S. VENKATESH, Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA, E-mail: venkatesh@.ee.upenn.edu

The concept of capacity is a fundamental one in the study of neural networks. On one hand, the computational capacity of a neural network estimates the largest number of instances of an arbitrary function that can be guaranteed to be loaded into the network -in a sense it is a measure of the largest problem size the network can handle reliably-; on the other hand, the capacity also turns out to be a critical parameter characterizing the ability of a network to learn and generalize from a set of examples of an otherwise unspecified function. In this paper we study computational and learning attributes of neural networks using two notions of capacity which have turned out to be most useful: random capacity and the VC-dimension. In the light of formal mathematical developments we will examine several commonly held notions ("folk theorems") on neural computation; among them: neural networks are fault-tolerant (robust); binary interconnections between

neurons suffice; the Hebbian rule can store 0.15*n memories; shallow networks are more efficient; and learning is hard.

Polynomial Time Neural Net Learning Algorithms

ERIC B. BAUM, NEC Research Institute, 4 Independence Way Princeton, NJ 08540, USA, E-mail: eric@research.nec.com

We describe some neural net training algorithms which are provably able to PAC learn (Valiant, 1984) in polynomial time classes of functions described by restricted but highly non-trivial classes of feedforward, depth 2 threshold networks. We also describe experiments in which variants of these algorithms rapidly learned target functions composed of random, layered, feedforward threshold networks with 200 inputs completely connected to 200 hidden units completely connected to an output unit or units.

Some Estimation and Approximation Theorems for Artificial Neural Networks

ANDREW R. BARRON, Dept. of Statistics and Electrical & Computer Eng., University of Illinois, 725 S.Wright Street, Champaign, IL 61820, USA, E-mail:barron@andrew.stat.uiuc.edu

We build feedforward networks using either sigmoidal or polynomial nodes as general-purpose response surface models. The issue we address is the proper use of training data to automatic select the size and structure of the network as well as to estimate the coefficients of the network. A theoretically attractive but practically infeasible approach is to globally optimize a complexity-penalized performance criterion. We discuss bounds on the statistical risk (generalization capability) that have been established in this case based on the index of resolvability.

A practical approach to the empirical construction of feedforward networks starts with one node and incrementally adds new nodes in a selection process guided by the performance criterion. Approximation-theoretically, these greedy algorithms for network construction are suboptimal compared to the global optimization of network structure. Nevertheless, recently developed theory demonstrates that satisfactory approximation can be obtained. In particular, in the case of integrated squared error, conditions are given such that the approximation error achieved by a k-step incremental synthesis (adding one node at each step) is within order $1/k$ of

the error achieved by the best network of size k in a given class of networks. Implications of these approximation results for statistically estimated networks are discussed.

Session 5

Shannon Theory

Multiterminal Filtering for Decentralized Detection Systems

K. KOBAYASHI AND T.S. HAN, Dept. Computer Sciences and Information Mathematics, University of Electro-Communications, Chofu, Tokyo 182, Japan, and Dept. Information Systems, Senshu University, Kawasaki 214, Japan

Recently there has been a growing interest in distributed detection systems. Systems made up of distributed sensors (seismographs, antennas, hydrophones, etc.) receive corrupted signals and have to transmit preprocessed data for inputs to a fusion center through channels of finite capacity. The fusion center is required to give the best decision on the existence of known or unknown signal under the constraints on transmission rates of channels.

In their paper, Tsitsiklis and Athans point out that the problem of synthesis of optimum decentralized detection system is NP-hard in general. After that, most research was restricted to the situation of independent noise, and/or concentrated to find suboptimum solutions by ad hoc methods.

It is quite important to recognize that the essential aspect in this kind of distributed detection problem will be clear by treating the problem asymptotically, and taking account of the correlation structure of probabilistic law itself.

Especially, in the additive Gaussian noise situation, we show that the optimum performance that would be obtained with no restriction on transmission rates, can be attained at essentially zero rate constraints by constructing a multi-terminal filter making full use of the correlation effects between the data into remote sensors. Moreover, we can use the Wiggins-Robinson fast algorithm for block Toeplitz matrix to construct our multi-terminal filter. It should be noted that our postulated correlated sources belong to a class of decomposable sources:

$$p^{\theta}_{X_1 X_2 \ldots X_M}(x_1, x_2, \ldots, x_M) = q_{X_1 X_2 \ldots X_M}(x_1, x_2, \ldots, x_M)\rho^{\theta}_{X_1}(x_1)\rho^{\theta}_{X_2}(x_2) \ldots \rho^{\theta}_{X_M}(x_M), \quad \theta \in \Theta.$$

On the other hand, in order to establish the optimum decentralized detection system for non-decomposable sources, we might need positive rates and many coding problems remain unsolved.

Capacity and Coding Theorems for the Discrete Noiseless Channel with Specified Cost-per-Symbol

B. MARCUS AND S. TUNCEL, IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA, and Mathematics Dept, University of Washington, Seattle WA 98195, USA

Let G be a finite directed graph and f be a vector-valued (cost) function defined on the edges of G. Define the cost per symbol of a cycle in G to be the average f-value (cost) of the edges that it traverses. We give a formula for the asymptotic growth rate of the number of cycles with specified cost per symbol. This generalizes Shannon's classical formula for the capacity of a discrete noiseless channel. It is also closely related to several recent papers on the capacity of a discrete noiseless channel with specified upper bound on the cost-per-symbol.

We use this formula as a tool for generalizing coding theorems in symbolic dynamics to coding theorems for Markov chains. In particular we consider two coding relations: by an imbedding of one Markov chain, P, to another, Q, we mean a 1-1 sliding block code from the sequences of P to the sequences of Q that preserves average log transition probabilities around any cycle, up to a uniform constant; by a finite-state code, we mean the same thing except that the 1-1 sliding block code is replaced by a finite-state encoder. We give sufficient conditions and constructions for imbeddings/finite state codes.

A Simple Proof of the Equality of the Maximal-Error and Average-Error Capacity Region for Broadcast Channels

F. WILLEMS, Technical University of Eindhoven, The Netherlands

A proof that for broadcast channels the maximal-error capacity region is equal to the average-error capacity region can be found in Csiszar and Koerner's book (1981). This proof however is far from being direct. By making an excursion to "stochastic encoders" and using Ahlswede's "elimination of correlation" technique (1978) they succeed in establishing the equality.

We give a direct proof of this equality, thereby avoiding the use of stochastic encoders. Instead we introduce a matrix-partitioning technique which enables us, at the expense of a negligible decrease in rate, to select those codewords for which the probability of error is acceptably low.

Reusable Memories in the Light of the old AV- and a New F-channel Theory

R. AHLSWEDE AND G. SIMONYI, University of Bielefeld, Germany

Arbitrarily varying channels were introduced as a model for transmission in cases of jamming. We show that their theory also naturally applies to memories and yields in a unified way some new and old Capacity Theorems for write-directional (and more general) memories with side information. If encoder and decoder have any side information it is still not understood what the optimal rates for many cycles are. We expect more insight from a theory of F-channels, which we have introduced for that purpose and started to analyse.

Deterministic Arbitrarily Varying Channels and List Decoding

M. PINSKER, Institute for Problems of Information Transmission Moscow, USSR

The capacity of some classes of deterministic arbitrarily varying channels is considered. The main attention will be paid to channels for which the input signals and the channel states form ellipsoids in Euclidean spaces. The capacity of such channels for random and deterministic codes is discussed. The possibility to use erasure and list decoding in arbitrarily varying channels is shown.

Arbitrarily Varying Channels

I. CSISZAR AND P. NARAYAN, Electrical Engineering Department and the Systems Research Center, University of Maryland, College Park, MD 20742, USA, and Mathematical Institute of the Hungarian Academy of Sciences, H-1364, POB 127, Budapest, Hungary

An arbitrarily varying channel (AVC) is a model of a communication channel with unknown parameters which may vary with time in an arbitrary and unknown manner during the transmission of a codeword. We consider the capacity of an AVC, with finite input and output alphabets and a finite set of states, for the average probability of error criterion. It is shown that this capacity is zero if and only if the AVC is symmetrizable; else, it equals the random coding capacity of the

AVC. Capacity is also determined by constraints on the transmitted codewords as well as on the channel state sequences, when it may be positive but less than the random coding capacity. Simple decoding rules are identified that attain the capacity of the noiseless binary (resp. arithmetic) adder AVC, whose output is the binary (resp. arithmetic) sum of the input and the state.

The talk will conclude with a brief discussion of recent results due to John Gubner on the capacity region of a two-sender multiple-access AVC.

Session 6

Recording Systems

Runlength Limited Codes for Mixed-Error Channels

O. YTREHUS, University of Bergen, Dept. Informatics, Thormohlenset 55, Bergen N-5008, Norway

The nature of errors in a magnetic recording channel is the matter of some discussion. The Binary Symmetric and the Peak Shift Channels are two of the "theoretic" channel models that have been proposed. However, the (small amount of) empirical evidence available suggests that a real channel is some combination of the two. On the Mixed-Error Channel, an error is either a Binary Symmetric error or a peak shift.

The block codes that are presented are (d,k)-constrained, and are t-error correcting on the Mixed-Error Channel. The code construction is an adaptation of the well-known technique of generalized convolutional codes. The typical code construction involves two outer codes:

O_1, which is a t-error correcting BCH code over GF(2^2) of block length N and dimension K, and

O_2, which is a Reed-Solomon code over GF(2^m) of length N and minimum distance $(t+1)$.

There are four disjoint inner codes of block length n, $I_0, \ldots, I_3$, so that each inner code is (d,k)-constrained, has 2^m codewords, and has minimum shift *and* Hamming distances of at least two.

The resulting code is (d,k)-constrained, corrects t errors on the Mixed-Error Channel, and has code rate

$$R = \frac{2K + m(N-t)}{Nn}.$$

The benefits of this new technique, as compared to a plain concatenation of a modulation block code and a Reed-Solomon code, are a reduction in decoding complexity and a slightly higher code rate. An alternative way of using this con-

struction is to construct lower bounds the size of a (d,k)-constrained t-mixed-error correcting block code. Specifically, codes are constructed that have higher rates than what can be achieved by using concatenations of current state-of-the-art sliding block codes and Reed-Solomon codes. (This research was supported by the Norwegian Council for Science and the Humanities, NAVF).

Bounds on Information Rates for the Peak-Shift Magnetic Recording Channel

S. SHAMAI (SHITZ), E. ZEHAVI, AND G. KAPLAN Dept of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel

A simple statistical model is suggested to account for single position peak (bit) shifts which were identified to be one of the major impairments in magnetic recording, employing peak detectors.

We investigate the capacity and the cut-off rate for this channel, where the channel inputs are the (d,k) codes, commonly used in magnetic recording. For $d \geq 2$, this channel is conveniently described in terms of phrase lengths where a bit shift causes a phrase either to shrink or to expand. The inherent correlation present in consecutive shift affected phrases manifests itself in memory which is introduced into the channel model. Sequences of nondecreasing lower bounds and nonincreasing upper bounds on the capacity are presented and investigated for a variety of interesting parameters. Lower bounds on the cut-off rate are evaluated as well and compared to the corresponding lower bounds on the capacity. Lower bounds on the zero-error capacity and the zero-error cut-off rate are also studied.

The channel model is extended to a concatenated scheme composed of a peak shift channel connected in tandem with the binary symmetric channel, capturing thus both major error generating mechanisms-the peak shifts as well as randomly generated errors. Lower and upper bounds on the capacity of this concatenated channel model with (d,k) input sequences are derived and compared to the corresponding lower bounds on the cut-off rate.

Combinatorial Bounds and Constructions for Error Correcting Constrained Codes

H.C. FERREIRA, Rand Afrikaans University, Cybernetics Laboratory, Fac. Eng., PO Box 524, Johannesburg 2000, South Africa

We present an overview and summary of some results achieved during recent years on the subject of error correcting runlength constrained or balanced codes /1-4/. These results involve combinatorial constructions, as opposed to the algebraic constructions usually employed for linear error correcting codes.

We first establish Gilbert type lower bounds on the minimum Hamming distance achievable with binary and ternary block codes which are balanced (and hence dc-free), or which are minimum and maximum runlength constrained /1,2/. These bounds also prove the existence of error correcting constrained codes.

Next we present a combinatorial construction for an error correcting balanced block code, synthesized with 2-bit tuples as elements /1/. This code prompted several other researchers to come up with improved constructions.

Error correcting constrained trellis codes are next investigated. We present the transformation of a linear convolutional code into a constrained trellis code with the same or larger distance properties /3/. In order to achieve such a transformation, the concept of a Hamming distance preserving mapping of the set of unconstrained binary symbols of the convolutional code, onto a set of symbols with the desired constraints, is introduced. Simple tests to determine if such mappings exist, and a tree search algorithm for finding such mappings are presented. The results obtained in this way are compared to related published work which also include several *ad hoc* codes.

Finally, we briefly describe techniques for systematically constructing different runlength constrained binary block codes capable of detecting and correcting single bit errors, peak shift errors, double adjacent errors, and multiple adjacent errors /4/. Contrary to some other reported block codes, the longer codes constructed in this way can be encoded and decoded with simple, structured logic circuits.

References

/1/ H.C. Ferreira, 'Lower Bounds on the Minimum Hamming Distance achievable with Runlength Constrained or DC free Block Codes and the Synthesis of a (16,8) $D_{min} = 4$, Dc-Free Block Code', IEEE Trans. Magn., vol. MAG-20, pp. 881-883, Sept. 1984.

/2/ H.C. Ferreira, J.F. Hope, and A.L. Nel, 'On Ternary Error Correcting Line Codes', IEEE Trans. Commun., vol. COM-27, pp. 510-515, May 1989.

/3/ H.C. Ferreira, D.A. Wright, and A.L. Nel, 'Hamming Distance Preserving Mappings and Trellis Codes with Constrained Binary Symbols', IEEE Trans. Inform. Theory, IT-35, pp. 1098-1103, Sept. 1989.

/4/ H.C. Ferreira and S. Lin, 'Error and Erasure Control (d,k) Block Codes', Submitted to IEEE Trans. Inform. Theory, 1990.

Coding and Data Detection in a Gigabit per Square Inch Magnetic Recording System

T.D. HOWELL, IBM Almaden Research Center, 650 Harry Road, K65/802, San Jose, CA 95120, USA

A magnetic disk system which records 1 Gb/in^2 has recently been demonstrated. It uses partial response class 4 signalling and maximum likelihood detection in place of the more conventional peak detection. Two aspects of the partial response channel which differ substantially from conventional channels are presented.

One aspect is the code. The partial response channel does not require that the magnetic transitions be kept far apart to avoid intersymbol interference as is done in peak detection channels. The lower run length constraint commonly used for this purpose is eliminated, and the code rate is increased from 1/2 or 2/3 to 8/9. An upper run length constraint is still needed to guarantee enough feedback in the timing and gain control loops. The code we used has run length constraints (0,4).

The class 4 partial response $(1 - D^2)$ channel can be viewed as two interleaved dicode $(1 - D)$ channels. The detection process operates independently on the two interleaved data subsequences. Additional constraints are imposed on the run lengths of zero samples in the interleaved subsequences: No more than four consecutive zeroes are allowed in each subsequence. This constraint limits the required length of the survivor sequence memories in the Viterbi decoder to about five stages each. Without the constraint about thirty stages would be required. The decoding delay is reduced by the same factor.

Another interesting aspect of the channel concerns the implementation of the Viterbi detector. We use a form of the Viterbi algorithm in which the difference between two path metrics is recursively updated at each clock interval. This path metric difference is easily seen to be bounded in absolute value by $L + 1$, where the samples produced by the analog-to-digital (A/D) converter range from $-L$ to L. It is most convenient to let L be a power of two so that certain thresholds used in the recursive update calculation correspond to simple bit patterns. This appears to require that the registers holding the path metric differences be one bit longer than the registers holding samples from the A/D converter. We present a theorem showing that no extra register width is required: with proper initialization, the metric differences can never exceed $L - 1$ in absolute value.

Trellis Codes and Sequence Estimation for Recording Channels

CHRIS HEEGARD AND MIGNON BELOGNIE, School of Electrical Engineering, Cornell University, 301 Phyllips Hall, Ithaca, NY 14853, USA heegard@ee.cornell.edu

This talk concerns the effectiveness of trellis coding and sequence estimation ideas applied to recording channels. The current approach to data detection in recording systems is to incorporate a "sampling detector" or a "peak detector" to extract an estimate of the encoded data (i.e., code sequence) followed by an "unencoder" (i.e., a sliding window mapping of the estimated code sequence onto data symbols). In many other communications systems, significant improvements in performance have been obtained by methods of sequence estimation; it is the goal of this report to indicate how such ideas can be used in recording systems.

The approach we have taken involves a variable length description of the run-length limited coding and decoding. We use a variable length trellis decoding to explore the effectiveness of various coding strategies on a Lorentz model for the magnetic recording channel. To account for the long memory of the system, we have combined these ideas with the Delayed Decision-Feedback Sequence Estimation. The approach can be easily extended to the problem of sequence estimation on more realistic magnetic and optical recording systems.

We show that for high density recording the typical run-length codes used in practice (i.e., "(2, 7)" and "(1, 7)") are ineffective in a sequence estimation environment. High rate codes which only limit the maximum run-lengths perform much better. This leads one to ponder how one should choose a code when sequence estimation is incorporated. One answer is the notion of a *Trellis Constrained, Run-length Code.* In such a code, in addition to the limiting of the maximum and minimum run-length, a constraint is imposed that insures that encoded sequences can be reliably distinguished by a sequence estimator. We argue that for densities of interest, significant improvements in performance can be obtained with this approach. (This work was supported in part by NSF grants NCR-8903931, ECS-8352220, IBM & AT&T.)

Spectral Null Codes and Number Theory

P.H. SIEGEL, IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA. The author is on sabbatical at the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093

A code with spectral null of order K at frequency f_0 is a collection of sequences S, over a finite, complex alphabet $A \subset C$, having the property that the average power spectral density $\Phi(f)$ and its derivatives $\Phi^{(k)}(f)$, $k = 1, \ldots, 2K-1$, assume the value zero at $f = f_0$, where f_0 is a rational submultiple of the symbol frequency. These codes underlie the technique of matched-spectral-null trellis coding /1/ which can improve the reliability of digital data recording and transmission over noisy, partial-response channels when the spectral null frequencies of the code and channel coincide.

The characterization of spectral null codes /2,3,1,4/ and the evaluation of their distance properties /5,1,4/ make use of several concepts and nontrivial results from number theory and algebra, including: cyclotomic polynomials; Newton's identities; equal-power-sum sets; quadratic residues, Legendre symbols, and the Gaussian sum formula; and the familiar Chinese remainder theorem. In return, the theory of spectral null codes has suggested some generalizations of these mathematical tools (possibly already known to experts).

This talk will describe some of these intriguing connections, and explore their possible relation to the spectral interpretation of error-control codes over finite fields /6/.

References

/1/ R. Karabed and P. Siegel, 'Matched-spectral-null codes for partial-response channels,' IEEE International Symposium on Information Theory, 1988; IEEE Information Theory Workshop, 1989; U.S. Patent 4,888,779, issued December 19, 1989; IBM Research Report, March 1990.

/2/ B. Marcus and P. Siegel, 'On codes with spectral nulls at rational submultiples of the symbol frequency', IEEE Trans. Info. Th., vol. IT-33, No. 4, pp. 557-568, July 1987.

/3/ C. Monti and G. Pierobon, 'Codes with a multiple spectral null at zero frequency', IEEE Trans. Info. Th., vol. IT-35, No. 2, pp. 463-471, March 1989.

/4/ E. Eleftheriou and R. Cideciyan, 'On codes satisfying M-th order running digital sum constraints', IBM Research Report, March 1990.

/5/ K.A.S. Immink and G. Beenker, 'Binary transmission codes with higher order spectral zeros at zero frequency', IEEE Trans. Info. Th., vol. IT-33, No. 3, pp. 452-454, May 1987.

/6/ R.E. Blahut, *Theory and Practice of Error Control Codes*, Addison-Wesley, 1983.

Session 7

Cryptology

Towards a Coding Theorem in Authentication Theory

ANDREA SGARRO, University Udine, Italy

Rather strong bounds now available on the probability of successful impersonation suggest the possibility of obtaining, for authentication codes, a coding theorem in the usual sense of Shannon theory. In particular, a normative performance parameter should be obtained whose meaning is the "capacity of resistance against impersonation". This problem is tackled, is shown to be equivalent to a graph-theoretic problem, and a partial solution is offered.

A Construction of Authentication Codes from Geometries of Classical Groups over Finite Fields

ZHE-XIAN WAN AND BEN SMEETS, Gr.Sch.Acad.Sci. Beijing, P.R. China, and University of Lund, PO Box 118, S-221 00 Lund, Sweden

Traditionally, only geometries of the linear group over finite fields have been investigated for use in constructing authentication codes. Constructions of new authentication codes will be given that are based on the use of geometries derived from other classical groups over finite fields.

An Information-Theoretic Approach to Stream Ciphers

CEES J.A. JANSEN AND DICK E. BOEKEE, Technical University of Delft, PO Box 5031, 2600 GA Delft, The Netherlands, and Philips Crypto BV, PO Box 218, 5600 MD Eindhoven, The Netherlands

that the equivocation $H(S|Z) \simeq k - 1$ bits $= H(S) - 1$. Since the encoding is done by generating k random bits, the system is secure to the extent that the random bit generator is truly random. (Joint work with L.H. Ozarow)

In 1949, Claude Shannon, the founder of information theory, published his paper 'Communication Theory of Secrecy Systems', /6/. his paper was a first inception to

Algebraic Coding for the Wire-Tap Channel

VICTOR K. WEI, Bellcore, 435 South Street, Morristown, NJ 07960, USA

Wyner and Ozarov-Wyner studied wire-tap channels of types I and II. They derived the channel capacities and studied several coding schemes with emphasis on asymptotics. The coding problem for wire-tap channels will be treated here in a more algebraic manner. It will be shown in particular that the performance of linear codes over the type II channel can be described in terms of a generalized Hamming weight. Several bounds on code parameters are derived from this perspective, and the exact hierarchy of corresponding generalized Hamming weights for Hamming codes and for Reed-Muller codes of all orders is determined. The proofs utilize the Kruskal-Katona Theorem from finite-set combinatorics. The obtained results can also be applied to several other cryptographic problems including bit extraction or t-resilient functions, known-plaintext attacks in public-key encryption, and variations of secret-sharing schemes.

A New Information-Theoretic Notion of Cipher Security and a Provably-Secure Randomized Cipher

UELI M. MAURER, Institute for Signal and Information Processing, Swiss Federal Institute of Technology, CH 8092 Zurich, Switzerland

A Cipher is defined to be *perfect with probability at least p* if $I(X;Y|A) = 0$, where X and Y denote the plaintext and cryptogram, respectively, and where A is an event that occurs with probability $P(A) \geq p$. By a strongly-randomized cipher, we mean that a random string is publicly available whose length is much greater than that of the plaintext. A strongly-randomized cipher will be exhibited whose secret key Z is small (compared to the length of the plaintext), i.e., that satisifies $H(Z) \ll H(X)$, and that is provably perfect with probability very close to 1 when the enemy's computational power is limited in a reasonable way. This result is somewhat surprising because, according to Shannon, perfect secrecy, i.e., $I(X,Y) = 0$, implies that $H(Z) \geq H(X)$. Moreover, although information-theoretic security usually implies that the enemy has infinite computing power, the constructed cipher is secure for an information-theoretic notion of security only when the enemy is computationally restricted.

Session 8

Source Coding

The Redundancy of the Ziv-Lempel Algorithm for Memoryless Sources

YURI M. SHTARKOV AND TJALLING J. TJALKENS, Institute for Problems of Information Transmission, Ermolovoy str. 19., 101447, Moscow, GSP-4. USSR, and Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

We discuss the redundancy of the Ziv-Lempel universal source coding algorithm. We consider this algorithm as a variable-to-fixed coding scheme and analyse its redundancy for memoryless sources as a function of the codeword length L. We show that the redundancy decreases as $O(1/\log L)$, whereas the optimal rate of decrease is of $O(\log(L)/L)$. Of course we cannot infer from this the behaviour of the redundancy for sources with memory, but to us, it is not obvious that the algorithm should have an optimal rate of convergence of the redundancy in this case. Because any reasonable measure of complexity will show an increase of complexity with the number of codewords we are of the opinion that the Ziv-Lempel algorithm will not be the optimal algorithm when we consider the redundancy as a function of the 'complexity'.

Variable-to-Fixed Length Codes Have Better Large Deviations Performance than Fixed-to-Variable Length Codes

NERI MERHAV AND DAVID L. NEUHOFF, AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974, USA

It is shown that for finite-alphabet, finite-state sources, the best V-F code provides a better large deviations performance than any F-V encoder with the same number of codewords. Specifically, we introduce a random variable, referred to as the empirical compression ratio (ECR), which is defined as the length (in bits) of the

/3/ C.J.A. Jansen, 'Investigations On Nonlinear Streamcipher Systems: Construction and Evaluation Methods, PhD. Thesis, Technical University of Delft, Delft, april 1989.
/4/ A. Lempel and J. Ziv, 'On the Complexity of Finite Sequences', IEEE Trans. on Info. Theory, vol. IT-22, no. 1, pp. 75-81, Jan. 1976.
/5/ R.A. Rueppel, 'New Approaches to Stream Ciphers', PhD. Thesis, Swiss Federal Institute of Technology, Zurich, 1984.
/6/ C.E. Shannon, 'Communication Theory of Secrecy Systems', Bell Systems Technical Journal, vol. 28, pp. 656-715, Oct. 1949.
/7/ H.E. Wanders, 'On the Significance of Golomb's Randomness Postulates in Cryptography', Philips Journal of Research, vol. 43, no. 2, pp. 185-222, 1988.

The Statistical Dependence between Output and Input of a Nonlinear Combiner with Feedback

INGEMAR INGEMARSSON AND AMUND HUNSTAD, Dept. Elec. Eng., Linkoping University, S-581 83 Linkoping, Sweden

Siegenthaler has shown that a memoryless combination of random binary sequences in general yields a statistical dependence between input and output bits of the combiner. Rueppel later showed how this can be avoided by introducing memory in the combiner in the form of feedback. It will be shown that feedback with unit delay (as considered by Rueppel) in general yields statistical dependence between *pairs* of bits at the input and output of the combiner. It will also be shown that if the delay in the feedback is n time units, then there is in general a statistical dependence between $(n+1)$-tuples of the input and output of the combiner. This fact can be used to design nonlinear combiners with feedback that will be sufficiently strong against a correlation attack.

Provability of Security and the Wire-Tap Channel

AARON WYNER, AT&T Bell Laboratories, Murray Hill, NJ 07974, USA

The wiretap channel is discussed with particular emphasis on the provability of its level of security. The following is a typical result: A k-bit message, S, is encoded into a binary $2k$-vector, X. An intruder is allowed to look at a k-bit subsequence, Z, of X of his choice. Then (when k is large) it is possible to make this encoding such that the equivocation $H(S|Z) \simeq k - 1$ bits $= H(S) - 1$. Since the encoding is done by generating k random bits, the system is secure to the extent that the random bit generator is truly random. (Joint work with L.H. Ozarow)

Algebraic Coding for the Wire-Tap Channel

VICTOR K. WEI, Bellcore, 435 South Street, Morristown, NJ 07960, USA

Wyner and Ozarov-Wyner studied wire-tap channels of types I and II. They derived the channel capacities and studied several coding schemes with emphasis on asymptotics. The coding problem for wire-tap channels will be treated here in a more algebraic manner. It will be shown in particular that the performance of linear codes over the type II channel can be described in terms of a generalized Hamming weight. Several bounds on code parameters are derived from this perspective, and the exact hierarchy of corresponding generalized Hamming weights for Hamming codes and for Reed-Muller codes of all orders is determined. The proofs utilize the Kruskal-Katona Theorem from finite-set combinatorics. The obtained results can also be applied to several other cryptographic problems including bit extraction or t-resilient functions, known-plaintext attacks in public-key encryption, and variations of secret-sharing schemes.

A New Information-Theoretic Notion of Cipher Security and a Provably-Secure Randomized Cipher

UELI M. MAURER, Institute for Signal and Information Processing, Swiss Federal Institute of Technology, CH 8092 Zurich, Switzerland

A Cipher is defined to be *perfect with probability at least p* if $I(X;Y|A) = 0$, where X and Y denote the plaintext and cryptogram, respectively, and where A is an event that occurs with probability $P(A) \geq p$. By a strongly-randomized cipher, we mean that a random string is publicly available whose length is much greater than that of the plaintext. A strongly-randomized cipher will be exhibited whose secret key Z is small (compared to the length of the plaintext), i.e., that satisifies $H(Z) \ll H(X)$, and that is provably perfect with probability very close to 1 when the enemy's computational power is limited in a reasonable way. This result is somewhat surprising because, according to Shannon, perfect secrecy, i.e., $I(X,Y) = 0$, implies that $H(Z) \geq H(X)$. Moreover, although information-theoretic security usually implies that the enemy has infinite computing power, the constructed cipher is secure for an information-theoretic notion of security only when the enemy is computationally restricted.

encoder output word divided by the length (in bits) of the encoder input word. As a measure of performance, we are interested in the exponential decay rate of the probability that the ECR exceeds a given threshold R in the range $H < R < 1$, where H is the entropy of the (binary) source. This is different from the commonly used performance measure of compression ratio, defined as the ratio between the *expected* output word length and the *expected* input word length, as it quantifies the rate of convergence of the ECR and provides insight on its tail behavior. It is shown that for *any* unifilar Markov source, the exponential decay rate of the probability that the ECR exceeds R, for the best V-F code, is $1/R$ times faster than that of the best F-V code with the same number of codewords, i.e., essentially the same storage requirements. Furthermore, the best performances in both F-V and V-F code classes are attained by universal codes which depend neither on the source nor on the value of R.

A Conjecture on Source Coding

RAYMOND W. YEUNG, AT&T Bell Laboratories, Holmdel, NJ 07733, USA

It is a folk belief that we should not consider probabilistic encoding and decoding schemes when designing communication networks. It is also a folk belief that for communication networks in which the channels are point-to-point, discrete, memoryless and independent of each other, optimization can be achieved asymptotically by designing source coding and channel coding separately. Shannon showed that the latter is true for point-to-point communication systems. In this paper we make a conjecture which would imply that the above folk beliefs are true for a rather general class of networks. We show that our conjecture holds for two non-trivial network configurations, including the classical multiterminal configuration and what we call the sequential data compression configuration. The latter is the configuration of a new multiterminal source coding problem of which we determine the admissible coding rate region. We also prove a necessary condition for an optimal coding scheme for the so-called two distortion criteria problem, which has remained unsolved for a long time.

A Universal Data Compression Scheme With Distortion

BIXIO RIMOLDI AND HIROSUKE YAMAMOTO, Dept. of Electrical Engineering, Washington University, One Brookings Drive, St. Louis, MO 63130, USA, and Dept. of Communications and Systems, University of Electro-Communications, 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182 Japan

A universal data compression scheme with distortion for finite alphabet sources is proposed and analyzed. The proposed scheme is an extension of the *move-to-front* scheme proposed by Bently et al. which is a universal noiseless encoding scheme. The encoding algorithm is as follows. Let $\{x_i\}_{i=1}^{NL}$ be the source output. The encoder chops it into source words $x_t^L = x_{(t-1)L+1}x_{(t-1)L+2} \ldots x_{tL}$ of length L, $t = 1, 2, \ldots, N$. The encoder has a buffer which is composed of M registers. Each register has length L and it can store one source word. First we clear the buffer. When $x_t^L (t \geq 1)$ is encoded, we search for the minimum j such that $d(x_t^L, W_j) \leq D$ where $d()$, D, W_j are a distortion measure function, distortion tolerance, the content of the j-th register, respectively. If we can find it in the buffer, then we send $B(j)$ as the codeword where $B()$ is a universal code of the positive integers. Next we move W_j into register 1 and shift the words of registers $1 \sim (j-1)$ to register $2 \sim j$. Otherwise we send $B(j+1)b(x_t^L)$ where m is the buffer size used by t and $b(x_t^L)$ is a binary representation of x_t^L. After we shift the words of registers $1 \sim m$ to register $2 \sim (m+1)$, we store x_t^L into register 1 as a new codewords.

For the foregoing scheme, the following theorem holds.

Theorem If the source is i.i.d., which is characterized by a probability distribution $P_X(x)$, then the average compression rate $\rho(X^{NL})$ is bounded by

$$\rho(X^{NL}) \leq \min_{P_{\hat{X}|X}(\hat{x}|x):Ed(X,\hat{X})\leq D} D(P_{\hat{X}|X}||P_{\hat{X}}|P_{\hat{X}}) + \varepsilon,$$

where $P_{\hat{X}}(x) = P_X(x)$ for all $x \in \chi$ and

$$D(P_{\hat{X}|X}||P_{\hat{X}}|P_X) = \sum_x P_X(x) D(P_{\hat{X}|X}(.|x)||P_{\hat{X}})$$

is the conditional informational divergence.

It can be easily shown from the theorem that if $P_X(x) = 1/|\chi|$ for all $x \in \chi$ and distortion measure function $d(x, \hat{x})$ is symmetric, the proposed scheme can achieve rates arbitrarily close to the rate-distortion function.

Identifiablity of Hidden Information Markovian Sources and Their Minimum Degrees of Freedom

H. ITO, S. AMARI, AND K. KOBAYASHI University of Tokyo, Japan, and University of Electro-Communications, Japan

Let us consider a finite state Markov chain and a function on its states to an other arbitrary set. Suppose that the function is not one-to-one and that we cannot see the states directly, then the function process of the original chain is no longer Markovian in general. We call this type of sources hidden Markovian information sources, which are very important and natural in both theoretical and practical sense. However, the most basic problem have been remained open for this type of sources, that is, when two sources are identical. In the paper, we complete this identifiability problem by means of a new linear algebraic consideration, in which we take the mathematical structure induced by given Markovian transition matrices carefully into account. Moreover, we reveal the minimum degrees of freedom for a given source. There exists some kind of special linear subspaces which respectively represent transient parts or ergodic parts like ordinary finite state Markov chain, and also exists a special one called zero subspace that is entirely different from that. We concretely construct the subspace corresponding to the minimum degree and prove that there exists an isomorphism on that subspace if and only if two sources are equivalent.

Authors Index