IEEE Information Theory Society Newsletter

Information Theory 50 1948-1998

Special Golden Jubilee Issue
Summer 1998

Editor: Michelle Effros
Guest Co-Editors: Anthony Ephremides & James Massey

ISSN 1059-2362

# 1948-1998 Information Theory: The First Fifty Years

*Anthony Ephremides and James Massey*
*Guest Co-Editors*

Information Theory is one of those rare scientific fields to which one can assign a definite beginning. The publication in 1948 of Claude E. Shannon's celebrated paper, "A Mathematical theory of Communication", marks the birth of Information Theory as clearly as the Declaration of Independence in 1776 marked the birth of a country. This special issue of the Newsletter is part of the IEEE Information Theory Society's celebration of the half-century mark in the history of our field. We are celebrating not only the birth of Information Theory, but also its remarkably rapid development and widespread application.

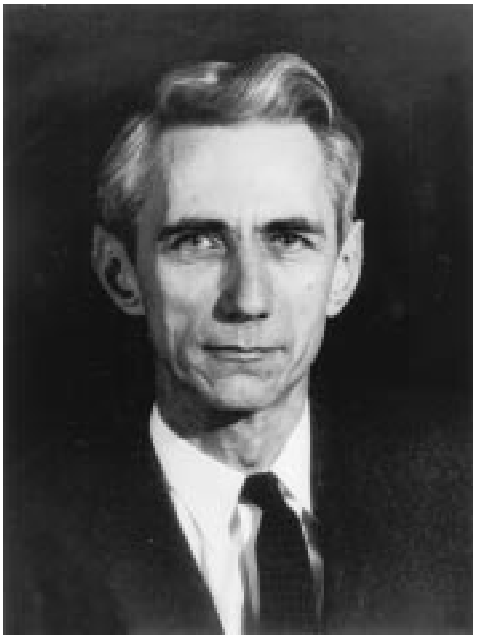


Photo by L. Zadeh

**Claude Elwood Shannon**
**The founder of Information Theory**

In this special issue we have attempted to document some aspects of this success story. The features we selected attest either to the eminent and profound impact of the field or to fond remembrances and personal commentaries.

In particular, we solicited and included here brief accounts of the impact of Information Theory on sister fields of science and engineering. We also solicited personal comments about the allure of Information Theory (and of Shannon's pioneering work in particular) by the individuals whom we have honored with the highest distinction bestowed by the Information Theory Society, namely the Shannon Award (whose recipients were reviously known under the rubric of "Shannon Lecturers"). We also thought it would be important to look to the future. After all, this year also marks the beginning of the next fifty years of Information Theory. Thus, we solicited personal comments on the outlook for our field from a few of the many talented researchers of the younger generation. These younger workers generally have an appreciation of Shannon's impact only through his published work and not through his equally influential persona.

This being a "newsletter", we also have included reports on the many special events and activities that mark the observation of this anniversary year.

Finally, we turned to the archives and picked a sampling of memorabilia and vignettes from the legacy of Claude Shannon, "one of the greatest scientific minds

Continued on next page

## 1948-1998 Information Theory: The First Fifty Years

Continued from page 1

of our time"—to borrow a felicitous description from a recent introduction of Shannon.

We thank all the contributors who responded generously to our call for help.

We hope that this special issue will be a fitting footnote to the tribute that all information theorists pay to our field and to its founder on this 50th year of its already colorful history. Because it was announced at the request of Mrs. Betty Shannon by the President of the Information Theory Society at the 1995 International Symposium on Information Theory, we know that her husband, Claude Shannon, is in good physical health but is suffering from Alzheimer's disease, that dreadful curse that respects neither Geniuses nor Statesmen. We hope nonetheless that that this special issue of the Newsletter will help to make him aware that he enjoys in abundance the admiration and affection of information theorists around the world.

IEEE
Information Theory
Society Newsletter

*IEEE Information Theory Society Newsletter* (USPS 360-350) is published quarterly by the Information Theory Society of the Institute of Electrical and Electronics Engineers, Inc.

Headquarters: 345 East 47th Street, New York, NY 10017.

Cost is $1.00 per member per year (included in Society fee) for each member of the Information Theory Society. Printed in the U.S.A. Periodicals postage paid at New York, NY and at additional mailing offices.

**Postmaster:** Send address changes to IEEE Information Theory Society Newsletter, IEEE, 445 Hoes Lane, Piscataway, NJ 08854.

## From the Editor

*Michelle Effros*

**Michelle Effros**

Planning for the celebration of the 50th anniversary of information theory has been underway for several years now. After consideration of a number of different locations, the site for ISIT '98 was announced by Dave Forney at the June 1996 meeting of the Board of Governors. An ad hoc committee for the 50th anniversary celebration was announced at the March 1997 meeting, and this committee, with Ezio Biglieri at its head, has been working steadily since. This special issue of the newsletter was approved at a July 1997 Board of Governors meeting. The thought and effort that have since gone into putting this issue together are considerable. All of the credit for the issue's success belongs to the guest editors of this special edition. It is therefore with great pleasure and enormous thanks that I welcome those editors, Tony Ephremides and Jim Massey, for this special issue of the IEEE Information Theory Society Newsletter.

## Table of Contents

# IEEE Information Theory Society Golden Jubilee Awards for Technological Innovation

## Presented on August 17, 1998 at the 1998 ISIT

Nominations for these awards were solicited through the Newsletter, e-mail, and through personal contacts. No limit was put on the number of awards. To receive an award the nominee had to have strong support from the committee. Comments from outside the committee were solicited from experts (as needed) and given substantial weight.

The goal was to recognize contributions that have been instrumental in important products and applications. The motivation for the awards was (is) that such contributions from members of our society may have been overlooked in the past and may not have received adequate recognition—although past recognition did not preclude a nominee from receiving an award. The committee that made the final selections consisted of

Thomas Ericson, president
Ezio Biglieri, 1st VP
Vijay Bhargava, 2nd VP
Sergio Verdu, 1st Past President
Jerry Gibson, 2nd Past President
(Chair of the committee)

The selected awardees are listed below in alphabetical order.

**1. Norman Abramson**
For the invention of the first random-access communication protocol.

**2. Elwyn Berlekamp**
For the invention of a computationally efficient algebraic decoding algorithm.

**3. Claude Berrou, Alain Glavieux, and Punva Thitimajshima**
For the invention of turbo codes.

**4. Ingrid Daubechies**
For the invention of wavelet-based methods for signal processing.

**5. Whitfield Diffie and Martin Hellman**
For the invention of public-key cryptography.

**6. Peter Elias**
For the invention of convolutional codes.

**7. G. David Forney, Jr.**
For the invention of concatenated codes and a generalized minimum-distance decoding algorithm

**8. Robert M. Gray**
For the invention and development of training mode vector quantization.

**9. David Huffman**
For the invention of the Huffman minimum-length lossless data-compression code.

**10. Kees A. Schouhamer Immink**
For the invention of constrained codes for commercial recording systems.

**11. Abraham Lempel and Jacob Ziv**
For the invention of the Lempel-Ziv universal data compression algorithm.

**12. Robert W. Lucky**
For the invention of pioneering adaptive equalization methods.

**13. Dwight O. North**
For the invention of the matched filter.

**14.Irving S. Reed**
For the co-invention of the Reed-Solomon error correction codes.

**15.Jorma Rissanen**
For the invention of arithmetic coding.

**16.Gottfried Ungerboeck**
For the invention of trellis coded modulation.

**17.Andrew J. Viterbi**
For the invention of the Viterbi algorithm.

# IEEE Information Theory Society Golden Jubilee Paper Awards

## Presented on August 17, 1998, at the 1998 ISIT

The IEEE Golden Jubilee Paper Awards Committee was formed to recognize outstanding articles published in the IEEE Transactions on Information Theory whose impact on the development of the fields of interest to the Information Theory Society is widely recognized. A particular focus of the committee was to select papers that have become classics in the field, but which were not fully appreciated at the time of publication. Thus, previous winners of the IT Society Paper Award were not eligible. The committee membership was comprised of:

Vijay Bhargava
Dan Costello, Chair
Imre Csiszar
Joachim Hagenauer
Dave Forney
Tom Kailath

Vincent Poor
Shlomo Shamai
Alexander Vardy
Victor Wei
Frans Willems
Jacob Ziv

The selected papers, along with a brief description of their contribution, are the following.

**1. M.J.E. Golay, "Notes on digital coding," Proc. IRE, vol. 37, p. 637, 1949.**

This paper, one year after Shannon's, introduced the single most significant code in algebraic coding theory. As Berlekamp said in the introduction to "Key Papers in Coding Theory," it is surely the best half-page contribution to coding theory ever, and is universally recognized as a classic.

**2. C.E. Shannon, "The zero-error capacity of a noisy channel", IEEE Trans. Information Theory, IT-2, pp. 8-19, 1956.**

This paper develops the idea of zero-error capacity, which has kept many a combinatorialist busy, and develops several tools for its analysis. As a bonus, Shannon proves that noiseless feedback does not increase the capacity of the discrete memoryless channel (and that it may indeed increase the zero-error capacity).

**3. R. Price, "A Useful Theorem for nonlinear devices having Gaussian inputs", IEEE Trans. Inform. Theory, IT-4, pp. 69-72, June 1958.**

Price found an interesting and unique property of Gaussian processes that allows for a simple calculation of, for example, the correlation function at the output of quite general nonlinearities. Price's formula encompasses a large number of previously derived special cases, e.g., including the hard and soft limiters, linear detectors, etc.

**4. R.G. Gallager, "Low-Density Parity-Check Codes," IRE Transactions on Information Theory, IT-8, pp. 21-28, January 1962**

In this seminal work, which is based on his PhD thesis, (see also the extended monograph: R.G. Gallager, "Low-Density Parity-Check Codes," MIT Press 1963) not only does Gallager come up with a coding scheme which exhibits extraordinary performance commensurate with what is today achievable with turbo-codes, but he specifies the iterative decoding with a posteriori probabilities, and he clearly discriminates between extrinsic and intrinsic information. In fact the theory Gallager was able to provide is in its depth, in some aspects, beyond what is known today for turbo-codes (including many hundreds of contributions which followed the classic ICC '93 paper by Berrou, Glavieux, and Thitimajshima, who introduced the modern concept of iteratively decoded turbo codes).

**5. R.G. Gallager, "A simple derivation of the coding theorem and some applications", IEEE Trans. Information Theory, IT-11, pp. 3-18, Jan. 1965.**

The third major approach to the proof of the achievability part of the coding theorem (along with Shannon's and Feinstein's) is developed in this paper. Further developed in Gallager's textbook with emphasis on error exponents, this is the way generations of information theorists have learned how to prove the fundamental theorem of information theory.

**6. T. Cover and P. Hart, "Nearest neighbor pattern classification", IEEE Trans. Inform. Theory, IT-13, pp. 21-27, 1967.**

The title describes the simple idea, which is of course not new. However, the issue is to analyze the properties of such a scheme. Cover and Hart showed that the nearest-neighbor rule has a risk that is less than twice the Bayes risk for all reasonable distributions and for any number of categories.

**7. J.L. Massey, "Shift-register synthesis and BCH decoding", IEEE Trans. Inform. Theory, IT-15, pp. 122-127, 1969.**

This is the "Massey part" of the Berlekamp-Massey algorithm, which is now taught in every course on coding theory pretty much everywhere. The paper develops a lucid, highly ingenious, theory of this algorithm in the framework of shift-register generation. It thus extends the applicability of this algorithm well beyond decoding BCH codes, in fact, beyond our field of information theory. Owing to Massey's insights, the algorithm is now used in cryptography and in control of linear systems.

**8. T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise", IEEE Trans. Inform. Theory, IT-15, pp. 350-361, 1969.**

This is the paper that established the famous "estimator-correlator" formula for the likelihood ratio in the problem of detecting stochastic signals in white Gaussian noise. The Kailath estimator-correlator representation is quite fundamental. In fact, after the matched filter (which it actually subsumes), it is the most fundamental result in signal detection theory, with implications in many other areas of stochastic analysis. It compactly sums up essentially the entire problem of signal detection in white Gaussian noise, and moreover it justifies many practical detection systems that operate as estimator-correlators (such as the RAKE receiver).

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

This paper (and its subsequent parts) essentially develops from scratch a profound and rigorous theory of convolutional codes as algebraic entities in the field of Laurent series. The subject matter of this paper is classic textbook material today. It was not fully appreciated at the time, because it didn't say much about free distance and the coding community was focused almost exclusively on rate 1 / n codes. But it has enjoyed a renaissance recently with the interest in systematic feedback encoders for TCM, group codes, and turbo codes.

**10. G.D. Forney, Jr, "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", IEEE Trans. Information Theory, IT-18, pp. 363-378, May 1972.**

This paper solves the open problem of optimum equalization of finite-length intersymbol interference channels. It also develops a method to upper bound the minimum bit error rate. It pioneers the concept of "whitened matched filter." The influence of this paper in the development of the theory and practice of digital communications is hard to overestimate.

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

The BCJR paper deserves full credit for the origin of the trellis theory of block codes. What's more, their elegant construction produces not just some trellis for a block code, but the minimal trellis!

**12. R.J. McEliece, E.R. Rodemich, H.C. Rumsey, and L.R. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," IEEE Trans. Inform. Theory, IT-23, pp. 157-166, 1977.**

This paper presents the best known upper bound on the number of codewords in a code of length n and distance d. The problem is described in [MacWilliams and Sloane, p.523] as "probably the most basic problem in coding theory." The significance of the result is well punctuated by the fact that no improvements to this bound are in sight today, 20 and some years later.

**13. H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," IEEE Trans. Inform. Theory, IT-23, pp. 371-377, 1977.**

The first paper to introduce multilevel coding ideas in the context of error-correction. The most important result of this paper is the proof that multistage decoding is a bounded-distance decoding algorithm. The paper also introduces constellation partitioning ideas akin to Ungerboeck's.

**14. R.M. Tanner, "A recursive approach to low-complexity codes," IEEE Trans. Inform. Theory, IT-27, pp. 533-547, 1981.**

This is a visionary paper that was truly ahead of its time. The paper establishes, for the first time, the profound connections between codes, graphs, and iterative decoding. There are several striking ideas in this paper regarding the importance and usefulness of extremal graphs, that remain overlooked even today.

**15. S. Verdu, "Minimum probability of error for asynchronous Gaussian multiple-access channels," IEEE Trans. Inform. Theory, IT-32, pp. 85-96, 1986.**

This is the paper that first described the field of multiuser detection. It went against the conventional thinking for receiver design in multiuser channels, by treating the multiple-access interference as a structured interference, rather than as an addition to the ambient noise level. The paper appeared at a very timely juncture in the field of multiuser communications, just as commercial CDMA systems were beginning to come into their own. Its influence has been quite significant. This is a rare paper that has had profound influence on the thinking of both researchers and practitioners in the field.

## The Historian's Column

*A. Ephremides*

This issue is in its entirety about history. It makes the presence of this column superfluous. Fifty years of vibrant presence of Information Theory in the world of science and engineering have left a clear, indelible mark. What more is there to say?

Perhaps, the right thing will be to look into the future. Gazing into the crystal ball is of course the antithesis of History. And yet, the reason we study History is to learn and understand in order to predict. Just as we measure the statistics of a channel in order to design the communication system.

A. Ephremides

Let us not confuse musing about the future with actually trying to predict it. I know better than venturing into prophesies and guesses. But I feel comfortable expressing some philosophical thoughts about the future. In a time of overwhelming (almost overbearing) domination by the past as we observe the 50th anniversary mark, the only escape by a historian will be to step aside and engage in some—anti-history. It's only fitting in moments like this not to be blinded and incapacitated by the greatness of the past. We need to keep moving. I remember when I was promoted to Full Professor and was basking in my self-congratulatory mood, a question by one of the staff members in our department caused me to awaken and experience a rough landing. "And now what?", he asked. And he got me thinking.

Indeed, Shannon's legacy is difficult to summarize. And the eventual successful application of his ideas makes Information Theory a textbook example of the sequence "theory-development-application-new theory-more development-additional applications-etc." But can we continue feeding from the same source? Is the richness inexhaustible? There is still much left that can be exploited, to be sure. but, perhaps, it is time to realize that we need new injections of energy and inspiration. The problems that we are confronting today in the broad area of communication systems have been compounded by a new beast that Shannon did not get a chance to tackle fully: Networks and multiuser systems. He did grasp the basis for it, since he did speak of feedback and secrecy. But, his thought and work did not anticipate the full impact of its growth. It was like passing from one dimension to multiple dimensions. The extension is not trivial. We are indeed today in desperate need of “A mathematical theory of networking”. Will it happen?

Instead of venturing an answer to this question, I would like to just make a few observations. First, the biggest conceptual advances in the field of Networking have come from Information Theorists. They still fall far short of the goal, but they are the only ones that have shed some light into the chaos. Yet, a lot of development out there by talented engineers of different backgrounds and persuasions has led to network systems, products, and services that have changed the way we live. But, it has been mostly chaotic and rudderless. It has been the occasional incisive thought by an Information Theorist that has injected a little order and sanity into this process. Secondly, the emergence of multi-user theories (for channel access for broadcasting, and for detection) have provided a conduit for the passage of Information-theoretic wisdom into the field of networking. We are still in need of the “missing link”, but there is light in the tunnel.

So, if I am to draw a conclusion about the future I can only express my own opinion but I have a hunch that many of you would share my view. I feel confident to conclude that, at least, “the dream is alive”!

# On the Impact of Fifty Years of Information Theory

*Information Theory has had rich interactions with many fields of science and application. To document these interactions, we solicited the next seven short notes from leading experts in both Information Theory and sister fields. They document the profound impact of our field across a wide spectrum of science and engineering.*

*The Eds.*

## The Impact of Information Theory on Communications

*Joachim Hagenauer*
*Munich University of Technology (TUM), Germany*

### Introduction

Talking about the impact of information theory on communications may seem at the first glance a little strange, because information theory is an offspring of communications theory. In fact, in Shannon’s groundbreaking paper entitled “A mathematical theory of communications(!)” which we celebrate as the starting point of information theory, the word “information theory” does not even appear. Nevertheless Shannon’s information theory is much more, it is “A philosophy of information from the point of view of communications”, as Bob Lucky put it. It grew out of communications, but it changed the view of communications engineers dramatically. Take for instance the impact of the channel coding theorem: Before Shannon, engineers could only imagine that a message could be reliably transmitted over a noisy channel by assigning unlimited power or unlimited bandwidth. Now they learned that they had to assign only a limited amount of redundancy via coding to achieve the same goal. Still, many of the engineers looked with some suspicion at the results of information theory, because its proofs are usually non-constructive and do not directly help in the actual design of a communications system. And, at least during the first 35 years, the absolute limits shown by information theory were still far away from what could be achieved in practice.

We will indicate in this contribution quite a number of examples where the promise of information theory did pay off in practice, in which information theory guided engineers-or in retrospect should have guided them-to achieve better communications.

### Deep Space and Satellite Communications

Deep Space Communications was the first area where engineers were led by information theory. One reason is that this channel can be accurately modeled as an additive white Gaussian noise channel with power density $N_0/2$. Information theory had results ready for this channel: Transmitting at the capacity limit means $R = C(RE_b/N_0)$ where $C(E/N_0)$ is the capacity of the Gaussian channel with average symbol energy $E$ and thus the minimal required energy per transmitted bit is $E_b$. Solving this equation for $E_b/N_0$ gives the so-called Shannon limit, which is shown in Table 1:

| Input | Output | Rate | $E_b/N_0\,\vert\,\text{min}$ |
|---|---|---|---|
| binary | soft | $\rightarrow 0$ | -1.6 dB |
| binary | soft | 0.5 | 0.2 dB |
| binary | binary | 0.5 | 1.8 dB |

Table 1: Shannon limit for the AWGN channel

At a BER of $10^{-5}$, uncoded binary transmission needs 9.6 dB, thus some 9 to 10 dB of power would be saved if Shannon’s promise could be made to come true. So the deep space channel and coding, guided by information theory, matched perfectly: A “marriage made in heaven”, as Jim Massey called it. The first pioneer spacecraft with sequential decoding achieved more than 3 dB coding gain, a first remarkable step. Later a concatenated scheme, namely an inner memory 6 convolutional code combined with an outer Reed-Solomon code of length *255* bytes with 8 bits each became the standard for deep space missions both of NASA and ESA. Again, information theory arguments had shown that concatenated schemes should perform well and practical considerations guided the match between the Viterbi and the RS decoder: The Viterbi decoder could easily accept soft values from the AWGN channel (a 1.6 dB advantage as indicated by the Table 1!) and its output bursts were easily handled by the RS-decoder. The scheme operates at 2.7 dB and therefore Shannon’s limit was only *2.5* dB away. Motivated by the failure of Galileos’s main antenna, researchers tried to squeeze out another few tenths of a dB by using elaborate iterative methods between inner and outer decoders, employing soft-output Viterbi algorithms and state pinning methods. With these receiver-only modifications almost another decibel was gained and Shannon’s limit became only 1.6 dB away, Thus, judging from information theory, deep space communications approached its limit and the engineers cannot hope for many more decibels of savings.

Motivated by this success, the same coding scheme also became a standard for digital TV transmitting MPEG2 compressed video via broadcast channels. Only the RS code was modified to a shorter blocklength.

For deep space, binary PSK was the best choice for the modulation format because bandwidth is not very limited. For satellites, the bandwidth requirements are more stringent. Therefore, 4-PSK was used, allowing uncoded 2 bits / symbol and less for coded systems because the redundant bits were spread out in time. Then came the invention of bandwidth efficient coded modulation by Imai and Ungerboeck. It is remarkable that Ungerboeck’s paper started with information theory arguments, specifically showing that operating at channel capacity with 2 bits / symbol saves 6.6 dB relative to uncoded 4-PSK if the redundancy is used to expand the symbol set to 8-PSK. This meant coding without bandwidth expansion. The first coded 8-PSK modem was built in a joint project between IBM’s G. Ungerboeck and the German Aerospace Research DLR and tested in 1983 via the INTELSAT satellite. The old modem was replaced by the new code and the satellite power could be reduced by 4 dB without any change in performance, bandwidth and data rate. This was so surprising that one of the ground station engineers suggested that some sort of foul play was involved, because previously he had learned that coding always expands bandwidth.

## Mobile Communications

During the last 15 years, mobile communications became the most extensively worked-on field in communications. Did information theory help in designing all these successful cellular systems? At least the engineers could have learned a lot from the results given in Table 2, and some probably did. Table 2 shows Shannon’s limit for a narrowband mobile channel in cities, where Rayleigh fading prevails. Channel state information (CSI) means that , the receiver measures the received power level, which has to be done anyway for power control in mobile systems.

| Input | Output | Rate | $E_b/\mathrm{N}_0 \mid$ min |
|---|---|---|---|
| binary | soft + CSI | $\rightarrow 0$ | -1.6 dB |
| binary | soft +CSI | 0.5 | 1.8 dB |
| binary | soft | 0.5 | 2.5 dB |
| binary | binary | 0.5 | 4.9 dB |

Table 2: Shannon limit for the Rayleigh channel

The comparison with the AWGN limits in Table 1 shows that for very low rate codes the fading effect can be totally removed if sufficient diversity is applied. That is exactly what cellular designers do, namely to introduce antenna diversity, coding diversity and multipath diversity. The second lesson to learn, which is as well applied in a mobile system, is to use soft decisions and channel state information instead of hard decisions, otherwise an asymptotic loss of 3.1 dB would occur.

## Error Control Coding in Communications Systems

Immediately after Shannon had shown that the capacity bound could be achieved with long random codes, the search for good codes decodable with limited complexity began. Hamming codes where the starting point of the widespread usage of codes over finite fields such as cyclic, BCH and Reed-Solomon codes. An elaborate theory showing the algebraic and asymptotic performance properties of these codes impressed many engineers; some, however, shied away from what they perceived as a rather esoteric field. Today long codes like Reed-Solomon codes over GF(256) are standard building blocks in communications systems. However, the block codes over finite fields did not come close to the capacity bounds in practical applications. Much more success in practice was enjoyed by probabilistic decoding of rather simplistic codes using Viterbi and sequential decoding. For the latter decoding method long random codes picked from any telephone book would do the job.

The real breakthrough for the attempts to approach the capacity bounds occurred 5 years ago with the invention of turbo decoding. The inventors and early contributors (Berrou, Glavieux, Battail, Lodge) were mostly not hard-core information theorists but electronic and communication engineers. However, they had understood the basic ideas of Shannon: First: Use long random codes-which they did by concatenating simple deterministic codes through random interleavers. Secondly: Use probabilistic decoding and all the information which is available to the receiver-they did it through iterative decoding while passing all the necessary information in the form of a *a posteriori* probabilities or log-likelihood values to the next stage of decoding. The asymptotic minimum distance code property became obsolete; at a BER of $10^{-5}$ the turbo decoders performed within 0.7 dB of Shannon's limit for rate 1/2. For higher rates around 0.9, simple Hamming or low density parity check codes were used as constituent codes and the respective Shannon limit was approached as close as 0.27 dB. In all cases simple codes with moderate complexity decoders were sufficient, but at the same time large interleavers as required by information theory were used. The promise of information theory is fulfilled after 50 years in the area of error control in communications systems.

## Multiuser and Network Communications

All the information theory and communications engineering examples discussed so far concern point-to-point communications links. The real communications world, however, has to accommodate many users, suffers from multiuser-interference and could benefit from cooperative coding and feedback. Some of the results in network information theory date back to Shannon, many of the basic results were created in the seventies by Cover, Schalkwijk, v.d Meulen, Ahlswede, Slepian, Wolf, Wyner et al. Only recently results in information theory for multiuser (CDMA) channels have become available i.e. by Verdu, Shamai and Rimoldi. Up to now, communications engineers have been slow to take advantage of these old and new findings. If we take the example of a cellular network, it was for a long time common practice to treat the interference from other users as noise, invoking the central limit theorem. It can be easily shown by information theory arguments that a lot is lost following this simplistic approach. Only during the last 5 years, more elaborate detection methods inspired by information theory were applied, such as joint detection and decoding, successive interference cancellation, rate splitting with soft cancellation and others. Also joint coding over multiple antennas, called space-time codes in SDMA systems, was applied recently, spurred by information theoretic thinking.

## Communications on Subscriber Loops with Modems and XDSL transmission

Let us turn to briefly to the influence of information theory on the devices which bring two-way communication links to our home. The dramatic step from 4.8 or 9.6 kbit/s to more than 30 kbit/s transmission with modems on the 3 kHz telephone channel was made possible to great extent by coded modulation, with its starting point in Ungerboek's information-theoretic arguments as mentioned above. These arguments show further that it does not pay to go beyond doubling the signal space through coding, a feature still observed in most modems. Further information theory arguments concern the limit on the so-called shaping gain of the signal constellation.

A hot area is currently the XDSL transmission over the existing twisted copper pairs to our home with rates up to 25 Mbit/s. Some schemes employ parallel channels via multitone multiplexing and borrow the well-known waterfilling argument from information theory in order to load the subchannels in an optimal way.

## Multimedia Communications

Multimedia is the new word for using source coding and transmission of voice, audio, image video and TV signals in an integrated fashion. Of course, information theory contributed a lot to this area of communications. Almost every standardized source compression scheme contains quantizers designed using rate-distortion principles. This is followed by some form of entropy coding, such as variable length coding, arithmetic coding , Huffman coding and all variations of it. But if one looks at the engineers' design of whole multimedia systems such as MPEG2 video and audio, there seems to be no general masterplan using information theory guidelines. It is a conglomerate of transform methods, signal processing, compression and multiplexing techniques. A similar situation prevails in speech coding; there is no top-down design following general principles, say from rate-distortion theory.

Between source and channel coding designers there has been a clearcut interface dating back to Shannon's famous separation theorem, which states that in the information-theoretic sense (unlimited delay and complexity) nothing can be gained by optimizing both tasks jointly. In standardization groups source and channel coding people used to meet separately. Only recently it was discovered that joint source and channel coding is beneficial when imperfect compression and rather bad channels are involved. Ideas like unequal error protection using source significance information, source controlled channel decoding using *a priori* and *a posteriori* knowledge about the source symbols, as well as elaborate concealing techniques, became more popular and resulted in a superior overall performance. Shannon himself had already suggested using residual source redundancy for better error correction and detection.

## Conclusions

We have listed a sizable number of communication systems where information theory played a crucial role, and there are many more. For instance, we have omitted here the whole area of cryptography. Nevertheless, it should be mentioned that there are wide areas in communications where information theory has had little impact on today's system design, such as optical communications, where the driving force was technology, or the entire Internet, where protocol and access design do not depend much on information theory. On the whole, however, there is no doubt that the discipline "Information Theory" has had a dramatic influence on the design of practical communications systems. It has always paid off for communications engineers to have a sound education in information theory.

# Applications of Information Theory in Probability and Statistics

*Imre Csiszár*
*Budapest*

Statistics, the science of extracting information from data, is a most natural field of applications of information theory (IT) ideas. Typical IT tools that have profound applications in probability and statistics are information measures, the method of types, and the concept of coding. Here we briefly review some of these applications. A more detailed but still incomplete review has been the author's Shannon Lecture at ISIT97, cf. the March 1998 issue of this Newsletter; the relevant references may also be found there. Information measures other than those basic for IT are outside the scope of this review, including Fisher's information that has been used in statistics since 1925.

## Hypothesis testing, large deviations

In the fifties, Kullback developed a unified approach to testing statistical hypotheses based on the information measure now known as information divergence (I-divergence) or relative entropy, denoted below by D(P | | Q). Its operational meaning for hypothesis testing was established in Chernoff's 1952 paper, attributing the result to Stein. Though I-divergence was formally introduced in 1951, it implicitly played a substantial role already in Wald's (1947) seminal work on sequential analysis. Hájek's dichotomy theorem was also a celebrated result at the time (1958), namely that Gaussian measures P and Q are either orthogonal or mutually absolutely continuous, according as D(P | | Q)+D(Q | | P) is infinite or finite.

The significance of I-divergence for probability and statistics was further enhanced by Sanov's work in 1957 on large deviation probabilities for the empirical distribution $\hat{P}_n$ of an i.i.d. sample X1,…, Xn from a distribution Q. Sanov's theorem in the discrete case, and Hoeffding's later results on hypothesis testing error exponents, may be viewed today as easy applications of the method of types. Historically, these results preceeded, and provided an impetus for, the development of that method in IT.

A general version of Sanov's theorem says, in modern terminology, that $\{\hat{P}_n\}$ satisfies the large deviation principle with good rate function D(· | | Q), if probability measures are equipped with the $\Upsilon$-topology. The simplest available proof of this (essentially due to Groeneboom, Oosterhoff and Ruymgaart) is inherently information-theoretic. The I-projection P* of Q onto a convex set Π of probability measures is defined as the minimizer of D(P | | Q) subject to P ∈ Π. Gibbs' conditioning principle asserts the convergence of the conditional distribution of X1 on the condition $\hat{P}_n \in \Pi$ to P*, as n → ∞. A rather simple calculation using properties of I-divergence suffices to prove a strong version of this principle (Csiszár 1984).

## Inference principles motivated by IT

The maximum entropy principle, pioneered by Jaynes and Kullback, says in its simplest form the following: If an unknown distribution P ought to be inferred when knowing only that P belongs to a set Π of probability measures determined by linear constraints, one should infer that P equals the I-projection P* onto Π of a default model distribution Q (if Q is uniform then P* has maximum entropy, hence the name).

Algorithms designed for computing P* include iterative scaling and generalized iterative scaling (or SMART); these have intuitive IT interpretations that suggest how to prove their convergence to P*. These algorithms may also be used to compute maximum likelihood (ML) estimates. For computing ML estimates from incomplete data, statisticians often use the EM algorithm; this has an IT interpretation as alternating I-divergence minimization.

Rissanen's minimum description length principle (MDL) says that for stochastic modeling of given data when a list of candidate models or model classes is available, the one leading to shortest description of the data should be chosen, assuming a code "ideal" for the chosen model (or class), and taking into account that also the latter has to be described. MDL gives statistical significance to the theory of source coding, and is a rich source of applications of IT to statistics.

## Other topics

Starting from the late fifties, several authors have used IT methods (basically the properties of I-divergence or of related functionals of probability measures) to prove various limit theorems in probability theory.

In statistics, the amount of information in the sample about the unknown parameter has been studied, starting with Rényi (1969) and Pinsker (1972), and the results were used to derive bounds on how well that parameter can be estimated. Recently, tight risk bounds on non-parametric density estimation were derived in a similar manner (Yu 1995, Yang and Barron 1997).

Hypothesis testing or estimation based on remote data may involve the feature that encoding the data before transmission is permissible, subject to rate constraints. Such problems jointly belong to IT and statistics. Their study began in the mid eighties, and offers many challenges for future research.

Fascinating recent applications of IT to probability theory involve the currently hot topic called measure concentration. An early measure concentration result known as the blow-

# Applications of Information Theory in Probability and Statistics

*Imre Csiszár*
*Budapest*

Statistics, the science of extracting information from data, is a most natural field of applications of information theory (IT) ideas. Typical IT tools that have profound applications in probability and statistics are information measures, the method of types, and the concept of coding. Here we briefly review some of these applications. A more detailed but still incomplete review has been the author's Shannon Lecture at ISIT97, cf. the March 1998 issue of this Newsletter; the relevant references may also be found there. Information measures other than those basic for IT are outside the scope of this review, including Fisher's information that has been used in statistics since 1925.

## Hypothesis testing, large deviations

In the fifties, Kullback developed a unified approach to testing statistical hypotheses based on the information measure now known as information divergence (I-divergence) or relative entropy, denoted below by $D(P || Q)$. Its operational meaning for hypothesis testing was established in Chernoff's 1952 paper, attributing the result to Stein. Though I-divergence was formally introduced in 1951, it implicitly played a substantial role already in Wald's (1947) seminal work on sequential analysis. Hájek's dichotomy theorem was also a celebrated result at the time (1958), namely that Gaussian measures P and Q are either orthogonal or mutually absolutely continuous, according as $D(P || Q)+D(Q || P)$ is infinite or finite.

The significance of I-divergence for probability and statistics was further enhanced by Sanov's work in 1957 on large deviation probabilities for the empirical distribution $\hat{P}_n$ of an i.i.d. sample X1,…, Xn from a distribution Q. Sanov's theorem in the discrete case, and Hoeffding's later results on hypothesis testing error exponents, may be viewed today as easy applications of the method of types. Historically, these results preceeded, and provided an impetus for, the development of that method in IT.

A general version of Sanov's theorem says, in modern terminology, that $\{\hat{P}_n\}$ satisfies the large deviation principle with good rate function $D(\cdot || Q)$, if probability measures are equipped with the ϒ-topology. The simplest available proof of this (essentially due to Groeneboom, Oosterhoff and Ruymgaart) is inherently information-theoretic. The I-projection P* of Q onto a convex set Π of probability measures is defined as the minimizer of $D(P || Q)$ subject to $P \in \Pi$. Gibbs' conditioning principle asserts the convergence of the conditional distribution of X1 on the condition $\hat{P}_n \in \Pi$ to P*, as $n \to \infty$. A rather simple calculation using properties of I-divergence suffices to prove a strong version of this principle (Csiszár 1984).

## Inference principles motivated by IT

The maximum entropy principle, pioneered by Jaynes and Kullback, says in its simplest form the following: If an unknown distribution P ought to be inferred when knowing only that P belongs to a set Π of probability measures determined by linear constraints, one should infer that P equals the I-projection P* onto Π of a default model distribution Q (if Q is uniform then P* has maximum entropy, hence the name).

Algorithms designed for computing P* include iterative scaling and generalized iterative scaling (or SMART); these have intuitive IT interpretations that suggest how to prove their convergence to P*. These algorithms may also be used to compute maximum likelihood (ML) estimates. For computing ML estimates from incomplete data, statisticians often use the EM algorithm; this has an IT interpretation as alternating I-divergence minimization.

Rissanen's minimum description length principle (MDL) says that for stochastic modeling of given data when a list of candidate models or model classes is available, the one leading to shortest description of the data should be chosen, assuming a code "ideal" for the chosen model (or class), and taking into account that also the latter has to be described. MDL gives statistical significance to the theory of source coding, and is a rich source of applications of IT to statistics.

## Other topics

Starting from the late fifties, several authors have used IT methods (basically the properties of I-divergence or of related functionals of probability measures) to prove various limit theorems in probability theory.

In statistics, the amount of information in the sample about the unknown parameter has been studied, starting with Rényi (1969) and Pinsker (1972), and the results were used to derive bounds on how well that parameter can be estimated. Recently, tight risk bounds on non-parametric density estimation were derived in a similar manner (Yu 1995, Yang and Barron 1997).

Hypothesis testing or estimation based on remote data may involve the feature that encoding the data before transmission is permissible, subject to rate constraints. Such problems jointly belong to IT and statistics. Their study began in the mid eighties, and offers many challenges for future research.

Fascinating recent applications of IT to probability theory involve the currently hot topic called measure concentration. An early measure concentration result known as the blowing up lemma has been a standard tool in IT, but an information theoretic proof of it was found only in 1986, by Marton. Recently, she was able to prove measure concentration theorems, via her IT method, under substantially weaker assumptions than independence; other approaches have not made this possible, so far.

## Shannon and Investment

*Thomas M. Cover*
*Stanford University*

The theory of growth rate optimal investment is beginning to look more and more like information theory. Although Shannon never published in this area, he gave a well-attended talk on the subject in the mid 1960s at MIT, and the influence of information theory has been substantial.

The first work with an information theoretic flavor in investment theory was Kelly's [1956] BSTJ paper in which he proved that the increase in the growth rate of wealth in a horse race due to side information was equal to the mutual information between the winner of the horse race and the side information. Breiman [1961] generalized this result and proved that Kelly gambling (gambling in proportion to the probability of winning) had a higher growth rate of wealth than any other investment scheme and that it minimized the time necessary for the wealth to achieve a distant goal. In the mid 1960s, Shannon gave a lecture on maximizing the growth rate of wealth and gave a geometric Wiener example.

At about this time, Shannon and Samuelson (a Nobel Prize winner-to-be in economics) held a number of evening discussion meetings on information theory and economics. It is not clear what was said in these meetings, but Samuelson seems to have become set in his views. He published several papers arguing strongly against maximizing the expected logarithm as an acceptable investment criterion. (It happens that maximizing the expected logarithm is the prescription for the growth-rate optimal portfolio.)

For example, Samuelson [1969] wrote, "Our analysis enables us to dispel a fallacy that has been borrowed into portfolio theory from information theory of the Shannon type." Samuelson goes on to argue that growth rate optimal policies do not achieve maximum utility unless one has a logarithmic uility for money. Of course this is the case, but it does not deny the fact that log optimal wealth has an objective property: it has a better growth rate than that achieved by any other strategy. Since growth rate optimal policies achieve a demonstrably desirable goal, growth rate optimal portfolios should only have a utility interpretation as an afterthought. In fact, Samuelson [1979] wrote a paper entitled, "Why we should not make mean log of wealth big though years to act are long." This is a two page paper in words of one syllable that makes the point that maximizing the expected log of wealth is not appropriate. The growth optimal portfolio literature has been slow to develop. It is possible that Samuelson's eloquent admonitions had their effect.

Thorp [1969] proved, among other things, that the growth rate optimal portfolio is not necessarily on the efficient frontier, thus showing the incompatibility of log optimality and the mean-variance theory of Markowitz. Thinking that a portfolio with several good properties must have more, Bell and Cover [1980, 1988] showed that the log optimal portfolio is also competitively optimal. Thus, the long run optimal portfolio is also optimal in the short run in the sense that it outperforms the wealth induced by any other portfolio, at least half the time. This leads to a similar result for Shannon coding, proving that the ideal code lengths log $(1/p(x))$ are competitively optimal as well as expected length optimal [Cover, 1991b].

The value of side information, first investigated by Kelly, was generalized by Barron and Cover [1988] to show that the increase in the growth rate of wealth is less than or equal to the mutual information.

Algoet and Cover [1988], generalizing Breiman, showed that the conditionally log optimal portfolio is asymptotically optimal in growth rate in ergodic markets and the growth rate converges to a constant. As a special case of this convergence argument, a new sandwich proof was given of the AEP.

There have also been some results in establishing universal portfolios—the counterpart to universal data compression. A universal portfolio attempts to do as well on the fly as if one had known ahead of time the precise sequence of stock market returns, and used the best constant rebalanced portfolio. (This restriction of the best constant rebalanced portfolio is naturally motivated since they are optimal for markets which have independent identically distributed investment opportunities.) Cover and Gluss [1986], Cover [1991a], and Ordentlich and Cover [1996, 1998] have proved at various levels of generality, that there exists a universal portfolio achieving a wealth $\hat{S}_n$ at time n such that $\hat{S}_n / S_n^* \geq 2/\sqrt{n+1}$ for every stock market sequence and for every n, where $S_n^*$ is the wealth generated by the best constant rebalanced portfolio with hindsight. In fact, the lower bound corresponds to the associated minimax regret lower bound for universal data compression. And since $S_n^*$, the best wealth achievable in hindsight, is expected to grow exponentially, the $\sqrt{n}$ term is asymptotically negligible in the exponent. Thus, one has the same asymptotic growth rate of wealth as if one had known the exact stock market sequence ahead of time.

The growth rate optimal portfolio theory development has gone hand in hand with the counterpart theorems in data compression and universal data compression. Shannon's influence on portfolio theory, although entirely indirect, has been substantial. If one identifies a result as being information theoretic if it involves entropy, mutual information, channel capacity, and asymptotic equipartition properties, one would have to say that this growing segment of portfolio theory is a proper subject of information theory.

## References

[1] J. Kelly, "A new interpretation of information rate," *Bell System Tech. Journal*, pp.917-926, 1956.

[2] H. Latané, "Criteria for choice among risk ventures," *Journal of Political Economy*, 67:144-155, 1959.

[3] L. Breiman, "Optimal gambling systems for favorable games," Fourth Berkeley Symposium, 1:65-78, 1961.

[4] P. Samuelson, "General proof that diversification pays," *Journal of Financial and Quantitative Analysis*, 2:1-13, 1967.

[5] P. Samuelson, "Lifetime portfolio selection by dynamic stochastic programming," *Rev. Econom. Statist.*, pp.239-246, 1969.

[6] E. Thorp, "Optimal gambling systems for favorable games," *Rev. Internat. Statist.* 37:273-293, 1969.

[7] P. Samuelson, "Why we should not make mean log of wealth big though years to act are long," *Journal of Banking and Finance III*, pp. 305-307, 1979.

[8] R. Bell and T. Cover. Competitive Optimality of Logarithmic Investment. *Mathematics of Operations Research*, 5(2):161–166, May 1980.

[9] T. Cover and D. Gluss. Empirical Bayes Stock Market Portfolios. *Advances in Applied Mathematics*, (7):170-181, 1986. Summary of this paper appears in: Proceedings of Conference Honoring Herbert Robbins, Springer-Verlag, 1986. Abstract and Summary appears in "Adaptive Statistical Procedures and Related Topics," IMS Lecture Notes Monograph Series, Vol. 8, ed. by J. Van Ryzin.

[10] A. Barron and T. Cover, A Bound on the Financial Value of Information. *IEEE Transactions of Information Theory*, 34(5): 1097-1100, September 1988.

[11] P. Algoet and T. Cover. Asymptotic Optimality and Asymptotic Equipartition Properties of Log-Optimum Investment. *The Annals of Probability*, 16(2):876-898, 1988.

[12] R. Bell and T. Cover. Game-Theoretic Optimal Portfolios. *Management Science*, 34(6): 724-733, June 1988.

[13] T. Cover. Universal Portfolios. *Mathematical Finance*, 1(1): 1-29, January 1991.

[14] T. Cover, On the Competitive Optimality of Huffman Codes. *IEEE Transactions on Information Theory*, 37(1): 172-174, January 1991.

[15] T. Cover and E. Ordentlich. Universal Portfolios with Side Information. *IEEE Transactions on Information Theory*, 42(2):348-363, March 1996.

[16] E. Ordentlich and T. Cover. The Cost of Achieving the Best Portfolio in Hindsight. To appear in *Mathematics of Operations Research*.

# Shannon and Cryptography

*James L. Massey*
*Prof. em., ETH-Zürich*

Martin Hellman, co-founder with Whitfield Diffie of "public-key cryptography", attributes his interest in cryptography to three sources, one of which was the fact that in 1970, Prof. Peter Elias of MIT introduced him to "Shannon's virtually forgotten 1949 paper on cryptography" (citation from the September 1981 IT-Newsletter article reporting that Diffie and Hellman had received the Donald G. Fink Prize Paper Award). Hellman has stated elsewhere that it was the following words of Shannon from that paper [1] which put him on the path to public-key cryptography: "The problem of good cipher design is essentially one of finding difficult problems, subject to certain other conditions. We may construct our cipher in such a way that breaking it is equivalent to (or requires at some point in the process) the solution of some problem known to be laborious."

But Shannon's contribution to cryptography go deeper that just being the godfather to public-key cryptography. His paper [1] is now generally credited with transforming cryptography from an art to a science. He was the first to give a system diagram of a secrecy system showing the message *M*, the cryptogram *E* (which was Shannon's notation for the Enciphered message) and the key *K* in the manner that (1) *M* and *K* determine E, (2) *E* and *K* determine *M* (as is necessary for decryption), and (3) the "enemy cryptanalyst" has access only to *E* but wants to find *M*. (To Shannon, *M* was always the totality of plaintext that is enciphered before the key is changed.)

One great service of Shannon was to say precisely what it

[1] The reader will be referred to books whenever possible.

[2] Like air, the parallels between Information Theory and Computational Complexity can be too ubiquitous and familiar to note. I apologize for my surely nermerous omissions.

[3] Shannon's technique of deriving lower bounds on the complexity of functions by counting appropriate configurations and taking the logarithm of the result has since been called in Complexity Theory "the information-theoretic argumment."

means for a cipher to be unbreakable. More precisely, he defined *perfect secrecy* to mean that $M$ and $E$ are statistically independent or, equivalently, that $H(M \mid E) = H(M)$. As my yellowed handwritten notes from one of Shannon's lectures in 1961 at MIT state, "It is obvious that $H(M \mid E) \leq H(K \mid M)$." Hence for perfect secrecy we must have $H(M) \leq H(K \mid M) \leq H(K)$, i.e., the length of the key in binary digits must be at least as great as the number of bits of information in the message that is being hidden. Patent offices around the world will be forever grateful for this proof that there is no unbreakable cipher with a short key. Shannon also demonstrated that the Vernam cipher or "one-time pad" in which a coin-tossing key is added bit-by-bit modulo-two to the message is unbreakable. Vernam knew this, too. In his 1926 paper [2] describing this cipher, he wrote that its unbreakability had been confirmed by field trials with the U.S. Army Signal Corps.

Shannon defined the *unicity distance* of a cipher as the amount of ciphertext that determines the key $K$ essentially uniquely and showed, for what he called a "random cipher", that this was closely approximated by $\frac{H(K)}{D}$, where $D$ is the per letter redundancy of the original language. This formula is still routinely used to estimate the unicity distance of virtually all ciphers.

Shannon also dealt with "practical secrecy" in [1], i.e., with ciphers that are difficult to break rather than impossible. To this end, he introduced the *work characteristic*, $W(N)$, of a cipher, which he defined as the "average amount of work to determine the key for a cryptogram of $N$ letters measured say in man hours". Forty-nine years later, no one has been able to determine the work characteristic, or to provide a nontrivial lower bound thereon, for any practical cipher. In a very real sense, we have not made much progress on the *foundations* of secrecy systems since Shannon's paper [1]. The reader will surely object to this assertion and ask "what about public-key cryptography, didn't Shannon miss this highly important development?" My answer is that the jury is still out (and is likely out to remain out for a long time) as to whether public-key secrecy systems are possible or just a *fata morgana*. One-way functions are the heart of this new kind of cryptography but there is no proof that such functions even exist. We are building a most imposing edifice on a foundation that has not been tested for its solidity. A new Shannon is sorely needed to test the soil for us.

Thirty years ago one might well have spoken of Shannon's "virtually forgotten 1949 paper", but no longer. At the EUROCRYPT meeting this May in Helsinki, several of the younger researchers in cryptography bombarded me with questions about Shannon, the man, and whether it would be possible to undertake a kind of pilgrimage to meet him. Cryptographic researchers today are very famililar with [1] and very appreciative of Shannon's contribution to their field. Whitfield Diffie insisted to me at this same meeting that it was Shannon's work in cryptography that led him to information theory, rather than the reverse. But I recall reading an interview in which Shannon related that secrecy systems provided him with a good application for the general techniques that he was developing in his information theory. That sounds more plausible to me, especially when I read in [1] that: "From the point of view of the cryptanalyst, a secrecy system is almost identical with a noisy communication system."

## References

[1] C.E. Shannon, "Communication Theory of Secrecy Systems", *Bell System Tech. J.*, vol. 28, pp. 656-715, Oct., 1949.

[2] G.S. Vernam, "Cipher Printing Telegraph Systems for Secret Wire and Radio Telegraphic Communications," *J. Amer. Inst. Electrical Eng.*, vol. 55, pp. 109-115, 1926.

# Information Theory and Computational Complexity: The Expanding Interface

*Christos Papadimitriou*
*Division of Computer Science, U.C. Berkeley*

Computational Complexity studies the intrinsic reasons why certain computations cannot be carried out efficiently [8].[1] Its affinity to Information Theory is therefore deep and broad. The two fields have close historical ties and technical parallels, they share methodology, and they often define one another in important ways. In this note I will identify and discuss some of the most important points of contact;[2] see [1] for a collection of articles on the same subject ca. 1988.

Perhaps the most important of the historical ties between the two fields is one of near-fraternity: Although Computational Complexity was launched in the 1960s [3], and focused on its central contradistinction between polynomial and exponential computations shortly thereafter, a 1949 paper by Claude Shannon [11] was one of its clearest precursors. Shannon showed that almost all Boolean functions in n inputs need an exponential number of gates for their computation. Identifying an explicit and natural family of Boolean functions that requires an exponential (actually, even superlinear) number of gates has since been a central problem in Computational Complexity, with the P≠ NP conjecture being a most important refinement.[3] The subarea of circuit complexity which studies the complexity of synthesizing reliable circuits from noisy materials can be seen as an extension of Shannon's work on channels; see [9] for a survey.

*Kolmogorov complexity* [6] is in some sense a hybrid of the two fields, in which computation is used as a means for defining information content. Kolmogorov complexity has in fact been used extensively in proofs of results in Computational Complexity, replacing complicated averaging arguments by simpler arguments involving “typical” inputs.

Another field in the interface is *communication complexity* [4]. Two agents, each having half of the input bits, must compute a Boolean function by exchanging as little information as possible. Communication complexity is interesting because it helps in the proof of lower bounds on the parallel complexity of problems, and on the area and time needed to compute functions on chips.

*Public-key cryptography* explicitly uses complexity in order to achieve something that is information-theoretically impossible (a one-way bijection); it is therefore a conscious *break* from Information Theory. Many of the cryptographic protocols deriving from public-key cryptography (e.g. signatures, mental poker, zero-knowledge proofs, see [8] push this break further. This culture has in fact produced protocols such as secret-sharing [12], which do not depend on complexity. Andy Yao has coined the term “computational information theory” [15] to describe arguments in which information is computationally hard to obtain (or predict). In the related field of (pseudo-)randomness [5,7] computation is often used to “remove information,” effectively producing random bits.

The methodological intersection and cross-fertilization of the two fields has been considerable and fascinating. Shannon’s paper [10] is one of the first applications of the *probabilistic method* for establishing the existence of favorable configurations, an ingredient of many complexity results. The probabilistic method was used to first establish the existence of *expanders* (bounded-degree graphs in which every cut is proportional to the smallest piece), of which however we subsequently obtained explicit constructions. Expander graphs are essential tools in Complexity; they also resulted recently in the discovery of *expander codes* [13], explicit and asymptotically good error-correcting codes with linear-time decoding (first defined, without performance guarantees, by Gallager in the early 1960s). Codes of a novel kind were essential in developing *Probabilistically Checkable Proofs* [2], one of the deepest results in Complexity to-date. In return, complexity concepts have invaded and enlightened coding theory, see for example [14]; these complexity results about codes may eventually find applications in cryptography.

There is more, and there will be much more. The fields of Information Theory and Computational Complexity are as intertwined and complementary as communication and computation. I believe that they will continue to be as central, flourishing, indispensible, and mutually enhancing as they have been in the past.

**Acknowledgment**: Many thanks to Jon Kleinberg, Len Schulman, and Umesh Vazirani for elightening discussions.

## References

[1] Y.S. Abu-Mostafa (ed.) *Complexity in Information Theory*, 1988, Springer-Verlarg, 1988. pp.1-15.

[2] S. Arora, CV. Lund, R Motwani, M. Sudan, M. Szegedy “Proof verification and hardness of approximation problems,” *Proc. 1992 FOCS*, pp.~2—13.

[3] Juris Hartmanis and Richard E.~Stearns “On the computational complexity of algorithms,” *Trans. of the AMS, 117*, pp.~285-306, 1965.

[4] Eyal Hushilevitz and Noam Nissan *Communication Complexity*, Cambridge Univ.Press, 1997.

[5] Russell Impagliazzo, David Zuckerman, “How to Recycle Random Bits,” *Proc. 30th FOCS 1989*, pp. 248-253.

[6] Ming Li and Paul Vitanyi, *An Introduction to Kolmogorov Complexity and its Applications*, 2nd ed., Springer-Verlag, New York, 1997.

[7] Michael Luby *Pseudorandomness and Cryptographic Applications*, Princeton Univ. Press 1995.

[8] Christos H. Papadimitriou *Computational Complexity*, Prentice-Hall, 1994.

[9] N. Pippenger’s “Developments in ‘the synthesis of reliable organisms from unreliable components’”, in *Proceedings of Symposia in Pure Mathematics, vol. 50*, p. 311-324, 1990.

[10] Claude Shannon “A mathematical theory of communication,” *Bell System Technical Journal, 27*, pp. 379-423, 623-656, 1948.

[11] Claude Shannon “The synthesis of two-terminal networks,” *Bell System Technical Journal, 28*, pp. 59-98, 1949.

[12] A. Shamir “How to share a secret,” *Communications of the ACM*, 22: 612-613, 1979.

[13] “Expander Codes,” Michael Sipser and Daniel A. Spielman, to appear in *IEEE Transactions on Information Theory*, extended abstract in *Proceedings of the 35th Annual IEEE Conference on Foundations of Computer Science*, 1994, pp. 566-576.

[14] Alexander Vardy, “Algorithmic complexity in coding theory and the minimum distance problem,” in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 92-109, El Paso, Texas, 4-6 May 1997.

[15] A.C.-C. Yao “Computational Information Theory,” in [1] pp.1-15, 1988.

# Interactions between Information Theory and Systems, Control and Signal Processing

*T. Kailath*
*Stanford University*

In the early 1950’s, the Information Theory Group, and the older Circuits group, were home to most of the mathematically oriented papers of the day, with those in the IT group having more of a probabilistic focus. By the mid-sixties the Automatic Control structure in the problem (see the review by Liu et al., Signal Proc. 50, 83-99, 1996).

### 5. *H*-∞ Criterion for Decision-Feedback Equalization

As a final example, we note that in the last decade control theorists have been studying an alternative to the mean-square-criterion to allow for imperfect knowledge and noise models and statistics. This is a form of minimax approach, protecting against the worst possible disturbances; it is called the *H*-∞ criterion for reasons too long to go into here. A new SIAM monograph by Hassibi, Sayed and Kailath gives a unified treatment of the mean-square error and *H*-∞ approaches by using an indefinite metric space formulation in place of the traditional Hilbert space setting for Wiener and Kalman filtering. Decision feedback equalization is a problem where the effect of decision errors is hard to model, so that mean-square analysis of such equalizers is generally based on the assumption of no decision error. Recent work by Erdogan et al., to be presented at the August 98 IT Symposium, carries out the analysis under the *H*-∞ criterion, which allows the results of errors to be modeled as an unknown.

## Quantum Information Theory

*A. R. Calderbank*
*Information Sciences Center, AT\&T Labs – Research*

Niels Bohr once remarked "Anyone who can contemplate quantum mechanics without getting dizzy has not properly understood it." The cascade of recent results in quantum information theory and quantum computing would have really made Bohr's head spin. According to Bennett and Shor [BS98]

*"It has become clear that an information theory based on quantum principles extends and completes classical information theory, somewhat as complex numbers extend and complete the reals. The new theory includes quantum generalizations of classical notions such as sources, channels and codes, and two complementary, quantifiable kinds of information – classical information and quantum entanglement. Classical information can be copied at will, but can only be transmitted forward in time, to a receiver in the sender's forward light cone. Entanglement, by contrast, cannot be copied, but can connect any two points in space-time. Conventional data processing operations destroy entanglement, but quantum operations can create it, preserve it, and use it for various purposes, notably speeding up certain classical computations and assisting in the transmission of intact quantum states."*

Classical bits take the values 0 or 1 at all times but quantum bits or qubits can occupy a superposition of the 0 and 1 states. This is not to say that the qubit has some intermediate value between 0 and 1. Rather the qubit is in both the 0 state and the 1 state at the same time to varying extents. Mathematically a *qubit* is a 2-dimensional Hilbert space, and a quantum state is a vector

$$\alpha|0\rangle+\beta|1\rangle,\ \text{where } |\alpha|^2+|\beta|^2=1.$$

A system with *N* different 2-state memory cells is realized as the tensor product of the individual 2-dimensional Hilbert spaces, so we are led to vectors

$$\sum_{v\in V}\alpha_v|v\rangle,\ \text{where } V=\mathbb{Z}_2^N \text{ and } \sum_{v\in V}|\alpha_v|^2=1$$

When the $\alpha|0\rangle+\beta|1\rangle$ is measured with respect to the basis $|0\rangle,\ |1\rangle$ the probability that the qubit is found in a particular state is the square of the absolute value of the corresponding amplitude. An isolated quantum system evolves in such a way as to preserve superpositions and distinguishability; mathematically such evolution is a unitary transformation (i.e. linear and inner-product-conserving), the Hilbert-space analog of rigid rotation in Euclidean space. Unitary evolution and superposition are the central principles of quantum mechanics.

To demonstrate that quantum information is quite different from classical information we examine the quantum coin flip (QCF) described by the unitary matrix

$$\frac{1}{\sqrt{2}}\qquad \begin{array}{c|cc} & |0\rangle & |1\rangle \\ \hline |0\rangle & 1 & -1 \\ |1\rangle & 1 & 1 \end{array}$$

We calculate the effect of placing two QCF gates in series.

The end-state of the system is obtained by computing the weight of each leaf and summing these weights. The coefficients of $|1\rangle$ interfere destructively and the coefficients of $|0\rangle$ interfere constructively. The result is $(QCF)^2 = NOT$. By contrast, any number of classical coin flips in series are equivalent in function to a single coin flip.

The first work connecting information theory and quantum mechanics was that of Rolf Landauer and Charles Bennett of IBM T. J. Watson Research Lab. Their motivation was Moore's law; every two years for the past 50 years, computers have become twice as fast while components have become twice as small. As the components of computer circuits become very small their description must be given by quantum mechanics. Bennett and Landauer wanted to understand how small components of circuits could be made and how much energy was required for computation. Over time there developed curiosity about the power of quantum computation. Then in 1994 Peter Shor was able to exploit quantum superposition to find a fast algorithm for factoring integers. This was the first example of an important problem that a quantum computer could solve faster than a classical computer, and an indication that quantum computation might be more powerful than classical computation. The effectiveness of quantum computing is founded on coherent quantum superposition or entanglement which allows exponentially many instances to be processed simultaneously. However no quantum system can be perfectly isolated from the rest of the world and this interaction with the environment causes *decoherence*: the environment measures the quantum system collapsing the wave packet.

In classical computing one can assemble computers that are much more reliable than any of their individual components by exploiting error correcting codes. In quantum computing this was initially thought to be precluded by the Heisenberg Uncertainty Principle (HUP) — observations of a quantum system, no matter how delicately performed cannot yield complete information on the system's state before observation. For example we cannot learn more about a single photon's polarization by amplifying it into a clone of many photons — the HUP introduces just enough randomness into the polarizations of the daughter photons to nullify any advantage gained by having more photons to measure. At first it was believed that the quantum no-cloning theorem makes error correction impossible in quantum communication and computing because redundancy cannot be obtained duplicating quantum bits. This is not so — only repetition codes are eliminated. The trick is to take quantum superposition + decoherence, to measure the decoherence in a way that gives no information about the original superposition, and then to correct the measured decoherence. For details, including a beautiful group theoretic framework for code construction see [2].

## References

[1] C. H. Bennett and P. W. Shor, Quantum Information Theory, *IEEE Trans. Inform. Theory*, to appear.

[2] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, Quantum Error Correction via Codes over GF(4), *IEEE Trans. Inform. Theory*, to appear.

# Reflections of Some Shannon Lecturers

*The following brief commentaries by thirteen recipients of the Shannon Award provide enlightening personal insight into the minds and attitudes of those who reached the pinnacle of achievement in our field.*

*The Eds.*

## Robert M. Fano
### (1976 Shannon Lecturer)

The following quotation is from the preface to my 1962 book, *Transmission of Information:* "My interest in Information Theory dates back to the Summer of 1947 when, after completion of my doctorate thesis, I began looking for greener research pastures. ... My curiosity was particularly aroused by Professor Wiener's frequent statement that the information associated with a message depended on the ensemble from which it was selected, and that its average value could be identified with the entropy of the ensemble. This notion was so strange to me that I felt the need of some operational justification for it. By March 1948 I had obtained a justification in terms of message encoding, which turned out to be very similar to a theorem already proved by C. E. Shannon, but still unpublished. It was in this connection that I had the pleasure of meeting Dr. Shannon for the first time. I was so impressed by the scope and profoundness of his work that I have followed in his footsteps ever since. This book is primarily an account of his work, and of work inspired by him, either directly or indirectly." I was particularly amazed by Shannon's noisy channel theorem, its engineering implications, and the scope of the underlying probabilistic phenomena. This had a major influence on my own work and that of my thesis students.

## Peter Elias
### (1977 Shannon Lecturer)

Fifty years ago I had completed a Master's program in computation and further coursework at Harvard and was looking for a doctoral thesis topic when Shannon's paper came out. It was an amazing piece of work. As the Russian mathematician Khinchin said in a 1956 paper, "Rarely does it happen in mathematics that a new discipline achieves the character of a mature and developed scientific theory in the first investigation devoted to it ... so it was with information theory after the work of Shannon." I was fascinated, finished a thesis in information theory in 1950 and have continued working in the domain ever since, the first three years as a Harvard postdoc and since 1953 at MIT. I joined a group that Bob Fano, who had explored some of the same questions, was starting in Jerry Wiesner's Research Laboratory of Electronics. Shannon came to MIT from Bell Labs for a visit in 1956, and came to stay in 1958: he gave a wonderful advanced topics course, opening new topics in many of the sessions, and was always open for discussion. It was a wonderful environment for graduate students and faculty. My favorite paper by Shannon since 1948 is "Prediction and Entropy of Printed English" — a delightful example of the playful diversity of his approach, particularly in the identical twin coding scheme for estimating the entropy of English. The talk I enjoyed most was his first Shannon Lecture, in Ashkalon in 1973, in which he dealt with the circularity of the occasion. He also made delightful gadgets. I liked best a box with a toggle switch on the front and a hinged cover. When you threw the switch up the cover opened slightly, an arm and hand came out, reached down, threw the switch back up and retreated into the box again as the cover closed. I miss that playfully creative mind.

## W. Wesley Peterson

### (1981 Shannon Lecturer)

Claude Shannon's book came to my attention about in 1953 when I was a graduate student at the University of Michigan studying vacuum tubes. I studied it from cover to cover and found it the most interesting material I had seen up to that time. I went to work for IBM, because I wanted to get involved with information and computing, and IBM gave me the unique and wonderful opportunity to attend the first course that Shannon taught at MIT, a truly inspiring experience for me.

Of course my work on error correcting codes was inspired directly by Shannon's work. I learned from him what an error correcting code is and he provided, with his fundamental theorem for the noisy channel, the goal for which I strived. I am happy that I could make some modest steps in that direction, and also that others have continued to make progress toward the goal that Shannon demonstrated for us.

Besides that, and maybe more important, Shannon taught us that at least some aspects of information can be dealt with quantitatively and I feel that from him I have acquired a much better understanding of this commodity information that I work with as a computer scientist. In the course of my teaching, I frequently find that I have insights that are a direct result of what I learned from him, for example in sorting, merging, searching, cryptography and cryptanalysis, and information coding.

## Irving S. Reed

### (1982 Shannon Lecturer)

I first heard of Claude Shannon and his work in early 1947 as a graduate student at Cal. Tech. Prof. E.T. Bell, after his course in mathematical logic, first told me of Shannon's 1938 AIEE paper on computer logic after my suggestion to Bell that logic statements could be realized by switches or relay logic. This landmark paper profoundly affected my early work on computer logic and architecture.

Later in 1952 at the MIT Lincoln Laboratory, physicist Ed Lerner lent me his copy of Shannon's famous paper (in book form), *The Mathematical Theory of Communication*, which laid the foundations of modern information theory. This paper made me painfully aware of how the reliability of the early digital processors could be improved, first by Shannon's coding theorem and more practically by the Hamming codes.

Because of my early work on processors for radar and communications, Shannon's information theory had an enormous impact on both my work and my personal development. Without doubt, Shannon's influence on me and on modern technology puts him in league with the greatest 20th Century scientist-engineers. In my case he led the way for me to consider seriously the possibility of algebraic error-correcting codes in the discovery and development of both the Reed-Muller and Reed-Solomon codes.

## Robert G. Gallager

### (1983 Shannon Lecturer)

Everything I needed to know about Information Theory, I learned from Claude Shannon. This includes not only his results about Information Theory, but also his way of doing research, which was beautifully balanced between a practical interest in how systems should work and theoretical beauty and generality. He never used a lot of high powered tools, but rather had an unsurpassed talent for finding the simplest non-trivial version of a problem, finding just the right way of looking at it, and then generalizing to a beautiful and general theory. In short, he gave us many existence proofs that research could be fun, simple, deep, and centrally important, all at the same time.

I remember once going to his office to talk about a puzzling problem I had been working on. He started throwing out pieces of it, one by one, until I was appalled at his "trivialization" of my problem. At a certain point, we could both understand it by inspection, and it was then easy to put all the complications back in.

## Solomon W. Golomb

### (1985 Shannon Lecturer)

I spent the summers of 1951 through 1954, while a graduate student in mathematics, working at the original Glenn L. Martin Co. in Middle River, Maryland. Two senior members of the group where I worked attended a special two-week course at M.I.T. (I think in 1953) and came back all enthused about Information Theory. I bought Woodward's book, *Probability, Information Theory and Applications to Radar*, and was pleasantly surprised to discover that this new branch of "engineering" was just as intelligible as mathematics.

In the spring of 1954, I drove from Harvard to New York to attend a special session on information theory "starring" Shannon, Elias, Huffman, *et al.*, at the annual I.R.E. convention. After spending a Fulbright Fellowship in Norway, I joined Caltech's Jet Propulsion Laboratory (JPL) in 1956. I spent most of the fall semester of 1959 at M.I.T. on leave from JPL. I sat in on Shannon's Information Theory course. I also had lunch at the same table with Shannon three times a week at the M.I.T. Faculty Club and had numerous discussions with him at other times as well. On one occasion, he asked if I could prove a result about probability distributions that he needed for some work he was doing concerning tails of martingales. He certainly got my attention and I was pleased to be able to present him with a proof the next day. (It turned out to be an easy proof once you found the right geometric way of looking at the problem.)

In June, 1985, I was the Shannon Lecturer at the ISIT in Brighton, England. Claude Shannon himself was in the audience! It was the first time that he had attended an ISIT since he gave the *first* Shannon Lecture in Ashkelon, Israel, in 1973. I chatted with him on a number of occasions in Brighton, but not on technical subjects of any depth. In general conversation he seemed totally lucid, but there were significant gaps in his remembrance of prior events.

For one of the most important technological innovators of the twenthieth century, Claude Shannon was remarkably modest and unassuming. I was one of many younger researchers whom he encouraged, and I certainly benefited significantly from my association with him.

## James L. Massey

### (1988 Shannon Lecturer)

I had the great luck as a graduate student at M.I.T. in the fall of 1960 to enroll in Professor Fano's course, Transmission of Information. His enthusiastic and entertaining lectures conveyed the beauty and excitement of information theory. I decided that this was the field for me and, in the following semester, enrolled in the course, Advanced Topics in Information Theory, which was taught by Shannon himself. I still treasure, and often re-read, my notes from those lectures. Shannon was a brilliant lecturer, if not in the classical style. His style of treating a subject was to present a sequence of increasingly complex examples, each having a solution that was obvious by inspection, until he had covered all the essential points, after which he would state the general theorem. I think that we graduate students all understood how to make the proofs then on our own. Whenever I have given a particularly good lecture since then, it has been because I consciously imitated Shannon's approach—the disasters have come when I forgot to do so.

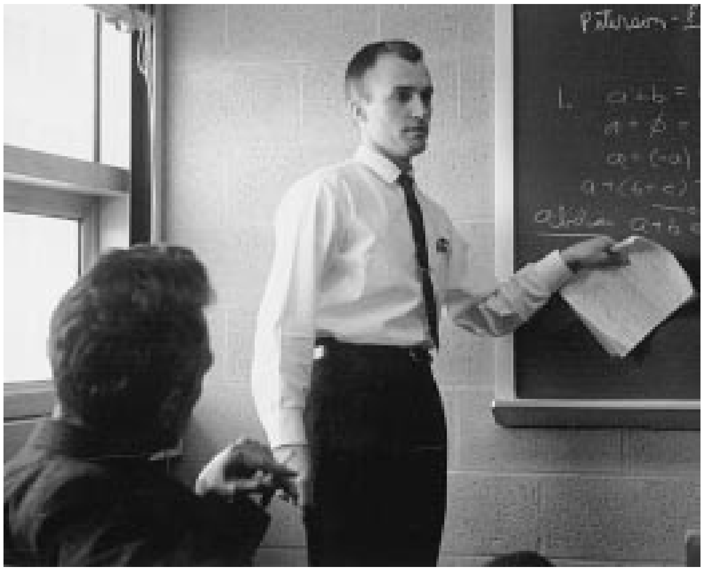


Photo courtesy of Jim Massey

Jim Massey as a young student under the watchful eye of C. Shannon.

Already then a very famous man, Shannon was nonetheless personally shy. He never seemed to me to be at ease in a crowd, but he was very relaxed and accessible to us struggling graduate students. Still it took me a long time to screw up my courage to the point where I dared to ask him to be a reader of my doctoral thesis, which he immediately agreed to do. He provided me with some excellent advice on my research—I also have copied the cross-examination technique by which he got me to explain what I was trying to do.

The second greatest honor of my life was being named a Shannon Lecturer. The greatest occurred during the 1986 ISIT in Ann Arbor, Michigan. Shortly before I gave my eminently forgettable talk, Claude and Betty Shannon entered the small lecture room, expressly to listen to me. To me this was one more proof that Claude Shannon is not only one of the great scientific figures of this century, but also a kind and generous human being.

## Thomas M. Cover

### (1990 Shannon Lecturer)

I bought two books with interesting titles the summer before starting graduate school at Stanford, Shannon's 1948 book on information theory and von Neumann and Morgenstern's book on game theory. Both books were extremely exciting. I spent over 100 hours using the game theory book to develop optimal strategies in various scenarios in poker. But Shannon's work seemed deeper and even more intriguing. I couldn't believe that something as intangible as information could be given such a satisfactory definition and have so many deep properties. I was also impressed by the relaxed and accessible writing style.

Shannon wrote his landmark paper as if the technical work had already been done and it was time to write an expository article on the subject. In fact, some of the greatest works in mathematics and physics have been written in this style. For example, Einstein, in his 1905 paper on relativity theory, had as one of his first equations: velocity = light path / time interval. And he described the notion of simultaneous events by saying, " 'The train arrives here at 7 o'clock,' means something like this: 'The pointing of the small hand on my watch to 7 and the arrival of the train are simultaneous events.' " This is craziness. Crazy or not, Shannon's paper is a great example of this tradition, in which no underlying intuition remains unrevealed. Indeed, the research literature in information theory is very readable, perhaps because of Shannon's influence.

As a result of Shannon's work, I have remained interested in the tangibility of information. One line follows the work of Kelly, in which Kelly shows that the increase in the growth rate of wealth from betting on a horse race is equal to the decrease in entropy. Another fascinating line of inquiry is the development of algorithmic complexity by Kolmogorov, Chaitin and Solomonoff in the mid 60s. The identification of the information in a sequence with its shortest computer de-

Photo courtesy of L. Zadeh

**Claude Shannon (right) with (from left) Joe Weizenbaum, Fredkin, and John McCarthy in April '68.**

scription leads to a concrete theory which is everywhere parallel to information theory.

Summing up, I would say that the results of Shannon, other than the major ones, of course, which have intrigued me the most are his proof that feedback does not increase capacity (which Gallager made transparent), Shannon's paper on the two-way channel, and Shannon's statement of the entropy power inequality.

There is certainly one piece of intriguing old business remaining. It is a quote from Shannon's 1959 paper on the fidelity criterion: "This duality can be pursued further and is related to a duality between past and future and notions of control and knowledge. Thus we may have knowledge of the past but cannot control it; we may control the future but have no knowledge of it." Shannon said he would write more about it in a subsequent paper, but the paper never appeared. I have tried, without success, to come up with an information theoretic statement on this subject equal to the summary.

## Andrew J. Viterbi
### (1991 Shannon Lecturer)

In my opinion, Shannon's contributions of 1948 and the subsequent decade are among the most remarkable and lasting theoretical achievements of the twentieth century. I have often remarked that the transistor and information theory, two Bell Laboratories breakthroughs within months of each other, have launched and powered the vehicle of modern digital communications. Solid state electronics provided the engine while information theory gave us the steering wheel with which to guide it.

Shannon theory not only establishes the limits on maximum efficiency of both source coding and channel coding, but it also points us in the right direction toward implementations which approach these limits. The vast majority of digital communication and broadcasting networks employ channel coding techniques based on Shannon theory. Furthermore, their designs are influenced by fundamental statistical ensemble concepts which were first expounded by Shannon in his founding 1948 paper. After nearly half a century of progressively more powerful coding techniques, we are within sight of the Shannon limit; almost error-free communication over a Gaussian channel can be achieved at above 80% of channel capacity, by iterative soft-decision decoding of concatenated codes. These so-called "turbo decoding" techniques require large block lengths and hence considerable delays to achieve such high efficiency, as clearly predicted by Shannon theory. As the very high speed data requirements of the Internet and similar multimedia applications become commonplace, delays which appear long in terms of bit times, become insignificant in real time, and such powerful methods will find wide usage. In a broader and more abstract sense, the spread spectrum techniques, which most digital wireless voice and data networks have already or will soon implement, are conceptually the logical extensions of Shannon theory.

Again we note that the sophisticated computation and memory intensive processing required for the implementation of Shannon theoretic concepts has become feasible and economically favorable through the amazing capabilities of solid state electronics to reduce size, power and cost while increasing speed, all by many orders of magnitude in the last three decades. This trend is likely to continue for some time to come.

## Elwyn R. Berlekamp
### (1993 Shannon Lecturer)

Claude Shannon has long been widely recognized as one of the foremost intellects of the 20th century. He discovered or invented Information Theory, which has become one of the key pillars of our digital society. He also made legendary contributions to topics now viewed as belonging to other fields, such as the applicability of Boolean algebra to the design of digital circuits, and the basic algorithm for computer-playing chess and checker programs.

He has also been a wonderful human being. He has been a major source of direct and indirect inspiration to me and to numerous others, and through a surprisingly small number of levels of indirection, to our entire community.

One unfamiliar with the man might easily assume that anyone who has made such an enormous impact must have been a promoter with a supersalesman-like personality. But such was not the case. He was actually a very modest man. Even though I worked with him directly over a period spanning several years, much of the influence he had on me was through others.

The earliest event in my career which I can remember occurred in 1946, when I was in the first grade and I learned to play the game called "Dots and Boxes." The second occurred around 1951, when I heard (by word of mouth) the problem of finding one off-weight coin mixed in with eleven

Photo courtesy of L. Zadeh

**Betty Shannon (right) with (from left) Fay Zadeh and Mrs. Simons; Shannon can be seen in the back.**

good coins, using only three weighings on a balance scale. That problem captured my interest in a big way. When I heard it had something to do with something called "Information Theory", I yearned to learn more about that subject.

Although there were no undergraduate information theory courses at MIT, my sophomore advisor was Peter Elias and he encouraged me to apply for a cooperative program at Bell Labs, where I began as a summer student in 1960. I was assigned to Ed David's department, where I became an apprentice to John L. Kelly, author of a paper originally entitled "Information Theory and Gambling" that had been published as "A New Interpretation of the Information Rate". Kelly was also very interested in games. I began learning Information Theory, and became acquainted with many other leading researchers then at Bell Labs including David Slepian, John Pierce, Ed Gilbert, Ed Moore, John Riordan, and Henry Pollak. Many of them had been former colleagues of Claude Shannon, and ALL of them regarded him with great awe. Back at MIT, I became acquainted with more faculty interested in information theory: Bob Gallager, Bob Fano, John Wozencraft, Irwin Jacobs, and eventually Claude Shannon himself. I oscillated between MIT and Bell Labs. At MIT, I was assigned a desk which was just being vacated by another student whom I then met very briefly; his name was Jim Massey. The first and only formal course I ever took in "Information Theory" was co-taught by Gallager and Fano. It was an extraordinarily eventful course, during which Gallager found a brilliant new proof of the coding theorem. David Forney was another student in that same class. I returned to Bell Labs in the summer of 1963 to work for David Slepian. During that summer he hired a new recruit named Aaron Wyner. Neil Sloane was recruited a few years later. Back at MIT, Gallager agreed to supervise my dissertation. The other members of my committee were Elias, Wozencraft, and Shannon.

I had had almost no direct contact with Shannon before he agreed to serve on my committee, but I soon found him to be quite open and accessible. Here are three memorable anecdotes from 1963-64:

- One day we crossed paths in MIT's infinite corridor, and he stopped to chat. I said I was going to the library to look up certain particular references including one that HE had published. He urged me NOT to do so; he said I'd learn more by first trying to solve the same problems from scratch by myself. This advice came as quite a shock to me.
- One day he warned me that this was NOT the time to buy any stocks. This also came as quite a shock, because he certainly knew that I was an impecunious student with insufficient funds to buy anything. But I soon realized that he had intellectual as well as financial interests in the markets, and he correctly sensed that I shared the former. He had designed an analog feedback circuit which was intended to simulate the market and its probable reactions to flows of funds moving in and out.
- The first time that I visited his home in Winchester was a truly unforgettable experience. His wife, who had also been a mathematician, was very supportive. There was a tightrope a couple feet above the ground, on which he and his children performed. His junior high daughter strapped a bungey chord from her belt around her unicycle, and then JUMPED ROPE on the unicycle! His garage contained several dozen unicycles, all home made, including some so small that no one had yet succeeded in riding them. He wanted me to help him learn how to juggle five balls. His hands were slightly smaller than average, and he had considerable difficulty getting started.

In 1964-1967 as a coauthor with Gallager and me, Shannon lived up to my very high expectations. He had so many ideas as to how the results might be generalized or improved that only a few of them came to fruition during the three-year period our pair of papers was in preparation. He was going so strong then that I could not imagine that this pair of papers would turn out to be his last publications in Information Theory.

In 1973 I was president of IT, the year we initiated the "Shannon Lecture", the forerunner of the Shannon Award. Of course Shannon was our first awardee; that conclusion was unanimous even before nominations opened. The symposium was in Israel, and it was my task to persuade him to accept. That turned out to be less difficult than we had feared. The bigger challenge, which came to me as another total surprise, turned out to be helping Shannon overcome his stage fright during the half-hour before his talk! Shannon feared that the audience expected a God-like performance beyond anything he could possibly deliver. Fortunately, however, he did manage to get started, and once he got into his prepared lecture on "Feedback", it went very well. In my opinion, it was the best Shannon Lecture among the dozen or so that I've attended.

## G. David Forney, Jr.

### (1995 Shannon Lecturer)

Claude Shannon had a quite direct and personal impact on my career in information theory, in the absence of which I might well be in some other field today.

I was introduced to information theory in a marvelous thermodynamics course taught at Princeton by John Wheeler. My term project for that course was a book report on Leon Brillouin's book on physics and information theory, in which Brillouin "explains" Maxwell's demon by accounting for the information that the demon requires to open and close his door properly.

As soon as I arrived as a graduate student MIT, I took the 6.574 course on information theory. This ultimately led to a master's thesis in information theory and quantum mechanics. I then spent an unhappy and unproductive six months wandering around in other areas like operations research looking for a Ph.D. thesis, having been advised that information theory was pretty much dead as a research topic.

However, in Spring Term 1964 I took Shannon's 6.575 course, "Advanced Topics in Information Theory." Shannon's teaching method was quite similar to Wheeler's: he talked about various problems that he had been interested in, what progress he had made, and what open questions still puzzled him. I started playing with some of these problems; I don't remember exactly which, and I don't remember making any great progress. Nonetheless, by the end of the term I was off and running in information theory again, and within the next year finished my doctoral thesis on concatenated codes.

It is clear to me that Shannon's course was the direct cause of my return to information theory. I am eternally grateful to him not only for founding this field, but also for luring me back into it. It has been a great place to work.

## Imre Csiszár

### (1997 Shannon Lecturer)

As a student of mathematics, I learned information theory from Alfred Renyi, an outstanding mathematician. Information theory was one of Renyi's favorite subjects, though he was more interested in breaking new ground than dealing with mainstream problems. Following his lead, I wrote my first papers on generalized information measures and their applications in statistics and probability. Only later, after having learned about channel coding theorems from Wolfowitz's book, did I come to read Shannon's classical paper.

I was fascinated to see that most problems then studied in information theory had actually been introduced in that paper and, even more amazingly, that Shannon also provided the main ideas for their solution. Trained as a mathematician in doing formal proofs, I have learned a lot from Shannon's paper in terms of developing the intuition necessary for serious research in information theory. Arguments short of formal proofs tend to be unconvincing to mathematicians and some, better than I, have been known to fail to really understand Shannon's paper, surely for this reason. I also struggled with this obstacle but already having had some limited technical knowledge in information theory helped me to overcome it, as I was able to visualize how the often informal arguments could be turned into formal proofs. Ever since I have strongly advised my students to read Shannon's paper towards the end of the course in order that they better benefit from it.

During my life as a scientist, my research interests have been concentrated on the one hand on strict-sense information theory as directly descended from Shannon' fundamental paper, a subject now known as "Shannon theory", and on the other hand on applications of information theory within mathematics. The first may surely be attributed to Shannon, through his paper, and the other to my teacher, Renyi. I am equally indebted to both.

## Jacob Ziv

### (1997 Shannon Lecturer)

I well remember my first visit to Shannon's own study.

I came to MIT to study for my D.Sc. degree in the fall of 1959. Being an R&D engineer, I already knew that it was Information Theory that I would like to learn and investigate. I had first encountered Claude Shannon's monumental contributions after reading and trying to understand Goldman's book on the subject. I was therefore excited when my wife, Shoshana, and I were invited one weekend in the fall of 1959 to an open-house party for all the new foreign students to take place at the Shannon's residence, a beautiful house on top of a hill.

If I remember well, Claude was out-of-town that weekend, but many of the EE faculty were there to host us and warmly greet us. The party took place on the Shannons' huge hillside lawn. We were all impressed by one of the many self-made gadgets: a cable car that took you all the way up to the house. But one could operate it only at dinner time! (A clear message to the Shannon kids to be home for dinner on time!)

After a warm welcome by the faculty, I decided to dare to have a look at Claude's own study. I was impressed by the sight of a huge blackboard behind his desk. The blackboard was covered by a green shade. I was suddenly facing a real dilemma: Should I dare to have a peek at some of Claude's most recent, yet unpublished great results? Finally, after a period of tense hesitation, I moved the shade slightly, only to find out that there was indeed a formula spelled on the blackboard, neatly written in big letters; $$ H = -\displaystyle\sum_i\, p_i \log p_i . $$ Claude was apparently ready for us, counting on the fact that at least some of us could not withstand the temptation \ldots \quad Since then, I never actually stop searching for many of the erased results on Shannon's own blackboard.

# A Great Mathematician's View of Shannon

## Andrei Nikolaevich Kolmogorov (1903-1987)

*[Editors' note: The following is a translation of Kolmogorov's foreword to the volume, Papers in Information Theory and Cybernetics by C. E. Shannon, that was published in Russian in 1963. Our Russian colleagues recommended publishing this foreword, not previously available in English, here as evidence of how long and deeply Shannon's work has been appreciated in that country.]*

In our age, when human knowledge is becoming more and more specialized, Claude Shannon is an exceptional example of a scientist who combines deep abstract mathematical thought with a broad and at the same time very concrete understanding of vital problems of technology. He can be considered equally well as one of the greatest mathematicians and as one of the greatest engineers of the last few decades. His role in the creation of cybernetics is quite singular. In contrast to Norbert Wiener, Shannon was not engaged in the dissemination of the new science or its systematization. What he did was to set up the fundamentals of information theory, which to a considerable extent predetermined the development of the general theory of discrete automata. These two subjects constitute large chapters in cybernetics and probably occupy the central position in that science.

The importance of Shannon's work for pure mathematics was not sufficiently appreciated at the outset. I recollect that, as late as 1954 during the International Congress of Mathematicians in Amsterdam, my American colleagues who were specialists in probability theory considered my interest in Shannon's work to be somewhat exaggerated and regarded his work to be primarily of engineering rather than mathematical value. Such views hardly need refutation now.

It is true that Shannon left to his followers the strict mathematical validation of his ideas in cases of real complexity. However, his mathematical intuition is amazingly correct. I know of only one case where his intuition seems to have deceived him: the formula for $\lambda$ at the end of Appendix 7 to his paper "The mathematical theory of communication" is open to doubt. [Editors' note: This is the last appendix in Shannon's 1948 paper and the formula in question is Shannon's expression for the "dimension rate", i.e., the average number of dimensions per second required to specify a member of an ensemble, in a general measure-theoretic setting.]

It seems unnecessary to advertise Shannon's work among the specialists in communication engineering today. Full publication of his works in Russian is extremely timely. Since the present volume is designed for an advanced reader, the editors have confined themselves to brief comments only where absolutely necessary. In view of the general prospects for the development of information theory and cybernetics, a short paper by Shannon himself, "the Bandwagon", seems to be of interest. A modest and businesslike approach to the achievements currently available in these fields is typical of Shannon.

The papers included in the present volume are divided into three sections. The first section and the last five articles of the last section have been edited by O. B. Lupanov. The second section, the first three articles of the third section, and the bibliography have been edited by R. L. Dobrushin.

**Claude Shannon playing chess with Russian World Champion M. Botvinnik.**

From *My Life and Travels with the Father of Fuzzy Logic* by Fay Zadeh, Photo ©1998 TSI Press, Albuquerque, NM USA

# Looking into the Future

## Views of Some Young Researchers

*As we transition to the post-Shannon period of Information Theory we felt it was important to hear from some promising younger researchers with regard to their views, motivation, and outlook as they entered and continue to work in their field. Here are their comments.*

*The Eds.*

## My Perspective on Information Theory

*Rajiv Laroia*
*Bell Laboratories, Lucent Technologies*

Attending the 50th anniversary celebration of Shannon's famous paper at Bell Labs recently, set me thinking about the huge impact that the field of information theory, started by Shannon fifty years ago, has had on the world of communications. Numerous researchers like Elias, Fano, Gallager, built on the foundations laid by Shannon and helped develop a theory that changed the way the world communicates. There is no doubt that information theory has strongly influenced the development of today's complex wireline and wireless telephone systems, magnetic and optical storage systems, optical communication systems, the ubiquitous internet.

Quite a remarkable achievement for a nonconstructive theory that tells us what is possible without telling us how. But knowing what is possible is like giving curious minds a puzzle for which the solution is yet unknown and telling them that the puzzle can be solved. Someone eventually figures it out and has a big impact. Researchers like Ungerboeck, Lempel & Ziv, took up the challenge and their work has resulted in significant advances in communication and storage technology.

Information theory has had a glorious fifty years, how much longer does it have to go? Is it dying or will it flourish for another fifty? Questions like these have sparked numerous debates in recent years. While no one has a definite answer for these questions, I do believe that the success of any field depends on its ability to make an impact on the lives of ordinary people who may never have heard about it. The half-century of success enjoyed by information theory was in no small part the result of such an impact. It could have many more great decades ahead if it continues to influence technology. If it only relies on the glory of its past, it might suffer the same fate as physics has over the last three decades. Helping information theory in this regard and providing it with numerous opportunities to make a significant impact is the emerging world of wireless communications where bandwidth is always a scarce resource and becomes even more precious as more and more people look to incorporate wireless communications into their everyday lives. It is for information theorists to ensure that their research will continue to drive technology and improve wireless communication systems. Visionaries like Viterbi, Forney, have ensured that the recent advances in information theory get incorporated into products and systems that touch the lives of ordinary citizens, whether through the use of cellular phones or modems to access the internet. Their efforts in this direction are perhaps even more significant than their enormous technical contributions. Continued success of information theory will require more information theorists to share their vision.

## The Accidental Theorist

*E. Telatar*
*Bell Laboratories, Lucent Technologies*

In a certain sense, I came to be in the Information Theory field by accident: While I was studying towards my BS in EE in Turkey, Erdal Arikan was a graduate student in the US and I was (and I still am) good friends with his younger brother Orhan. Erdal was studying information theory at MIT, and his brother told me what he knew of Erdal's field of research. Erdal had explained some aspects of information theory and network theory to Orhan, and I was intrigued by these even when told second hand. I checked out a copy of *Shannon and Weaver* from the library and was completely fascinated by it. I have to admit that I got an illegal copy made by the local photocopy store—I realized much later that buying the book would have been cheaper. Until then my exposure to EE was more via Electromagnetic Theory, Control Theory and the like with a heavy dose of partial differential equations. The communication theory courses were more concerned with analog systems also. Shannon's paper was quite an eye opener as a first exposure to discrete mathematics with its rather different methods of proof. Accordingly, the next year, when I was applying to graduate schools in the US, I listed Information Theory as one of my fields of interest (besides EM and Control). It goes to show my ignorance that I did not realize my good fortune when Bob Gallager extended a research assistantship to me—I had not heard of him until then. Nor had I heard of the predictions of Information Theory's demise; it was not the last time my ignorance helped me along. Even if I had heard these predictions, that whatever was to be done was done by Shannon and nothing was now left, I suspect I would have done nothing different. I was simply attracted to the beauty of Shannon's ideas, and did not particularly care what other people thought of the prospects.

As for the future, I don't think any of us has much of a chance of predicting what topics will become important. Given that, we might as well work on problems that we find interesting; I suspect that those of us making predictions are simply proposing the problems that interest them now as the topics of the future. Nonetheless, I am confident that the understanding and insight one gains from Information Theory will keep our field indispensable for communication practitioners for many years to come.

## My View of the Future

*Michelle Effros*
*Cal Tech*

My work in information theory is inspired by two strong personal leanings. First, I want to work on important problems, where my own definition of important seems somehow tied to the notion of applicability or "practice" in its broadest sense: application to technological innovation, to improving our basic mathematical understanding, and so on. Second, I value answers that are absolute and provable truths. Many seem to argue the implicit conflict between these two desires. Barbs about "ad hoc" practical solutions and "obscure" theoretical results are sadly common and, even more sadly, not unbacked by supporting examples. While information theory is not immune to this debate (nor to the potential pitfalls of which the debate reminds us), with its broad scope of interest, participation, and applicability, information theory is one field in which theory and practice not only coexist but in fact nourish each other.

In fact, history shows us that major advances are not the sole property of either a purely theoretical or purely practical bent. As applications change, they inspire us to pose new theoretical questions or to re-examine old ones in new ways. Examples are numerous, but to look at just one, I point to the case of multi-resolution coding. While multi-resolution or progressive transmission source codes have been around for over a decade, applications like the world wide web, mobile communications systems, and diversity coded satellite communications, with their need for embedded source descriptions at a wide variety of rates, have inspired new multi-resolution source codes and subsequent theoretical results. Inspiration is, of course, not a one way street. One has only to imagine what the field of communications would look like today without Shannon's original works and the field that grew from them to see the monumental impact that theory can have on practice.

As an information theorist, I find myself constantly seeking out the boundary between theory and practice: using the theory to build better practical systems and using applications to gain insight into the right questions to ask and what their answers might look like. In fact, it is the interplay between theory and practice that I find to be most personally rewarding and to which I look with the greatest excitement in terms of future research. In addition to the more traditional problems, internet-based communications, mobile communications, optical storage media, investing, and biology are just a few examples of the disparate fields in which information theory can provide answers and because of which information theorists are asking new questions or reconsidering old ones from a new light.

On this fiftieth anniversary of information theory, then, I put forward a few hopes for the future of our field. I hope that the diversity that currently exists within the field of information theory flourishes; that we continue both to tackle the established, mainstay problems of our field and to seek out the new challenges; and that we encourage and support the cycling evolution of problems between theory and practice.

## What Information Theory Taught Me

*David Tse*
*University of California Berkeley*

It is difficult not to fall in love with information theory, if not at first sight. But everyone does so for a different reason. For me, it is for the research philosophy it embodies as much as for the elegance of the results. And indeed it is the philosophy which has a broad and deep impact on my research outlook.

I remember vividly my first lecture in information theory. Bob Gallager told us that if knowledge is represented as a tree, then the role of theory is to shrink the tree rather than to grow it. To a first-year graduate student, that sounded really counter-intuitive! I had thought that my job was to add my own little branch to the tree! Only after some experience in research do I now understand him better. While new facts are constantly added to form branches and twigs of this tree, good theory provides a unified and coherent understanding of the facts so that things do not become unwieldy. This is the “simple model” approach to research. By using a set of basic examples, models and problems, the essence of a wide range of phenomenon can be captured, and insights developed. Information theory is such a spectacular successful example of this approach. The power of the discrete memoryless channel was such an eye-opener for me!

I am doing a reading course with my students on Shannon’s 1948 paper, and it really strikes us how original the problem formulations must have been at that time. It is perhaps not too off the mark to say that the formulations have as much an impact on how we think about communication problems as the solutions themselves. Given that transmission rate and error probability are both performance measures of a communication system, how can one even start to ask questions about the fundamental tradeoffs between them? Shannon taught me that by setting up the right questions, one cannot only have nice answers but more importantly get to the heart of the matter. This certainly works hand-in-hand with the simple model approach.

In this day and age of increasing pressure on “applied” research and fast turnaround time for solutions, I believe that the research philosophy information theory exemplifies is more important rather than less. With the proliferation of problems, models and results, it is crucial to be able to see through the fluff, seek simple models and ask the right set of questions that can stand the test of time. Otherwise research done today will be useless tomorrow! Based on the spectacular success of information theory, I still keep faith in this fundamental approach to research. This is despite teaching in a university on the edge of Silicon Valley madness!

Who knows what terrain my research journey will take me in the future, and who knows how many coding theorems I will still be using in a few years’ time? But one thing for sure: the research philosophy I learned from information theory will always be with me.

# Shannon Day at Bell Labs

*E. Telatar*
*Bell Laboratories, Lucent Technologies*

## Conference Report:

To celebrate the 50th anniversary of the publication of Claude Shannon’s “A Mathematical Theory of Communication,” the Mathematical Sciences Research Center of Bell Laboratories, Lucent Technologies, held a one day symposium on May 18th. About 350 participants attended the event, which was held at the Bell Labs in Murray Hill, NJ.

The day opened with the viewing of a short film, courtesy of Betty Shannon, featuring our hero, Claude Shannon, in his younger days. Made in the 1950s by AT&T, the film shows Shannon demonstrating Theseus, the electronic maze solving mouse, named after the classical Greek hero that defeated the Minotaur in the labyrinths of Crete. Shannon explains the relay circuits that govern the operation of the machine and presents it as an application of the research on telephone switches and networks going on then at Bell Labs. Shannon came across as a very able speaker—the expression on his face when he puts the mouse in a maze with no solution is priceless. Those interested in the operation of Theseus can see Sloane and Wyner, *Claude Elwood Shannon: Collected Works*, IEEE Press, 1993, pp. 681–7.

Jacob Ziv was the first technical speaker of the day. In his talk entitled “Entropy Has Many Faces,”’ Prof. Ziv demonstrated the importance of entropy in our understanding of information, choice and uncertainty. He began with a definition of entropy for stationary sources. This led to a discussion of some of the numerous roles of entropy: The entropy of an ergodic source is intimately related to the number of “typical sequences”’ from this source; the entropy of a source is a lower bound on the average number of bits per input letter achieved by any noiseless encoder; the entropy of an ergodic source is closely related to the self-information of a string from the set of “typical sequences;” the entropy of a source can be estimated by finding the length of the longest source output string that matches a substring of an earlier database of source output letters; the entropy of an $\ell$-th order Markovian source provides us with a lower bound on the length of the output string needed to be observed in order to get a good estimate of the probability distribution corresponding to the source. In short, the talk highlighted the significance of entropy in information theory.

In “Lessons Learned from Claude Shannon,” Robert Gallager shared his insights into Shannon’s success as a researcher: Shannon used curiosity as his main criterion in selecting research problems. He was interested in understanding the design of systems at a fundamental level. His approach to studying a problem was to simplify it as much as possible, but no further. In order to find the simplest coherent way of investigating a problem, Shannon often turned to simple toy examples and back-of-the-envelope calculations. Gallager noted that this kind of research is now an endangered species. If Shannon were starting his career today would he get tenure at a major research university or a position at a major research lab? "Probably not," Gallager surmised, but "given his presentation abilities demonstrated in the short film he would certainly find a job in marketing." The real legacy of Shannon's research, beyond all the beautiful theorems, is the demonstration that systems can be made understandable if we take the time to understand them.

Jack Wolf, in "Shannon Theory and Magnetic Recording" focused on modulation codes, and in particular on $(d,k)$ codes. Shannon's capacity theorem for noiseless channels yields, in straight-forward fashion, the capacity of these codes. The difficulty arises when one considers the two-dimensional analogs of $(d,k)$ codes, binary 2D arrays that satisfy a $(d,k)$ constraint in each row and column. Although Shannon hints at such problems in his paper in the form of conditions for the existence of crossword puzzles, almost nothing is known about the capacity of systems obeying two-dimensional constraints. One exception is a recent result by Kato and Zeger: unless $k = d + 1$, the capacity of a 2D system obeying a $(d,k)$ constraint is zero. Prof. Wolf considered a further specialization to 2D, $(d,\infty)$ codes, and showed numerical results by Blahut and Weeks. His conclusion is that there is much to be done in this field.

David Forney summarized the development of coding for the additive Gaussian noise channel in "Euclidean Space Coding." Advocating $\mathrm{SNR}_{\mathrm{norm}}$ as the more appropriate measure as compared to $E_b / N_0$, Dr. Forney distinguished two regimes of operation: bandwidth limited and power limited. Noting that in the power limited operation Turbo codes can get very close to capacity, he focused on the bandwidth limited regime. His survey considered the route to capacity via lattice codes and trellis codes. More recently, the discovery of turbo codes and the rediscovery of low density parity check codes lead to another such route. For non-white Gaussian noise channels water-pouring suggests a natural strategy via multiple carrier modulation. There is another way via equalization: precoding techniques that can be combined with coding allow us to do just as well on an arbitrary linear Gaussian channel as on an ideal Gaussian channel. Dr.Forney concludes that the "challenge set up by Shannon for the Gaussian channel is solved modulo 0.05dB."

**Some of the speakers awaiting the start of the session.**

Neil Sloane surveyed "The Sphere Packing Problem" from the point of view of "what would we tell Shannon about the progress in sphere packings." The sampling theorem gives us a geometric representation of communication over band limited channels. The design of good codes is then equivalent to finding dense packings in high dimensional spaces. Dr. Sloane summarized what is known about the densest packings, and certain constructions of lattice packings. He showed a simple construction which give many of the best known packings, and concluded with a list of "most wanted lattices."

The afternoon session started with a tribute to Aaron Wyner who had conceived of the symposium and brought this "all-star" cast of speakers together. His friends Toby Berger, Jack Wolf and Jacob Ziv gave presentations that reflected upon Aaron in "Remembering Aaron Wyner." The presentations were not only about Aaron's fundamental contributions to the field but also his warm personality and his devotion to his colleagues and friends. To summarize these talks is beyond the ability of this writer—you simply had to be there.

Jim Massey's talk was entitled "Shannon Theory and Contemporary Cryptology." Shannon, in his 1949 paper "Communications Theory of Secrecy Systems", defined a perfect secrecy system to be one in which the conditional entropy of the message given the cryptogram is equal to the entropy of the message. The standard example of a perfect secrecy system is the one-time pad also known as the Vernam cipher. Prof. Massey gave an elementary proof that for a perfect secrecy system the length of the key pad (in bits) has to be at least as large as the length of the plaintext (in bits). He then discussed Shannon's concept of typical ciphers and their unicity distance, as well as the use of compression to generate ideal ciphers. Finally, Prof. Massey pointed out that many modern systems are still based on Shannon's original principles of "confusion" and "diffusion." He ended with the question whether Shannon had missed the application of public key cryptosystems. His answer: Maybe.

Emre Telatar summarized the "Current Topics in Information Theory at Bell Labs." The analysis and fundamentals of wire-less systems constitute a high priority at Bell Labs. Aaron Wyner investigated a one-dimensional cellular system, reducing the problem to its bare essentials, in the best tradition of Shannon. Extending Gallager's work on wide-band fading channels, Emre Telatar and David Tse showed that under mild assumptions on the fading process white-like signals perform poorly on wide-band systems. Finally, much attention is paid to the use of multiple-antennas in improving the performance over fading channels. Another important topic in Bell Labs is turbo codes. The questions of why these codes perform well and why the iterative decoding process achieves almost maximum-likelihood performance are both being investigated. Dr. Telatar then pointed out that more abstract topics are still important in Bell Labs. He mentioned list decoding, zero-error codes, and improved bounds on the redundancy of universal codes.

In "Shannon Theory and Turbo Decoding" Andrew Viterbi started with a discussion of error exponents. In principle en-

Crowd of attendees commingles around the registration desk.

A view of the amphitheater at the start of the day.

coding is easy since almost all long codes are good, whereas decoding is hard. Algebraic block codes, convolutional codes and concatenated codes are attempts to introduce structure into the codes to make decoding easy. Concatenated codes deserve special attention—they contain the genesis of Turbo codes. The traditional form of concatenated coding suffers from the hard decisions the inner decoder makes. Dr. Viterbi took note of the BJCR and SOVA algorithms for making soft decisions and presented a simplified MAP decoder based on the Viterbi algorithm. The rest of the talk was about bounds on error probability. Bounding techniques based on Gallager's Ph.D. thesis that improve upon the union bound and the application of these techniques to turbo codes were considered. These techniques give bounds based on the distance profile of the code. Dr. Viterbi conjectured that turbo codes' heavy tails at the low-weight end of the distance profile will not allow serial turbo codes to achieve capacity. Dr. Viterbi singled out "how good is iterative decoding?" as an open problem.

Tom Cover outlined the connections between "Shannon Theory and Investing." After discussing the concepts of entropy, mutual information, AEP, and relative entropy, Prof. Cover moved on to the early work of Kelly and Breiman in log-optimal portfolios that maximize growth rate of wealth. Thorpe's work shows that log-optimal portfolios are not on the efficient frontier of undominated mean and variance pairs of returns. However, this poses no contradiction since log-optimality is the appropriate criterion for an investor who is willing to bet his entire capital at each time whereas the traditional measures of mean and variance of return apply to an investor who bets a constant amount each time. After discussing the hostility of traditional economists such as Samuelson to "portfolio theories borrowed from information theory of Shannon type" Prof. Cover moved on to present newer results. These include competitive optimality of log-optimal investing and universal portfolios both of which have counterparts in source coding. Prof. Cover's concluding slide pointed out the analogies between information theory and investing and let the audience draw its own conclusion as to the similarity.

In the final talk, "The Evergreen Legacy of Shannon to Communications Today," Robert Lucky shared his thoughts on the past and future of communication theory. Drawing a distinction between the "golden years" with "the years that we have now—which will remain nameless" he asked "why do we need to rationalize a Nobel prize by saying 'it is good for the telephones'?" Works such as Shannon's stand on their own beauty and do not need to be rationalized. Even if you never apply them, the intellectual stimulation one gets is sufficient reason to learn them. Dr. Lucky reflected upon the changing realities of the communication systems such as the switch from an analog system with additive Gaussian noise to a digital network with signal-dependent quantization noise. Asking "is bandwidth precious," he outlined the new challenges to communication theory in many-to-many communication, in understanding queuing, and in dealing with latency. He concluded that "the world has changed a lot since Shannon, but all these years the brilliance and beauty of it has shone on us all."

The talks clearly showed the diversity and vivacity of our field. The challenge is on the next generation to make the next 50 years as good as the first. The viewgraphs of the talks and Shannon's paper can be found at http://cm.bell- labs.com/shannonday.

# EUROPE OBSERVES GOLD JUBILEE

Among the many special events during this year a special colloquium was organized in the Netherlands to commemorate the 50th anniversary of the birth of Information Theory. It took place in June in Amsterdam, (too close in time to permit us to give you a full report) organized by Kees Immink and Sergio Verdu. It was sponsored by the royal Netherlands Academy of Arts and Sciences. There was a total of fifty invited attendees of whom sixteen gave in-depth presentations and led discussion on all aspects of the field. More details (and pictures) will be forthcoming in the future issues of the Newsletter. However, here is the program of the colloquium.

## Academy Colloquium
## Information Theory: The first 50 years and beyond

June 17-19, 1998
Amsterdam, The Netherlands

Venue: Barbizon Palace, Prins Hendrikkade 59-72, Amsterdam

### Programme

### Tuesday, June 16

| | |
|---|---|
| 19.00 hrs | Get-together<br>Venue: Brewery-Café “Maximiliaan”, Kloveniersburgwal 6/8, Amsterdam. |

### Wednesday, June 17

| | |
|---|---|
| 08.30-09.00 hrs | Registration |
| 09.00-09.15 hrs | Welcome by Prof Dr. J.E. Mooy on behalf of the Royal Akademie |
| 09.15-10.00 hrs | dick Blahut, Unviersity of Illinois, USA<br>Two-Dimensional Information Theory: A Task for the Coming Century |
| 10.00-10.45 hrs | Robert Calderbank, AT\T Labs Research, USA<br>The Future of Wireless Communications |
| 10.45-11.15 hrs | Imre Csiszar, Hungarian Academy of Sciences, Hungary<br>Arbitrarily Varying Channels: A Survey |
| 12.00-13.00 hrs | Lunch |
| 13.00-13.45 hrs | Ingrid Daubechies, Princeton University, USA<br>From Nonlinear Approximation Theory to Rate-Distortion Bounds |
| 13.45-14.30 hrs | G. Dave Forney, Jr., Motorola, Mansfield, USA<br>On Euclidean-Space Coding |
| 14.30-15.15 hrs | Tea break |
| 15.15-16.00 hrs | Joachim Hagenauer, TU Munich, Germany<br>Approaching Shannon’s Limit |

### Thursday, June 18

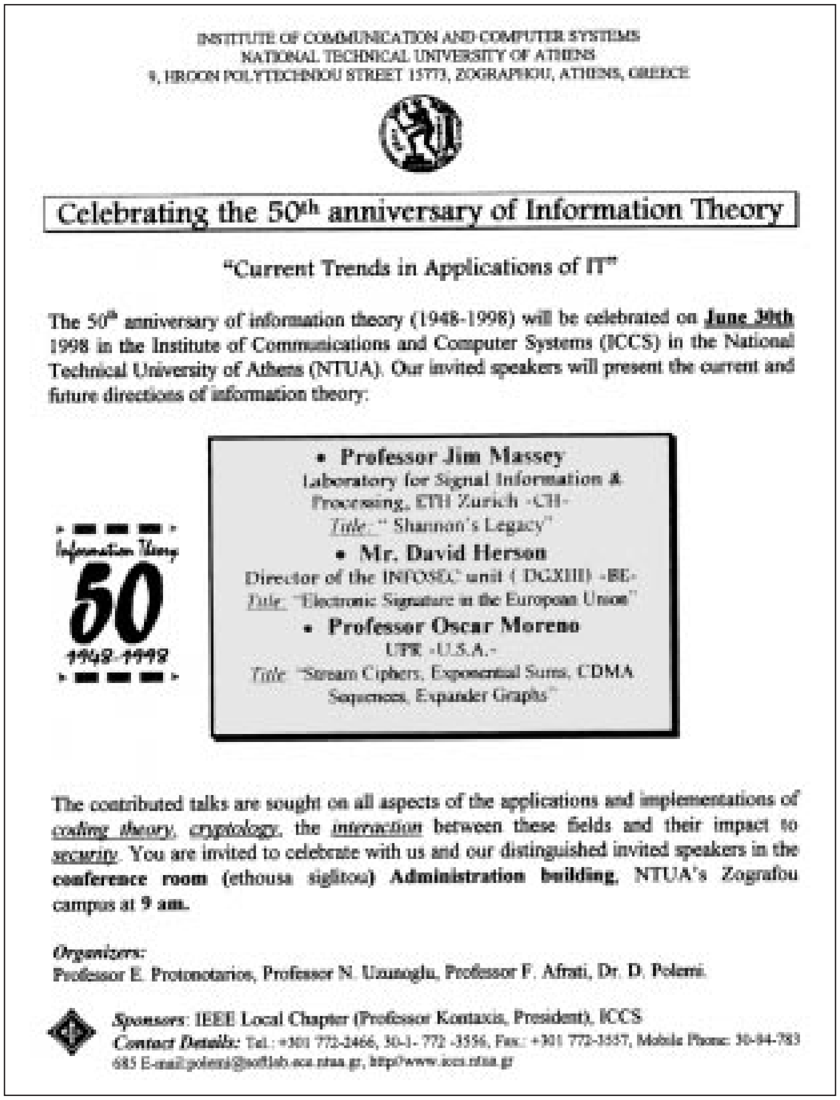
| | |
|---|---|
| 09.00-09.45 hrs | Te Sun Han, University of electro-communications, Tokyo<br>From the Method of Types Toward Information-Spectrum Methods |
| 0.945-10.30 hrs | Shu Lin, University of Hawaii, Honolulu, USA<br>Soft-Decision Decoding of Binary Linear Codes |
| 10.30-11.00 hrs | Coffee break |
| 11.00-11.45 hrs | Jack van Lint, TU Eindhoven, The Netherlands<br>Algebraic Geometry and Coding Theory |
| 11.45-13.00 hrs | Lunch |
| 13.00-13.45 hrs | James Massey, ETH, Zurich, Switzerland<br>Cryptography—A Theoretical Mess |
| 13.45-14.30 hrs | Shlomo Shamai (Shitz), Haifa, Technion, Israel<br>Information theoretic Perspective of Fading Channels |

| | |
|---|---|
| 14.30-15.15 hrsTea Break | |
| 15.15-16.00 hrs | Gottfried Ungerboeck, IBM Research, Zurich, Switzerland<br>New Challenge for Channel Coding: The Downstream Channel in PCM Modems |
| 20.00 | Banquet Doelen Hotel |

## Friday, June 19

| | |
|---|---|
| 09.00-09.45 hrs | Sergio Verdu, Princeton University, USA<br>Channel Capacity: Open Problems |
| 09.45-10.30 hrs | Victor Wei, chinese University of Hong Kong, China<br>Information theoretical Aspects of Electronic Money |
| 10.30-11.00 hrs | Coffee break |
| 11.00-11.45 hrs | Frans Willems, TU Eindhoven, The Netherlands<br>Weighting and Maximizing Methods in Noiseless Source Coding |
| 11.45-12.30 | The First Fifty Years |
| 12.30-13.30 hrs | Lunch |
| 14.00-15.00 hrs | Canal tour. Accompanied persons very welcome. |
| 15.00 | Visit to Maritime Museum |

Also, a special colloquium was organized at the National Technical University of Athens on June 30th to commemorate the Golden Jubilee of Information Theory. Below is the cover page of the announcement for that colloquium.



INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
9, HROON POLYTECHNIOU STREET 15773, ZOGRAPHOU, ATHENS, GREECE

**Celebrating the 50th anniversary of Information Theory**

"Current Trends in Applications of IT"

The 50th anniversary of information theory (1948-1998) will be celebrated on **June 30th** 1998 in the Institute of Communications and Computer Systems (ICCS) in the National Technical University of Athens (NTUA). Our invited speakers will present the current and future directions of information theory:

- **Professor Jim Massey**
  Laboratory for Signal Information & Processing, ETH Zurich -CH-
  *Title:* " Shannon's Legacy"
- **Mr. David Herson**
  Director of the INFOSEC unit ( DGXIII) -BE-
  *Title:* "Electronic Signature in the European Union"
- **Professor Oscar Moreno**
  UPR -U.S.A.-
  *Title:* "Stream Ciphers, Exponential Sums, CDMA Sequences, Expander Graphs"

The contributed talks are sought on all aspects of the applications and implementations of *coding theory*, *cryptology*, the *interaction* between these fields and their impact to *security*. You are invited to celebrate with us and our distinguished invited speakers in the **conference room** (ethousa sigitou) **Administration building**, NTUA's Zografou campus at **9 am.**

*Organizers:*
Professor E. Protonotarios, Professor N. Uzunoglu, Professor F. Afrati, Dr. D. Polemi.

*Sponsors*: IEEE Local Chapter (Professor Kontaxis, President), ICCS
*Contact Details:* Tel.: +301 772-2466, 30-1- 772 -3556, Fax: +301 772-3557, Mobile Phone: 30-94-783 685 E-mail:polemi@softlab.ece.ntua.gr, http//www.iccs.ntua.gr

# Special Issue of the IEEE Transactions on Information Theory

*Sergio Verdu*

To mark the fiftieth anniversary of the publication of "A Mathematical Theory of Communication" in July-October 1948, the IEEE Transactions on Information Theory will publish a special commemorative issue "Information Theory: 1948-1998" in October 1998.

Guest Editor Sergio Verdu has commissioned 25 retrospective articles from some of the foremost authorities in the fields of coding theory, Shannon theory, digital communications, data compression, networks, signal processing, statistical inference, and pattern recognition.

The table of contents of this special issue are as follows:

S.I. Amari (Riken, Tokyo) and T.S. Han (University Electrocommunications, Tokyo) "Statistical Inference under Multiterminal Data Compression."

A. Barron (Yale University), J. Rissanen (IBM, Almaden), and B. Yu (UC Berkeley) "The Minimum Description Length Principle in Modeling and Coding."

C. Bennett (IBM, T.J. Watson) and P. Shor (ATT Florham Park) "Quantum Information Theory".

T. Berger (Cornell University) and J. Gibson (Texas A&M) "Lossy Data Compression".

E. Biglieri (University of Torino), J. Proakis (Northeastern University), S. Shamai-shitz "Fading Channels."

I. Blake (Hewlett-Packard, Palo Alto), C. Heegard (Cornell University), T. Hoeholdt "Algebraic Geometry Codes."

R. Calderbank (ATT, Florham Park) "Coding Theory and Coding Practice."

D. Costello (Notre-Dame), J. Hagenauer (T. University Munich), H. Imai (University Tokyo) "Applications of Error Control Coding."

T. Cover (Stanford University) "Comments on Broadcast Channels."

I. Csiszar (Hungarian Academy of Sciences) "The Method of Types."

D. Donoho (Stanford University), I. Daubechies (Princeton University), R. Devore (Princeton University), M. Vetterli, "Data Compression and Harmonic Analysis."

P. Delsarte (University of Louvain) and V. Levenshtein (Russian Academy of Sciences,) "Association Schemes and Coding Theory."

A. Ephremides (University of Maryland) and B. Hajek (University of Illinois) "Information theory and Communication Networks: An Unconsummated Union."

M. Feder (Tel Aviv University) and N. Merhav (Technion) "Universal Prediction."

D. Forney (Motorola) and G. Ungerboeck (IBM, Zurich) "Modulation and Coding for Linear Gaussian Channels."

R. Gray (Stanford University) and D. Neuhoff (University of Michigan) "Quantization."

K. Immink (Philips, Eindhoven), P. Siegel (UC San Diego), J. Wolf (UC San Diego) "Codes for Digital Recorders."

T. Kailath (Stanford University) and V. Poor (Princeton University) "Detection of Stochastic Processes."

J. Korner (University of Rome) and A. Orlitsky (UC San Diego) "Zero-Error Information Theory."

S. Kulkarni (Princeton University), G. Lugosi (University Pompeu Fabra, Barcelona), S.V. Venkatesh, (University of Pennsylvania) "Learning and Pattern Recognition."

A. Lapidoth (MIT), P. Narayan (University of Maryland) "Reliable communication under Channel Uncertainty."

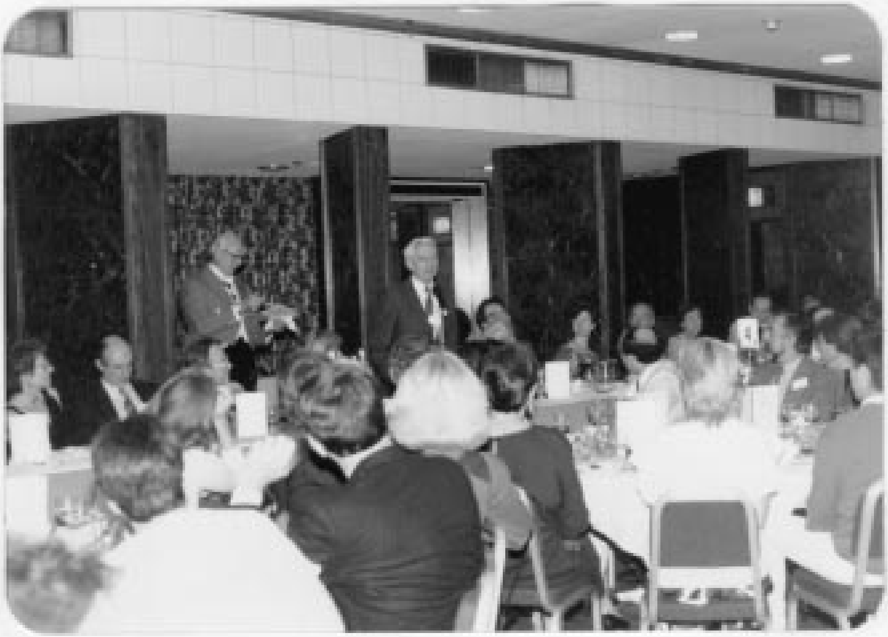
J. O'Sullivan (Washington University), R. Blahut (University of Illinois), D. Snyder (Washington) "Information Theoretic Methods Image Formation."

P. Shields (University of Toledo) "The Interactions Between Ergodic Theory and Information Theory."

A. Wyner (Bell Labs, Murray Hill), J. Ziv (Technion) and A. Wyner (University of Cal) "On the Role of Pattern Matching in Information Theory."

S. Verdu (Princeton University) "Five Decades of Shannon Theory."

# THE FOUNDER



## CLAUDE SHANNON
## BRIGHTON 1985

Courtesy of Toby Berger

# The 1998 International Symposium on Information Theory

The "main event" in the activities of our Society (other than the publication of the Transactions) is, of course, the Symposium. As this issue is in the process of reaching you, the 1998 Symposium is taking place in Cambridge, MA (August 16-21). A detailed report on this unique event will be reported in the next regular issue of the Newsletter. Here we simply pay tribute to it as the "capstone" event in the series of commemorative activities of the Society during this golden jubilee year.

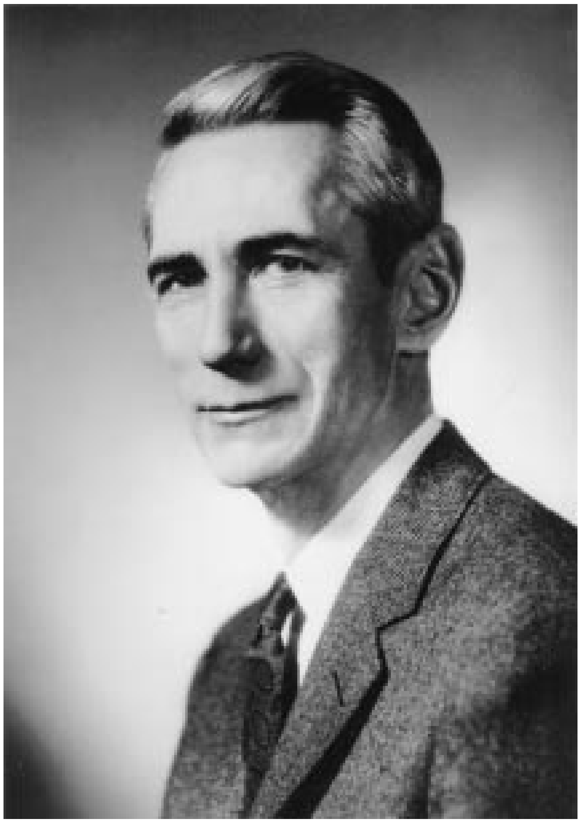


Photo courtesy of Jim Massey (given to him by the late Prof. Fritz Borgnis of the ETH Zurich

Another portrait of Claude Shannon from the mid '50's.

IEEE
445 Hoes Lane, P.O. Box 1331
Piscataway, NJ 08855-1331 USA

Information Theory
Society Newsletter