IEEE

IEEE Information Theory Society Newsletter

Vol. 53, No. 1, March 2003 | Editor: Lance C. Pérez | ISSN 1059-2362

LIVING INFORMATION THEORY

The 2002 Shannon Lecture

by Toby Berger
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853

1 Meanings of the Title

The title, "Living Information Theory," is a triple entendre. First and foremost, it pertains to the information theory of living systems. Second, it symbolizes the fact that our research community has been living information theory for more than five decades, enthralled with the beauty of the subject and intrigued by its many areas of application and potential application. Lastly, it is intended to connote that information theory is decidedly alive, despite sporadic protestations to the contrary. Moreover, there is a thread that ties together all three of these meanings for me. That thread is my strong belief that one way in which information theorists, both new and seasoned, can assure that their subject will remain vitally alive deep into the future is to embrace enthusiastically its applications to the life sciences.

2 Early History of Information Theory in Biology

In the 1950's and early 1960's a cadre of scientists and engineers were adherents of the premise that information theory could serve as a calculus for living systems. That is, they believed information theory could be used to build a solid mathematical foundation for biology which always had occupied a peculiar middle ground between the hard and the soft sciences. International meetings were organized by Colin Cherry and others to explore this frontier, but by the mid-1960's the effort had dissipated. This may have been due in part to none other than Claude Shannon himself, who in his guest editorial, *The Bandwagon*, in the March 1956 issue of the *IRE Transactions on Information Theory* stated:

> Information theory has ... perhaps ballooned to an importance beyond its actual accomplishments. Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in ... biology, psychology, linguistics, fundamental physics, economics, the theory of the organization, ... Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger. While we feel that information theory is indeed a valuable tool in providing fundamental insights into the nature of communication problems and will continue to grow in importance, it is certainly no panacea for the communication engineer or, *a fortiori*, for anyone else. Seldom do more than a few of nature's secrets give way at one time.

More devastating was Peter Elias's scathing 1958 editorial in the same journal, *Two Famous Papers*, which in part read:

> The first paper has the generic title *Information Theory, Photosynthesis and Religion...* written by an engineer or physicist ... I suggest we stop writing [it], and release a supply of man power to work on ... important problems which need investigation.

The demise of the nascent community that was endeavoring to inject information theory into mainstream biology probably was occasioned less by these "purist"

continued on page 6

From the Editor

Lance C. Pérez

In this issue of the *IEEE Information Theory Society Newsletter* we are fortunate to feature an article entitled "Living Information Theory" by Toby Berger, the 2001 Shannon Award Winner. This is a fascinating and provocative article that expands on Toby's well received Shannon lecture at the 2002 International Symposium on Information Theory in Lausanne, Switzerland.

The Information Theory Society news continues to be bittersweet. Another IT sage, Valery N. Koshelev, has passed away and is rememebered in this issue. At the same time, Michael B. Pursley, a past president of the IT society, has been honored with the IEEE Communications Society Edwin Howard Armstrong Achievement Award.

This issue of the Newsletter also marks the first appearing under the new IT Society President Han Vinck. Han's first President's Column appears on page 4.

Please help make the Newsletter as interesting and informative as possible by offering suggestions and contributing news. The deadlines for the 2003 issues of the newsletter are as follows:

Lance C. Pérez

Issue	Deadline
June 2003	April 22, 2003
September 2003	July 15, 2003
December 2003	October 15, 2003

Electronic submission, especially in ascii and Word formats, is encouraged. I may be reached at the following address:

Lance C. Pérez
Department of Electrical Engineering
209N Walter Scott Engineering Center
University of Nebraska-Lincoln
Lincoln, NE 68588-0511
Phone: (402)472-6258
Fax: (402)472-4732
Email: lperez@unl.edu

Sincerely,
Lance C. Pérez

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: 3 Park Avenue, 17th Floor, New York, NY 10016-5997.

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.

Table of Contents

Prague Floods and Information Theory: Urgent Call for Help

The recent floods in the Czech Republic have had serious consequences for our Czech colleagues. The most important library for the mathematical sciences in the country - situated in the Karlin district at the Faculty of Mathematics and Physics of Charles University - was heavily damaged and around 400 journals and 10-50,000 book titles lost. Among the losses are Acta Informatica, Journal of Computing and Information Technology, Problems of Control and Information Theory, IEEE Transactions on Information Theory, IEEE Transactions on Neural Networks, IEEE Transactions on Signal Processing and many more journals in our or in neighbouring areas. Also, several books written by members of our society have been lost. For further details, see http://www.mff.cuni.cz/povoden/ where you can also see lists of lost books and journals and get information on how to donate money or send books and/or journals.

The undersigned participated in EMS2002, the 24th European Meeting of Statisticians, held jointly with the 14th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. The conference was held in Prague immediately after the flood and participants witnessed some of the damage done.

At the personal level, what affected us most was the news that Igor Vajda has lost so much, if not everything of his private property. Igor Vajda is a key member of our Society, Fellow of IEEE, and a main contributor to the Information Theory/Statistics literature. He was also a key member of the EMS2002 Scientific Committee. However, due to the dramatic events, he could not attend the EMS2002 himself.

Igor and his wife, Zdenka, have lived for many years in what was a charming small house surrounded by fruit trees along the Vltava river in Dolanky north of Prague. The flood waters reached the roof of their house destroying thereby all of the interior and did serious damage to the house itself and also devastated their garden. Essentially everything they had built up over a lifetime was destroyed. Igor and Zdenka are now starting to reconstruct their lives. It will take several months before they will be able to return to the house, since serious reconstruction of it needs to be done. Igor can hardly resume his scientific work at UTIA, the Institute of Information Theory and Automation in Prague, where he holds a senior position.

The website http://siprint.utia.cas.cz/igor/ contains information on the losses suffered by Igor. Let us add that during the first critical hours, Igor and Zdenka — with several friends and neighbours — tried to save the most important things, but the water flooded the places assumed to be secure in even the most pessimistic predictions. Thereafter, they could only look on powerlessly as their property vanished under water. For illustration, the last 100-year flood — in 1890 — achieved a flow of 3970 m^3/sec as compared to 5300 m^3/sec on August 14, 2002. The photographs available at the website help to convey the true extent of the catastrophe.

The website lists a bank account number, set-up by Igor's friends and former students in Prague, to which any donation can be made. As Igor's insurance contracts and the common support of the state appear to be entirely insufficient to cover such a disaster, any financial support would be of great moral and practical value for Igor and his wife. The website also contains a list of damaged and lost books, journals and papers collected by Igor over the years. Some of them are indispensable for Igor's work in the immediate future. Any copy of the listed titles would be very helpful for him. The list will be updated continuously.

During EMS2002 a sum was collected both in support of the library at Charles University and in support of Igor, personally. We thought that friends of Czech mathematics and mathematicians — and of Igor in particular — who were not present at EMS2002 should also learn about the impact of the flood. They too may want to be given an opportunity to offer their help.

Edward van der Meulen
Katholieke Universiteit Leuven, Belgium

Flemming Topsøe
University of Copenhagen, Denmark

President's Message

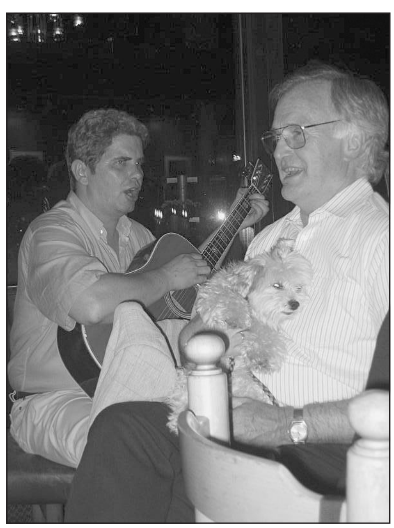
Han Vinck

In my inaugural message I first want to thank all the volunteers for the work they have done and continue to do for our society. A special thanks goes to Tom Fuja, the past president. Without his understanding of finances, the functioning of IEEE and his way of handling complicated problems, it would have been impossible to have order in the chaos of the past year.

A subcommittee of the Board of Governors is examining our "bad" financial situation and will present suggestions to balance the 2004 budget at the next Board meeting. It is clear that we have to restructure our income policy. IEEE charges and politics are unpredictable and therefore, we have to find a solid basis for our finances. The IT membership costs can be split up into the following parts:

Transactions	paper version (includes mailing $ 20)	$ 40
	production	$ 60
	related activities. Newsletter, Xplore	$ 12
Society costs	(BoG)	$ 8
IEEE charges (including recovery) for the year 2003		$ 40
Total costs		$ 160

You may wonder why your membership fee for the year 2002, that includes a paper version of the transactions, is only $30. Actually, one may argue that we give our product (the Transactions), which is created by IT volunteers, to IEEE and allow them to make money with it. In the past this was very successful, and money was no problem. However, we have seen IEEE lose a lot of their assets in recent years, and the "fat" they need to survive is present in the societies. The discussion has thus reduced to the ownership of our reserves. We are forced by IEEE to balance the budget for the year 2004 with additional IEEE recovery charges. We are very lucky to have the expertise of Marc Fossorier as treasurer. To reduce costs we may split the membership into an electronic and a paper version. This would take away the $40 load from our budget for delivering a paper version of the transactions to your desk. We need new initiatives to increase income. One idea is to produce a DVD with all our transactions, as is done by the Communication Society. IEEE wants us to increase the income from conferences. We need to discuss this at our next Board meetings. It is our policy to encourage and enable students and young faculty to participate at low cost. We also have a large percentage of our members from universities and developing countries with limited travel budgets. I will report on this item in the next newsletter. Further information can also be found in the minutes of the Board meetings, prepared by Aaron Gulliver.



Jim Massey and Martin Bossert singing a duet at the Breisach workshop.

It is a pleasure to be a member of this society, where personal communication still plays an important role. For me, a highlight for the year 2002 was the meeting in Breisach on "Concepts in Information Theory, a tribute to Jim Massey". It showed that the "old" spirit is still alive! The main goal of the meeting was to stimulate discussions on the principles of Information Theory. We need more of this type of meeting. Other highlights of the last year were ISIT2002 in Lausanne and the ITW in Bangelore. The Bangalore workshop had a large number (115) of local participants, illustrating the intention of our society to be as international and as global as possible.

A further illustration of this fact is that we now have 3 presidents (past, present and future) on the Board from outside USA. Joachim Hagenauer and Han Vinck from Region 8 and Hideki Imai from Region 10. We want to bring the benefits of society membership to as many individual engineers and scientists as possible. These benefits include reduced conference fees, publications, electronic library and other services offered by the IT Society. We have a program called the Membership Fee Subsidy Program (MFSP), which will both complement the IEEE Minimum Income offering and provide a significant additional benefit for qualified individuals. With the IT Membership Fee Subsidy Program, the IT society will pay 50% of the IEEE and IT dues for individuals qualifying for the Minimum Income option, up to a maximum of 20 individuals (IEEE pays the other 50%). Those society members receiving the IT-FSP benefits are obliged to promote IT activities. For any questions concerning the program, please contact Steven McLaughlin, the newly elected second vice-president and IT membership development chair. Affiliate membership in the Information Theory Society is available to engineers and scientists not interested to be an IEEE member. For further information, see our new home page (www.ieeeits.org) maintained by Aaron Gulliver.

From left to right: Han Vinck, Daniel Elias, Ellen Elias-Bursac and Vijay Bhargava

An important activity for the future of our society is the foundation of chapters. Chapters can be used to coordinate local activities and to present information theory to the local community. Personally, I participate in two such initiatives: the working community on Information Theory in the Benelux and the Germany chapter on Information Theory. They were very successful in organizing an IT workshop (1990, Veldhoven) and ISIT1997 (Ulm), but many other events took place on a regular basis. The benefits for chapters are numerous, but one has to be creative. To start with, a new chapter is supported by the society with a startup grant of $1000. For obvious reasons, IEEE Regions and IEEE Sections encourage the activities of chapters and it is easy to get support from these organizations too. Another advantage that is not often used is the Disinguished Lecturers program (see our Web site). Coordination of activities in Information Theory makes our field stronger and visible to the "outside" world. Writing a report of the chapter's activity for our newsletter is very much welcomed. Thanks to Lance Pérez, the newsletter is in excellent shape now. To improve contacts between the chapters, a lunch for the chapter chairpersons is organized at each ISIT. For a detailed report about this meeting in Lausanne, please see our December newsletter. We further encourage chapters with the "Best Chapter Award". This award, which includes $1000, is given to the chapter that was the most active during the previous year. The Japan chapter won this award in 2002. If you pay a visit to their web site, you can see why!

I look forward to the ISIT2003 in Yokohama, June 29 - July 4. Japan has a very active and strong community in Information Theory. They organize the SITA conference every year in December (about 400 mainly Japanese participants), and the International Symposium ISITA every two years (2002 in Xian, China). It is a great pleasure to have Hideki Imai as first vice-president of our Society and Ryuji Kohno on the Board of Governors. It reflects the Asian strength and representation from Region 10. A stronger input from outside USA is necessary in other society activities to reflect the composition of our membership. Just before and after the ISIT2003 there are several workshops that might be of interest. There is the Workshop on Coding, Cryptography and Combinatorics in Huang Shan (Yellow Mountain) City, China, June 23-28, webpage: http://www.ustc.edu.cn/conference/ccc, the IT Workshop in Hong Kong July 6-10, organized by Victor Wei and Raymond W. Yeung (http://personal.ie.cuhk.edu.hk/~info-itw/CFP.pdf), and the 3rd Asia-Europe Workshop on Coding and Information Theory, June 25-27, in Kamogawa, China Prefecture, organized by Kingo Kobayashi, http://main.math-sys.is.uec.ac.jp/ae3/announce.html.

Incoming presidents of our society are encouraged to "experience" the IEEE TAB meetings. I participated in the IEEE Board of Directors meeting in Toronto, June 2002. At this occasion, two children of Peter Elias, Daniel Elias and Ellen Elias-Bursac, were present at the IEEE Award Ceremony to receive the Richard W. Hamming Medal, awarded for exceptional contributions to information sciences and systems, in the name of their father Peter. It was very impressive to meet them and get some background information about this great scientist. At the same meeting, Vijay Bhargava received the IEEE Graduate Teaching Award and Muriel Medard the IEEE Leon K. Kirchmayer Prize Paper Award. Congratulations also to T.J. Richardson, R.L. Urbanke, M.G. Luby, M. Mitzenmacher, M.A. Shokrollahi, and D.A. Spielman for the 2002 IEEE Information Theory Society Paper Award and M.L. Honig and W. Xiao for the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award.

The year 2003 promises to be an exciting year from a scientific and a society point of view. I look forward to serve as your president and hope to meet many of you at our society activities.

LIVING INFORMATION THEORY

The 2002 Shannon Lecture

Continued from page 1

information theory editorials than by the relatively primitive state of quantitative biology at the time.

1. The structure of DNA was not determined by Crick and Watson until five years after Shannon published *A Mathematical Theory of Communication*.
2. Although some measurements had been made of neural spikes on single axons of dissected animals (most notably on the giant squid axon), it was not possible as it is today to record accurately and simultaneously in vivo the pulse trains of many neighboring neurons.
3. It was not possible to measure time variations in the concentrations of chemicals at sub-millisecond speeds in volumes of submicron dimensions such as those which constitute ion channels in neurons. This remains a stumbling block, but measurement techniques capitalizing on fluorescence and other phenomena are steadily progressing toward this goal.

We offer arguments below to support the premise that matters have progressed to a stage at which biology is positioned to profit meaningfully from an invasion by information theorists. Indeed, during the past decade some biologists have equipped themselves with more than a surface knowledge of information theory and are applying it correctly and fruitfully to selected biological subdisciplines, notable among which are genomics and neuroscience. Since our interest here is in the information theory of sensory perception, we will discuss neuroscience and eschew genomics.

3 Information Within Organisms

At a fundamental level information in a living organism is instantiated in the time variations of the concentrations of chemical and electrochemical species (ions, molecules and compounds) in the compartments that comprise the organism. Chemical thermodynamics and statistical mechanics tell us that these concentrations are always tending toward a multiphase equilibrium characterized by minimization of the Helmholtz free energy functional. On the other hand, complete equilibrium with the environment never is attained both because the environment constantly changes and because the organism must exhibit homeostasis in order to remain "alive". A fascinating dynamic prevails in which the organism sacrifices internal energy in order to reduce its uncertainty about the environment, which in turn permits it to locate new sources of energy and find mates with whom to perpetuate the species. This is one of several considerations that strongly militate in favor of looking at an information gain by a living system never in absolute terms but rather always relative to the energy expended to achieve it.

There is, in addition, an intriguing mathematical analogy between the equations that govern multiphase equilibrium in chemical thermodynamics and those which specify points on Shannon's rate-distortion function of an information source with respect to a fidelity criterion [9]. This analogy is not in this writer's opinion just a mathematical curiosity but rather is central to fruitfully "bringing information theory to life." We shall not be exploring this analogy further here, however. This is because, although it provides an overarching theoretical framework, it operates on a level which does not readily lead to concrete results apropos our goal of developing an information-theoretically based formulation of sensory perception.

An information theorist venturing into new territory must treat that territory with respect. In particular, one should not assume that, just because the basic concepts and methods developed by Shannon and his disciples have proved so effective in describing the key features of man-made communication systems, they can be applied en masse to render explicable the long-standing mysteries of another discipline. Rather, one must think critically about information-theoretic concepts and methods and then apply only those that genuinely transfer to the new territory. My endeavors in this connection to date have led me to the following two beliefs:

- Judicious application of Shannon's fundamental concepts of entropy, mutual information, channel capacity and rate-distortion is crucial to gaining an elevated understanding of how living systems handle sensory information.
- Living systems have little if any need for the elegant block and convolutional coding theorems and techniques of information theory because, as will be explained below, organisms have found ways to perform their information handling tasks in an effectively Shannon-optimum manner without having to employ coding in the information-theoretic sense of the term.

Is it necessary to learn chemistry, biochemistry, biophysics, neuroscience, and such before one can make any useful contributions? The answer, I feel, is "Yes, but not deeply." The object is not to get to the point where you can think like a biologist. The object is to get to the point where you can think like the biology. The biology has had hundreds of millions of years to evolve via natural selection such that most of that which it does is done in a nearly optimum fashion. Hence, thinking about how the biology **should** do things is often effectively identical to thinking about how the biology **does** do things and is perhaps even a more fruitful endeavor.[1]

Information theorists are fond of figuring out how best to transmit information over a "given" channel. When trespassing on biological turf, however, an information theorist must abandon the tenet that the channel is given. Quite to the contrary, nature has evolved the channels that function within organisms in response to needs for specific information residing either in the environment or in the organism itself - channels for sight, channels for sound, for olfaction, for touch, for taste, for blood alcohol and osmolality regulation, and so on. Common sense strongly suggests that biological structures built to sense and transfer information from certain sites located either outside or inside the organism to other such sites will be efficiently "matched" to the data sources they service. Indeed, it would be ill-advised to expect otherwise, since natural selection rarely chooses foolishly, especially in the long run. The compelling hypothesis, at least from my perspective, is that all biological channels are well matched to the information sources that feed them.

4 Double Matching of Sources and Channels

Matching a channel to a source has a precise mathematical meaning in information theory. Let us consider the simplest case of a discrete memoryless source (dms) with instantaneous letter probabilities $\left\{p(u), u\in U\right\}$ and a discrete memoryless channel (dmc) with instantaneous transition probabilities $\left\{p\left(y|x\right), x\in X, \in Y\right\}$. Furthermore, let us suppose that the channel's purpose is to deliver a signal $\{Y_k\}$ to its output terminal on the basis of which one could construct an approximation $\{V_k\}$ to the source data $\{U_k\}$ that is accurate enough for satisfactory performance in some application of interest. Following Shannon, we shall measure said accuracy by means of a distortion measure $d: U\times V\rightarrow[0,\infty]$. $\{V_k\}$ will be considered to be a suffciently accurate approximation of $\{U_k\}$ if and only if the average distortion does not exceed a level deemed to be tolerable which we shall denote by D. Stated mathematically, our requirement for an approximation to be sufficiently accurate is

$$\lim_{n\rightarrow\infty} En^{-1}\sum_{k=1}^{n} d\left(U_k, V_k\right)\leq D.$$

In order for the dmc $\left\{p\left(y|x\right)\right\}$ to be *instantaneously matched* to the combination of the dms $\{p(u)\}$ and the distortion measure $\{d(u,v)\}$ at fidelity D, the following requirements must be satisfied:

1. The number of source letters produced per second must equal the number of times per second that the channel is available for use.
2. There must exist two transition probability matrices $\{r(x|u), u\in U, x\in X\}$ and $\{w(v\mid y), y\in Y, v\in V\}$, such that the end-to-end transition probabilities

$$q\left(v|u\right):=\sum_{x\in X}\sum_{y\in Y} r\left(x|u\right)p\left(y|x\right)w\left(v|y\right), (u,v)\in U\times V$$

solve the variational problem that defines the point $(D,R(D))$ on Shannon's rate-distortion function of the dms $\{p(u)\}$ with respect to the distortion measure $\{d(u,v)\}$.

Readers not conversant with rate-distortion theory should refer to Section 10 below. If that does not suffice, they should commune at their leisure with Shannon [4], Jelinek [10], Gallager [11] or Berger [9]. However, the two key examples that follow should be largely accessible to persons unfamiliar with the content of any of these references. Each example is constructed on a foundation comprised of two of Shannon's famous formulas. Moreover, each exhibits not only matching of the channel to the source but also matching of the source to the channel. This phenomenon of *double matching* is central to the approach we take in the remainder of this paper to the generation of an information theory of living systems.

4.1 Double Matching Example 1

This example uses: (1) the formula for the capacity of a binary symmetric channel (BSC) with crossover probability α, namely

$$C = 1-h(\alpha) = 1+\alpha\log_2\alpha + (1-\alpha)\log_2(1-\alpha) \text{ bits/channel use,}$$

where we assume without loss of essential generality that $\alpha \leq 1/2$, and (2) the formula for the rate-distortion function of a Bernoulli-1/2 source with respect to the error frequency distortion measure $d(x, y) = 1 - \delta(x, y)$, namely

$$R(D) = 1 - h(D) = 1+D\log_2 D + (1 - D)\log_2(1 - D)$$
$$\text{bits/source letter, } 0\leq D\leq 1/2.$$

Shannon's converse channel coding theorem [1] establishes that it is not possible to convey more than nC bits of information from the channel's input to its output via n uses of the channel. Similarly, his converse source coding theorem [4] establishes that it is not possible to generate an approximation $V_1, \ldots, V_n$ to source letters $U_1, \ldots, U_n$ that has an average distortion $En^{-1}\sum_{k=1}^{n} d(U_k, V_k)$ of D or less unless that representation is based on $nR(D)$ or more bits of information about these source letters. Accordingly, assuming the source resides at the channel input, it is impossible to generate an approximation to it at the channel output that has an average distortion any smaller than the value of D for which $R(D) = C$, even if the number n of source letters and channel uses is allowed to become large. Comparing the above formulas for C and $R(D)$, we see that no value of average distortion less than α can be achieved. This is true regardless of how complicated an encoder we place between the source and the channel, how complicated a decoder we place between the channel and the recipient of the source approximation, and how large a finite delay we allow the system to employ. It is easy to see, however, that $D = \alpha$ can be achieved simply by

[1]Elwyn Berlekamp related at IEEE ISIT 2001 in Washington, DC, a conversation he had with Claude Shannon in an MIT hallway in the 1960's the gist of which was:
CES: Where are you going, Elwyn?
EB: To the library to study articles, including some of yours.
CES: Oh, don't do that. You'd be better off to just figure it out for yourself.

connecting the source directly to the channel input and using the channel output as the approximate reconstruction of the source output. Hence, this trivial communication system, which is devoid of any source or channel coding and operates with zero delay, is optimum in this example. There are two reasons for this:

Reason One: The channel is instantaneously matched to the source as defined above with the particularly simple structure that $X = U$, $V = Y$, $r(x \mid u) = \delta(u, x)$ and $w(v \mid y) = \delta(y, v)$. That is, the source is instantaneously and deterministically fed into the channel, and the channel output directly serves as the approximation to the source.

Reason Two: The source also is matched to the channel in the sense that the distribution of each U_k, and hence of each X_k, is $p(0) = p(1) = 1/2$, which distribution maximizes the mutual information between a channel input and the corresponding output. That is, the channel input letters are i.i.d. with their common distribution being the one that solves the variational problem that defines the channel's capacity.

4.2 Double Matching Example 2

The channel in this example is a time-discrete, average-power-constrained additive white Gaussian noise (AWGN). Specifically, its k^{th} output Y_k equals $X_k + N_k$, where X_k is the k^{th} input and the additive noises N_k are i.i.d. $N(0,N)$ for $k = 1, 2, \ldots$. Also, the average signaling power cannot exceed S, which we express mathematically by the requirement

$$\lim_{n \to \infty} E n^{-1} \sum_{k=1}^{n} X_k^2 \leq S.$$

Shannon's well-known formula for this channel's capacity is

$$C = \frac{1}{2} \log_2 \left(1 + \frac{S}{N}\right) \text{ bits/channel use.}$$

The source in this example produces i.i.d. $N(0, \sigma^2)$ symbols $\{U_k\}$. The squared error distortion measure, $d(u, v) = (v - u)^2$, is employed, so the end-to-end distortion is the mean-squared-error,

$$\text{MSE} = \lim_{n \to \infty} E n^{-1} \sum_{k} \left(V_k - U_k\right).$$

Shannon's celebrated formula for the rate-distortion function of this source and distortion measure combination is

$$R(D) = (1/2) \log_2(\sigma^2/D),\ 0 \leq D \leq \sigma^2.$$

The minimum achievable value of the MSE is, as usual, the value of D that satisfies $R(D) = C$, which in this example is

$$D = \sigma^2 \Big/ \left(1 + \frac{S}{N}\right).$$

As in Example 1, we find that this minimum value of D is trivially attainable without any source or channel coding and with zero delay. However, in this instance the source symbols must be scaled by $\lambda := \sqrt{S} / \sigma$ before being put into the channel in order to ensure compliance with the power constraint. Similarly, V_k is produced by multiplying Y_k by the constant $\mu := \sqrt{S\sigma / (S + N)}$, since this produces the minimum MSE estimate of U_k based on the channel output. Hence, the channel is instantaneously matched to the source via the deterministic transformations $r(x \mid u) = \delta(x - \lambda u)$ and $w(v \mid y) = \delta(v - \mu y)$. Moreover, the source is matched to the channel in that, once scaled by λ, it becomes the channel input which, among all those that comply with the power constraint, maximizes mutual information between itself and the channel output that it elicits. Thus, the scaled source drives the constrained channel at its capacity.

It can be argued validly that, notwithstanding the fact that Examples 1 and 2 deal with source models, channel models, and distortion measures all dear to information theorists, these examples are exceptional cases. Indeed, if one were to modify $\{p(u)\}$ or $\{p(y \mid x)\}$ or $\{d(u, v)\}$ even slightly, there no longer would be an absolutely optimum system that is both coding-free and delay-free. Achieving optimal performance would then require the use of coding schemes whose complexity and delay diverge as their end-to-end performance approaches the minimum possible average distortion attainable between the source and an approximation of it based on information delivered via the channel. However, if the perturbations to the source, to the channel and/or to the distortion measure were minor, then an instantaneous system would exist that is only mildly suboptimum. Because of its simplicity and relatively low operating costs, this mildly suboptimum scheme likely would be deemed preferable in practice to a highly complicated system that is truly optimum in the pure information theory sense.

I have argued above for why it is reasonable to expect biological channels to have evolved so as to be matched to the sources they monitor. I further believe that, as in Examples 1 and 2, the data selected from a biological source to be conveyed through a biological channel will drive that channel at a rate effectively equal to the its resource-constrained capacity. That is, I postulate that double matching of channel to source and of source to channel in a manner analogous to that of Examples 1 and 2 is the rule rather than the exception in the information theory of living systems. Indeed, suppose selected stimuli were to be conditioned for transmission across one of an organism's internal channels in such a way that information failed to be conveyed at a rate nearly equal to the channel's capacity calculated for the level of resources being expended. This would make it possible to select additional data and then properly condition and transmit both it and the original data through the channel in a manner that does not increase the resources consumed. To fail to use such an alternative input would be wasteful either of information or of energy, since energy usually is the constrained resource in question. As explained previously, a fundamental characteristic of an efficient organism is that it always should be optimally trading information for energy, or vice versa, as circumstances dictate. The only way to assure that pertinent information will be garnered at low latency at the maximum rate per unit of power expended is not only to match the channel to the source but also to match the source to the channel.

5 Bit Rate and Thermodynamic Efficiency

We shall now discuss how increasing the number of bits handled per second unavoidably increases the number of joules expended per bit (i.e., decreases thermodynamic efficiency). To establish this in full generality requires penetrating deeply into thermodynamics and statistical mechanics. We shall instead content ourselves with studying the energy-information tradeoff implicit in Shannon's celebrated formula for the capacity of an average-power constrained bandlimited AWGN channel, namely

$$C(S) = W \log\left(1 + \frac{S}{N_0 W}\right),$$

where S is the constrained signaling power, W is the bandwidth in positive frequencies, and N_0 is the one-sided power spectral density of the additive white Gaussian noise. Like all capacity-cost functions, $C(S)$ is concave in S. Hence, its slope decreases as S increases; specifically, $C'(S) = W/(S + N_0W)$. The slope of $C(S)$ has the dimensions of capacity per unit of power, which is to say (bits/second)/(joules/second) = bits/joule. Since the thermodynamic efficiency of the information-energy tradeoff is measured in bits/joule, it decreases steadily as the power level S and the bit rate $C(S)$ increase. This militates in favor of gathering information slowly in any application not characterized by a stringent latency demand. To be sure, there are circumstances in which an organism needs to gather and process information rapidly and therefore does so. However, energy conservation dictates that information handling always should be conducted at as leisurely a pace as the application will tolerate. For example, recent experiments have shown that within the neocortex a neural region sometimes transfers information at a high rate and accordingly expends energy liberally, while at other times it conveys information at a relatively low rate and thereby expends less than proportionately much energy. In both of these modes, and others in between, our hypothesis is that these coalitions of neurons operate in an information-theoretically optimum manner. We shall attempt to describe below how this is accomplished.

6 Feedforward and Feedback: Bottom-Up and Top-Down

Before turning in earnest to information handling by neural regions, we first need to generalize and further explicate the phenomenon of double matching of sources and channels. So far, we have discussed this only in the context of sources and channels that are memoryless. We could extend to sources and/or channels with memory via the usual procedure of blocking successive symbols into a "supersymbol" and treating long supersymbols as nearly i.i.d., but this would increase the latency by a factor equal to the number of symbols per supersymbol, thereby defeating one of the principal advantages of double matching. We suggest an alternative approach below which leads to limiting the memory of many crucial processes to at most first-order Markovianness.

It has long been appreciated that neuromuscular systems and metabolic regulatory mechanisms exhibit masterful use of feedback. Physiological measurements of the past fifteen or so years have incontrovertibly established that the same is true of neurosensory systems. Describing signaling paths in the primate visual cortex, for example, Woods and Krantz [8] tell us that "In addition to all the connections from V1 and V2 to V3, V4 and V5, each of these regions connects back to V1 and V2. These seemingly backward or reentrant connections are not well understood. · · · Information, instead of flowing in one direction,now flows in both directions. Thus, later levels do not simply receive information and send it forward, but are in an intimate two-way communication with other modules." Of course, it is not that information flowed unidirectionally in the visual system until some time in the 1980's and then began to flow bidirectionally. Rather, as is so often the case in science, measurements made possible by new instrumentation and methodologies have demanded that certain cherished paradigms be seriously revised. In this case, those mistaken paradigms espoused so-called "bottom-up" unidirectionality of signaling pathways in the human visual system (HVS) [6] [5].

Instead of speaking about feedforward and feedback signaling, neuroscientists refer to *bottom-up* and *top-down* signaling, respectively. Roughly speaking, neurons whose axons carry signals principally in a direction that moves from the sensory organs toward the "top brain" are called bottom-up neurons, while those whose axons propagate signals from the top brain back toward the sensory organs are called top-down neurons. Recent measurements have revealed that there are roughly as many top-down neurons in the HVS as there are bottom-up neurons. Indeed, nested feedback loops operate at the local, regional and global levels.[2] We shall see that a theory of sensory perception which embraces rather than eschews feedback reaps rewards in the form of analytical results that are both simpler to obtain and more powerful in scope.

7 Neurons and Coalitions

The roughly 10^{10} neurons in the human visual system (HVS) constitute circa one-tenth of all the neurons in the brain. HVS neurons are directly interconnected with one another via an average of 10^4 synapses per neuron. That is, a typical HVS neuron has on its dendritic tree about 10^4 synapses at each of which it taps off the spike signal propagating along the axon of one of the roughly 10^4 other HVS neurons that are afferent (i.e., incoming) to it. Via processing of this multidimensional input in a manner to be discussed below, it generates an ef-

[2]Shannon's topic for the inaugural Shannon Lecture in June 1973 was *Feedback*. Biological considerations, one of his many interests [3], may have in part motivated this choice.

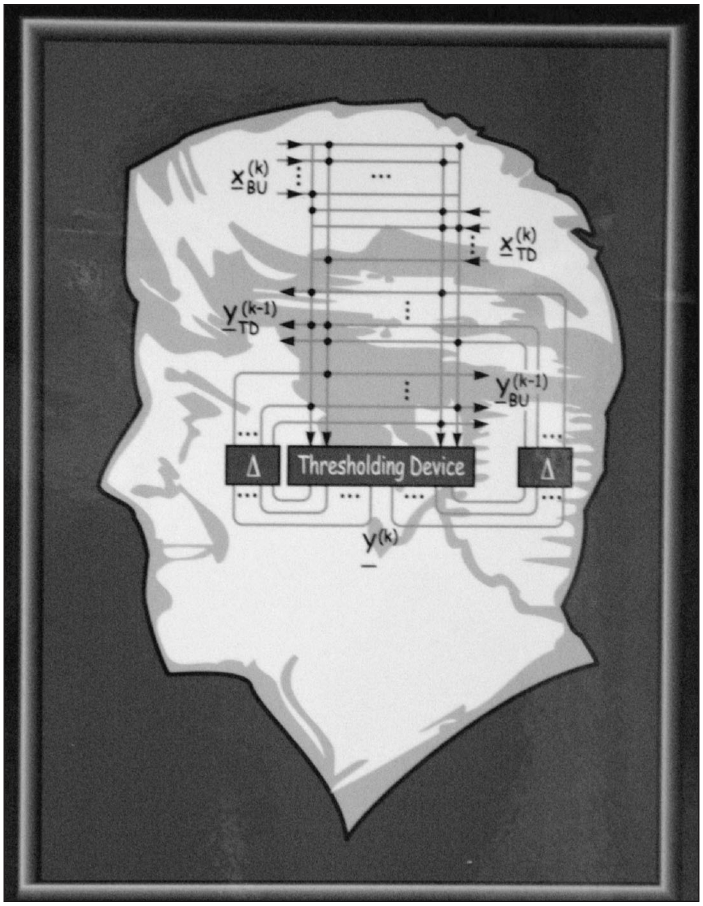


Figure 1: Time-Discrete Model of Neural Coalition.

ferent (i.e., outgoing) spike train on its own axon which propagates to the circa 10^4 other HVS neurons with which it is in direct connection. The $10^{10} \times 10^{10}$ matrix whose (i, j) entry is 1 if neuron i is afferent to neuron j and 0 otherwise thus has a 1's density of 10^{-6}. However, there exist special subsets of the HVS neurons that are connected much more densely than this. These special subsets, among which are the ones referred to as V1, V2 . . . V5 in the above quote, consist of a few million to as many as a few tens of millions of neurons and have connectivity submatrices whose 1's densities range from 0.1 to as much as 0.5. Common sense suggests and experiments have verified that the neurons comprising such a subset work together to effect one or more crucial functions in the processing of visual signals. We shall henceforth refer to such subsets of neurons as "coalitions". Alternative names for them include neural "regions", "groupings" and "contingents".

Figure 1 shows a schematic representation of a neural coalition. Whereas real neural spike trains occur in continuous time and are asynchronous, Figure 1 is a time-discrete model. Its time step is circa 2.5 ms, which corresponds to the minimal duration between successive instants at which a neuron can generate a spike; spikes are also known as an *action potentials*. The time-discrete model usually captures the essence of the coalition's operation as regards information transfer. Any spike traveling along an axon afferent to a coalition of neurons in the visual cortex will reach all the members of that coalition within the same time step. That is, although the leading edge of the spike arrives serially at the synapses to which it is afferent, the result is as it were multicasted to them all during a single time slot of the time-discrete model.[3]

The input to the coalition in Figure 1 at time k is a random binary vector $\underline{X}(k)$ which possesses millions of components. Its i^{th} component, $X_i(k)$ is 1 if a spike arrives on the i^{th} axon afferent to the coalition during the k^{th} time step and is 0 otherwise.[4] The afferent neurons in Figure 1 have been divided into two groups indexed by BU and TD, standing respectively for bottom-up and top-down. The vertical lines represent the neurons of the coalition. The presence (absence) of dark dot where the i^{th} horizontal line and m^{th} vertical line cross indicates that the i^{th} afferent axon forms (does not form) a synapse with the m^{th} neuron of the coalition. The strength, or weight, of synapse (i, m) will be denoted by W_{im}; if afferent axon i does not form a synapse with coalition neuron m, then $W_{im} = 0$. If $W_{im} > 0$, the connection (i, m) is said to be *excitatory*; if $W_{im} < 0$, the connection is *inhibitory*. In a primate visual cortex about five-sixths of the connections are excitatory.

The so-called post-synaptic potential (PSP) of neuron m is built up during time step k as a weighted linear combination of all the signals that arrive at its synapses during this time step. If this sum exceeds the threshold $T_m(k)$ of neuron m at time k, then neuron m produces a spike at time k and we write $Y_m(k) = 1$; if not, then $Y_m(k) = 0$. The thresholds do not vary much with m and k with one exception. If $Y_m(k) = 1$, a refractory period of duration about equal to a typical spike width follows during which the threshold is extremely high, making it virtually impossible for the neuron to spike. In real neurons, the PSP is reset to its rest voltage after a neuron spikes. One shortcoming of our time-discrete model is that it assumes that a neuron's PSP is reset between the end of one time step and the beginning of the next even if the neuron did not fire a spike. In reality, if the peak PSP during the previous time step did not exceed threshold and hence no action potential was produced, contributions to this PSP will partially carry over into the next time step. Because of capacitative leakage they generally will have decayed to

[3] The time-discrete model mirrors reality with insufficient accuracy in certain saturated or near-saturated conditions characterized by many of the neurons in a coalition spiking almost as fast as they can. Such instances, which occur rarely in the visual system but relatively frequently in the auditory system, are characterized by successive spikes on an axon being separated by short, nearly uniform durations whose sample standard deviation is less than 1 millisecond. Mathematical methods based on Poisson limit theorems (a.k.a. mean field approximations) and PDE's can be used to distinguish and quantify these exceptional conditions [25] [26].

[4] Most neuroscientists today agree that the detailed shape of the action potential spike as a function of time is inconsequential for information transmission purposes, all that matters being whether there is or is not a spike in the time slot in question.

one-third or less of their peak value a time step ago, but they will not have vanished entirely.

Every time-discrete system diagram must include a unit delay element in order to allow time to advance. In Figure 1 unit delays occur in the boxes marked Δ. Note, therefore, that the random binary vector $\underline{Y}(k-1)$ of spikes and non-spikes produced by the coalition's neurons during time step k–1 gets fed back to the coalition's input during time step k. This reflects the high interconnection density of the neurons in question that is responsible for why they constitute a coalition. Also note that, after the spike trains on the axons of each neuron in the coalition are delivered to synapses on the dendrites of selected members of the coalition itself, they then proceed in either a top-down or a bottom-up direction toward other HVS coalitions. In the case of a densely connected coalition, about half of its efferent axons' connections are local ones with other neurons in the coalition. On average, about one quarter of its connections provide feedback, mostly to the coalition immediately below in the HVS hierarchy although some are directed farther down. The remaining quarter are fed forward to coalitions higher in the hierarchy, again mainly to the coalition directly above. This elucidates why we distinguished the external input $\underline{X}$ to the coalition as being comprised of both a bottom-up (BU) and a top-down (TD) subset; these subsets come, respectively, mainly from the coalitions directly below and directly above.

Neuroscientists refer not only to bottom-up and top-down connections but also to *horizontal* connections [15], [16], [17]. Translated into feedforward-feedback terminology, horizontal connections are local feedback such as $\underline{Y}(k-1)$ in Figure 1, top-down connections are regional feedback, and bottom-up connections are regional feedforward. Bottom-up and top-down signals also can be considered to constitute instances of what information theorists call *side information*.[5]

In information theory parlance, the neural coalition of Figure 1 is a time-discrete, finite state channel whose state is the previous channel output vector. At the channel input appear both the regional feedforward signal $\{\underline{X}_{BU}(k)\}$ and the regional feedback signal $\{\underline{X}_{TD}(k)\}$. However, there is no channel encoder in the information-theoretic sense of that term that is able to operate on these two signals in whatever manner suits its fancy in order to generate the channel input. Rather, the composite binary vector process $\{\underline{X}(k)\} := \{(\underline{X}_{BU}(k),\underline{X}_{TD}(k))\}$ simply enters the channel by virtue of the axons carrying it being afferent to the synapses of the channel's neurons. We shall subsequently see that there is no decoder in the information-theoretic sense either; as foretold, the system is coding-free.

My use of the adjective "coding-free" is likely to rile both information theorists and neuroscientists - information theorists because they are deeply enamored of coding and neuroscientists because they are accustomed to thinking about how an organism's neural spike trains serve as coded representations of aspects of its environment. In hopes of not losing both halves of my audience at once, allow me to elaborate. Certainly, sensory neurons' spike trains constitute an encoding of environmental data sources. However, unless they explicitly say otherwise, information theorists referring to coding usually mean channel coding rather than source coding. Channel coding consists of the intentional insertion of cleverly selected redundant parity check symbols into a data stream in order to provide error detection and error correction capabilities for applications involving noisy storage or transmission of data. I do not believe that the brain employs error control coding (ECC).[6]

8 Mathematical Model of a Neural Coalition

It's time for some mathematical information theory. Let's see what has to happen in order for an organism to make optimum use of the channel in Figure 1, i.e., to transmit information through it at a rate equal to its capacity. (Of course, we are not interested just in sending any old information through the channel - we want to send the "right" information through the channel, but we shall temporarily ignore this requirement.) A channel's capacity depends on the level of resources expended. What resources, if any, are being depleted in the course of operating the channel of Figure 1? The

[5] In [2] Shannon wrote, "Channels with feedback from the receiving to the transmitting point are a special case of a situation in which there is additional information available at the transmitter which may be used as an aid in the forward transmission system."

[6] There is a possibility that the brain employs a form of *space-time coding*, with the emphasis heavily on space as opposed to time. Here, the space dimension means the neurons themselves, the cardinality of which dwarfs that of the paucity of antennas that comprise the space dimension of the space-time codes currently under development for wireless communications. Think of it this way. In order to evolve more capable sensory systems, organisms needed to expand the temporal and/or the spatial dimensionality of their processing. Time expansion was not viable because the need to respond to certain stimuli in only a few tens of milliseconds precluded employing temporal ECC techniques of any power because these require long block lengths or long convolutional constraint lengths which impose unacceptably long latency. Pulse widths conceivably could have been narrowed (i.e., bandwidths increased), but the width of a neural pulse appears to have held steady at circa 2 ms over all species over hundreds of millions of years, no doubt for a variety of compelling reasons. (Certain owls' auditory systems have spikes only about 1 ms wide, but we are not looking for factor of 2 explanations here.) The obvious solution was to expand the spatial dimension. Organisms have done precisely that, relying on parallel processing by more and more neurons in order to progress up the phylogenic tree. If there is any ECC coding done by neurons, it likely is done spatially over the huge numbers of neurons involved. Indeed, strong correlations have been observed in the spiking behaviors of neighboring neurons, but these may simply be consequences of the need to obtain high resolution of certain environmental stimuli that are themselves inherently correlated and/or the need to direct certain spike trains to more locations, or more widely dispersed locations, than it is practical for a single neuron to visit. There is not yet any solid evidence that neurons implement ECC. Similarly, although outside the domain of this paper, we remark that there is not yet any concrete evidence that redundancies in the genetic code play an ECC role; if it turns out they do, said ECC capability clearly also will be predominately spatial as opposed to temporal in nature.

answer lies in the biology. Obviously, energy is consumed every time one of the channel's neurons generates an action potential, or spike.[7] It is also true that when a spike arrives at a synapse located on a dendrite of one of the channel's neurons, energy usually is expended in order to convert it into a contribution to the post-synaptic potential.[8] This is effected via a sequence of electrochemical processes the end result of which is that vesicles containing neurotransmitter chemicals empty them for transportation across the synaptic cleft. This, in turn, either increases or decreases the post-synaptic potential (equivalently, the post-synaptic current), respectively as the synapse is an excitatory or an inhibitory one. The expected energy dissipated in the synapses in Figure 1 at time k therefore depends on $\underline{X}(k)$ and $\underline{Y}(k-1)$, while that dissipated in the axons depends on $\underline{Y}(k)$. The average energy dissipated in the coalition at time k therefore is the expected value of one function of $(\underline{X}(k), \underline{Y}(k-1))$ and another function of $\underline{Y}(k)$, with the effect of quantal synaptic failures (see footnote concerning QSF) usually well approximated by multiplying the first of these two functions by s. For purposes of the theorem we are about to present, it suffices to make a less restrictive assumption that the average resources expended at time k are the expected value of some function solely of $(\underline{X}(k), \underline{Y}(k-1), \underline{Y}(k))$. We may impose either a schedule of expected resource depletion constraints as a function of k or simply constrain the sum of the k^{th} expected resource depletion over some appropriate range of the discrete time index k. Average energy expenditure, which we believe to be the dominant operative constraint in practice, is an important special case of this general family of resource constraint functions.

Let $PSP_m(k)$ denote the post-synaptic potential of the m^{th} neuron in the coalition at time k. Then the output $Y_m(k)$ of this neuron, considered to be 1 if there is a spike at time k and 0 if there isn't, is given by

[7] Actually, the energy gets consumed principally during the process of re-setting chemical concentrations in and around the neuron after each time it spikes so as to prepare it to fire again should sufficient excitation arrive. The actual transmission of a spike is more a matter of energy conversion than of energy dissipation.

[8] Spikes arriving at synapses often are ignored, a phenomenon known as *quantal synaptic failure* (QSF). Its name notwithstanding, QSF actually is one of natural selection's finer triumphs, enhancing the performance of neural coalitions in several ingenious respects the details of which can be found in the works of Levy and Baxter[13] [14]. Let $S_{im}(k)$ be a binary random variable that equals 1 if QSF does not occur at synapse (i,m) at time k and equals 0 if it does; that is, $S_{im} = 1$ denotes a quantal synaptic success at synapse (i,m) at time k. Often, the $S_{im}(k)$'s can be well modeled as Bernoulli-s random variables, i.e., as being i.i.d. over i, m and k with common distribution $P(S = 1) = 1 - P(S = 0) = s$; in practice, $s \in_{\approx} [0.25, 0.9]$. The phenomenon of QSF then may be modeled by multiplying the spike, if any, afferent to synapse (i,m) at time k by $S_{im}(k)$. This is seen to be equivalent to installing what information theorists call a Z-channel [12] at every synapse. Were it not for QSF, the coalition channel would be effectively deterministic when viewed as an operator that transforms $\{\underline{X}\}$ into $\{\underline{Y}\}$, since the signal-to-noise ratio on neural channels usually is quite strong. However, if the channel is viewed as an operator only from $\{\underline{X}_{BU}\}$ to $\{\underline{Y}\}$, with $\{\underline{X}_{TD}\}$ considered to be random side information, QSF may no longer be its dominant source of randomness.

$$Y_m(k) = U(PSP_m(k) - T),$$

where $U(\cdot)$ is the unit step function. The above discussion of Figure 1 lets us write

$$PSP_m(k) = \sum_l X_l(k) W_{lm} Q_{lm}(k) S_{lm}(k) + \sum_i Y_i(k-1) W_{im} Q_{im}(k) S_{im}(k),$$

where W_{im} is the signed weight of synapse (i,m), $Q_{im}(k)$ is the random size of the quantity of neurotransmitter that will traverse the synaptic cleft in response to an afferent spike at synapse (i,m) at time k, if synaptic quantal failure does not occur there then, and $S_{im}(k)$ equals 0 or 1, respectively, in accordance with whether said quantal synaptic failure does or does not occur. The spiking threshold $T = T_m(k)$ of neuron m at time k varies with m and k, though usually only mildly with m.

Note that this channel model is such that successive channel output vectors $\underline{Y}(k)$ are generated independently, conditional on their corresponding input vectors $\underline{X}(k)$ and local feedback vectors $\underline{Y}(k-1)$; that is,

$$p\left(\underline{y}_1^n \middle| \underline{x}_1^n, \underline{y}_0\right) = \prod_{k=1}^{n} p\left(\underline{y}_k \middle| \underline{x}_k, \underline{y}_{k-1}\right). \quad (1)$$

As information theorists, one of our inclinations would be to investigate conditions sufficient to ensure that such a finite-state channel model with feedback has a Shannon capacity. That is, we might seek conditions under which the maximum over channel input processes $\{\underline{X}(k)\}$ of the mutual information rate between said input and the output process $\{\underline{Y}(k)\}$ it generates, subject to whatever constraints are imposed on the input and/or the output, equals the maximum number of bits per channel use at which information actually can be sent reliably over the constrained channel. Today, however, we shall focus only on the mutual-information-rate-maximizing constrained input process and the structure of the joint (input,output) stochastic process it produces.

Temporarily assume that there is a genuine encoder at the channel input which, for purposes of generating input $\underline{x}_k$ at time k remembers all the past inputs $\underline{x}_1^{k-1}$ and all the past local feedback (i.e., past output) values $\underline{y}_0^{k-1}$; here, $\underline{y}_0$ represents the "initial" state. Obviously, the maximum mutual information rate achievable under these circumstances is an upper bound to that which could be achieved when only $\underline{y}_{k-1}$ is available at the channel input at time k, with $\underline{x}_1^{k-1}$ and $\underline{y}_0^{k-2}$ by then no longer being available there. We will show that this upper bound can be met even when access is denied to said past input and local feedback vectors. This, in turn, helps demystify how a real neural network structured as in Figure 1 can be information-theoretically optimum despite its not possessing a classical encoder at the network's input. The postulated full-memory encoder can generate any input process whose probabilistic structure is described by a member of the set $P\left(\underline{X}_1^n\right)$ of probabilistic distributions of the form

$$\prod_{k=1}^{n} p\left(\underline{x}_k \middle| \underline{x}_1^{k-1}, \underline{y}_0^{k-1}\right). \tag{2}$$

Now define another set of input distributions on $\underline{X}_1^n$, $P^*\left(\underline{X}_1^n\right)$, having all probability mass functions of the form

$$\prod_{k=1}^{n} p\left(\underline{x}_k \middle| \underline{y}_{k-1}\right). \tag{3}$$

Compared with (2), this set contains only those input distributions for which, given $\underline{Y}_{k-1}$, $\underline{X}_k$ becomes conditionally independent of all the previous inputs $\underline{X}_1^{k-1}$ and all the previous outputs $\underline{Y}_0^{k-2}$.

9 Statement and Proof of Main Theorem

Our main result is stated as the following theorem.

Theorem 1 *The maximum mutual information rate between the channel's input and output processes is attained inside $P^*\left(X_1^n\right)$, uniformly in the initial conditions $\underline{Y}_0$. Moreover, if we restrict the distribution of the inputs X_1^n to $P^*\left(X_1^n\right)$, let Y_1^n denote the corresponding output, and let $\underline{Y}_0$ denote the initial channel state, then we have*

1. $\{\underline{Y}_k, k = 0,1,\ldots,n\}$ *is a first-order Markov chain,*
2. $\{(\underline{X}_k, \underline{Y}_k), k = 1, 2, \ldots, n\}$ also is a first-order Markov chain.

Remarks: (i) $\{(\underline{X}_k)\}$ is not necessarily a first-order Markov chain, though we have not yet endeavored to construct an example in which it fails to be. Since $\{\underline{X}_k\}$ depends, in part, on bottom-up information derived from the environment, it is unrealistic to expect it to exhibit Markovianness, especially at precisely the time step duration of the model. (ii) The theorem's Markovian results help explain how many neural regions can be hierarchically stacked, as is the case in the human visual system, without necessarily engendering unacceptably large response times. (iii) The theorem reinforces a view of sensory brain function as a procedure for recursively estimating quantities of interest in a manner that becomes increasingly informed and accurate.

Proof of Theorem 1[9]

We suppress underlining of vectors and abuse notation by writing $X_1^n \in P\left(X_1^n\right)$ to indicate that X_1^n is distributed according to some distribution in $P\left(X_1^n\right)$. Furthermore, the expression $\left(X_1^n, Y_1^n\right) \in P\left(X_1^n, Y_0\right)$ means that $X_1^n \in P\left(X_1^n\right)$ and Y_1^n is the output that corresponds to input X_1^n and channel initial state $\underline{Y}_0$. First we establish the Markovianness of the output process.

Lemma 1 *If $\left(X_1^n, Y_1^n\right) \in P^*\left(X_1^n, Y_0\right)$, then Y_0^n is a first-order Markov chain.*

Proof: For all k we have

$$p\left(y_k \middle| y_0^{k-1}\right) = \sum_{x_k} p\left(y_k \middle| x_k, y_0^{k-1}\right) p\left(x_k \middle| y_0^{k-1}\right).$$

Since $X_1^n \in P^*\left(X_1^n\right)$,

$$p\left(x_k \middle| y_0^{k-1}\right) = p\left(x_k \middle| y_{k-1}\right).$$

Thus, with reference to the conditional memoryless of the channel (cf. equation (1)), we have

$$p\left(y_k \middle| y_0^{k-1}\right) = \sum_{x_k} p\left(y_k \middle| x_k, y_{k-1}\right) p\left(x_k \middle| y_{k-1}\right) = p\left(y_k \middle| y_{k-1}\right). \tag{4}$$

Remark: Depending on the input pmf's $p\left(x_k \middle| y_{k-1}\right)$, $k = 1, 2, \ldots$, Y_0^n can be either a homogeneous or a nonhomogeneous Markov chain. If the pmf $p\left(x_k \middle| y_{k-1}\right)$ does not vary with k, then Y_0^n is homogeneous; otherwise, it's nonhomogeneous.

We next derive an upper bound on the mutual information given Y_0 between the two components of any $\left(X_1^n, Y_1^n\right) \in P\left(X_1^n, Y_0\right)$, which is needed for the proof.

$$\begin{aligned} & I\left(X_1^n; Y_1^n \middle| Y_0\right) \\ &= H\left(Y_1^n \middle| Y_0\right) - H\left(Y_1^n \middle| X_1^n, Y_0\right) \\ &\overset{(a)}{=} \sum_{k=1}^{n} H\left(Y_k \middle| Y_0^{k-1}\right) - \sum_{k=1}^{n} H\left(Y_k \middle| X_1^n, Y_0^{k-1}\right) \\ &\overset{(b)}{=} \sum_{k=1}^{n} H\left(Y_k \middle| Y_0^{k-1}\right) - \sum_{k=1}^{n} H\left(Y_k \middle| X_k, Y_{k-1}\right) \\ &\overset{(c)}{\le} \sum_{k=1}^{n} H\left(Y_k \middle| Y_{k-1}\right) - \sum_{k=1}^{n} H\left(Y_k \middle| X_k, Y_{k-1}\right) \\ &= \sum_{k=1}^{n} I\left(X_k; Y_k \middle| Y_{k-1}\right), \end{aligned} \tag{5}$$

where (a) is the chain rule; (b) follows from the channel property $p\left(y_k \middle| x_1^n, y_0^{k-1}\right) = p\left(y_k \middle| x_k, y_{k-1}\right)$ for all k and all $n > k$; and (c) follows from the fact that increasing conditioning can only reduce entropy. Notice that the inequality in (c) be-

[9] This proof is joint work with Yuzheng Ying [7].

comes equality when Y_0^n is a Markov chain. Therefore, it follows from Lemma 1 that, for any $\left(X_1^n, Y_1^n\right) \in P^*\left(X_1^n, Y_0\right)$

$$I\left(X_1^n; Y_1^n \middle| Y_0\right) = \sum_{k=1}^{n} I\left(X_k; Y_k \middle| Y_{k-1}\right). \tag{6}$$

We now show that, for any $\left(X_1^n, Y_1^n\right) \in P\left(X_1^n, Y_0\right)$, there's a $\left(\hat{X}_1^n, \hat{Y}_1^n\right) \in P\left(X_1^n, Y_0\right)$ such that

$$I\left(\hat{X}_1^n; \hat{Y}_1^n \middle| Y_0\right) \geq I\left(X_1^n; Y_1^n \middle| Y_0\right). \tag{7}$$

This assertion says that $I_n(Y_0)$ is attained inside $P^*\left(X_1^n\right)$. Since $\left(X_1^n, Y_1^n\right)$ is a pair of channel (input,output) sequences under the initial state Y_0,

$$\begin{aligned} & p_{X_1^n, Y_1^n | Y_0}\left(x_1^n, y_1^n \middle| y_0\right) \\ &= \prod_{k=1}^{n} p_{X_k | X_1^{k-1}, Y_0^{k-1}}\left(x_k \middle| x_1^{k-1}, y_0^{k-1}\right) \\ &\cdot\ p_{Y_k | X_1^k, Y_0^{k-1}}\left(y_k \middle| x_1^k, y_0^{k-1}\right) \\ &= \prod_{k=1}^{n} p_{X_k | X_1^{k-1}, Y_0^{k-1}}\left(x_k \middle| x_1^{k-1}, y_0^{k-1}\right) \\ &\cdot\ p_{Y_k | X_k, Y_{k-1}}\left(y_k \middle| x_k, y_{k-1}\right), \end{aligned} \tag{8}$$

where the last equality follows from (1). We now construct a $\left(\hat{X}_1^n, \hat{Y}_1^n\right) \in P^*\left(X_1^n, Y_0\right)$ distributed according to the pmf

$$\begin{aligned} & p_{\hat{X}_1^n, \hat{Y}_1^n | \hat{Y}_0}\left(x_1^n, y_1^n\right) = \\ & \prod_{k=1}^{n} p_{\hat{X}_k | \hat{Y}_{k-1}}\left(x_k \middle| y_{k-1}\right) \\ & \cdot\ p_{\hat{Y}_k | \hat{X}_k, \hat{Y}_{k-1}}\left(y_k \middle| x_k, y_{k-1}\right), \end{aligned} \tag{9}$$

where we set $p_{\hat{X}_k, \hat{Y}_{k-1}}(\cdot|\cdot)$ equal to $p_{X_k, Y_{k-1}}(\cdot|\cdot)$ so that for all $k \geq 1$

$$p_{\hat{X}_k | \hat{Y}_{k-1}}\left(x_k \middle| y_{k-1}\right) = p_{X_k | Y_{k-1}}\left(x_k \middle| y_{k-1}\right), \tag{10}$$

and

$$p_{\hat{Y}_k | \hat{X}_k, \hat{Y}_{k-1}}\left(y_k \middle| x_k, y_{k-1}\right) = p_{Y_k | X_k, Y_{k-1}}\left(y_k \middle| x_k, y_{k-1}\right). \tag{11}$$

$\hat{Y}_0$ therein is just an alias for the random variable Y_0. Unlike X_k in (8), $\hat{X}_k$ is restricted to be conditionally independent of $\left(\hat{X}_1^{k-1}, \hat{Y}_0^{k-2}\right)$, given $\hat{Y}_{k-1}$. Thus, $\hat{X}_1^n \in P^*\left(X_1^n\right)$. Equation (11), together with $\hat{Y}_0 = Y_0$, assures us that $\hat{Y}_1^n$ is indeed the output from our channel in response to input $\hat{X}_1^n$ and initial state Y_0; i.e., $\left(\hat{X}_1^n, \hat{Y}_1^n\right) \in P^*\left(X_1^n, Y_0\right)$. It is obvious that the joint pmf of $\left(\hat{X}_1^n, \hat{Y}_0^n\right)$ is different from that of $\left(X_1^n, Y_0^n\right)$. However, we have the following lemma.

Lemma 2 *Assume* $\left(X_1^n, Y_1^n\right) \in P\left(X_1^n, Y_0\right)$. *Let* $\left(\hat{X}_1^n, \hat{Y}_1^n\right)$ *be defined as in (9) and let* $\hat{Y}_0 = Y_0$. *Then*

$$\begin{aligned} & p_{\hat{Y}_k, \hat{X}_k, \hat{Y}_{k-1}}\left(y_k, x_k, y_{k-1}\right) = \\ & p_{Y_k, X_k, Y_{k-1}}\left(y_k, x_k, y_{k-1}\right) \quad \forall k \geq 1. \end{aligned} \tag{12}$$

Proof: It follows from (10) and (11) that

$$\begin{aligned} & p_{\hat{Y}_k, \hat{X}_k | \hat{Y}_{k-1}}\left(y_k, x_k \middle| y_{k-1}\right) = \\ & p_{Y_k, X_k | Y_{k-1}}\left(y_k, x_k \middle| y_{k-1}\right) \quad \forall k \geq 1. \end{aligned} \tag{13}$$

Since $\hat{Y}_0 = Y_0$, we may write

$$\begin{aligned} & p_{\hat{Y}_1, \hat{X}_1, \hat{Y}_0}\left(y_1, x_1, y_0\right) \\ &= p_{\hat{Y}_1, \hat{X}_1 | \hat{Y}_0}\left(y_1, x_1 \middle| y_0\right) p_{Y_0}\left(y_0\right) \\ &= p_{Y_1, X_1 | Y_0}\left(y_1, x_1 \middle| y_0\right) p_{\hat{Y}_0}\left(y_0\right) \\ &= p_{Y_1, X_1, Y_0}\left(y_1, x_1, y_0\right). \end{aligned} \tag{14}$$

That is, the lemma statement (12) holds for $k = 1$, from which it follows by marginalization that

$$p_{\hat{Y}_1}\left(y_1\right) = p_{Y_1}\left(y_1\right).$$

The same arguments as in (14) now can be used to verify that (12) holds for $k = 2$. Repeating this argument for $k = 3$, and so on, establishes the desired result for all $k \geq 1$.

Since $\left(\hat{X}_1^n, \hat{Y}_1^n\right) \in P^*\left(X_1^n, Y_0\right)$, we know from (6) that

$$I\left(\hat{X}_1^n; \hat{Y}_1^n \middle| \hat{Y}_0\right) = \sum_{k=1}^{n} I\left(\hat{X}_k; \hat{Y}_k \middle| \hat{Y}_{k-1}\right).$$

Next, recall from (5) that $I\left(X_1^n; Y_1^n \middle| Y_0\right) \leq \sum_{k=1}^{n} I\left(X_k; Y_k \middle| Y_{k-1}\right)$, and observe from Lemma 2 that

$$I\left(\hat{X}_k; \hat{Y}_k \middle| \hat{Y}_{k-1}\right) = I\left(X_k; Y_k \middle| Y_{k-1}\right), \ \forall k \geq 1, \tag{15}$$

so

$$\sum_{k=1}^{n} I\left(\hat{X}_k; \hat{Y}_k \middle| \hat{Y}_{k-1}\right) = \sum_{k=1}^{n} I\left(X_k; Y_k \middle| Y_{k-1}\right). \tag{16}$$

Therefore, $I\left(\hat{X}_1^n; \hat{Y}_1^n \middle| \hat{Y}_0\right) \geq I\left(X_1^n; Y_1^n \middle| Y_0\right)$, which is (7).

To show that the joint (input,output) process is Markovian when $X_1^n \in P^*\left(X_1^n\right)$, we write

$$\Pr\left(X_k, Y_k \middle| X_1^{k-1}, Y_1^{k-1}\right)$$
$$= \Pr\left(Y_k \middle| X_k, X_1^{k-1}, Y_1^{k-1}\right) \Pr\left(X_k \middle| X_1^{k-1}, Y_1^{k-1}\right)$$
$$\overset{(a)}{=} \Pr\left(Y_k \middle| X_k, Y_{k-1}\right) \Pr\left(X_k \middle| Y_{k-1}\right)$$
$$= \Pr\left(Y_k \middle| X_k, X_{k-1}, Y_{k-1}\right) \Pr\left(X_k \middle| Y_{k-1}, X_{k-1}\right)$$
$$= \Pr\left(X_k, Y_k \middle| X_{k-1}, Y_{k-1}\right),$$

where (a) follows from (1) and the condition $X_1^n \in P^*\left(X_1^n\right)$. Theorem 1 is proved.

For a broad class of constraints on the channel input and/or output, $I_n(Y_0)$ still is attained inside $P^*\left(X_1^n\right)$. Specifically, for all constraints on expected values of functions of triples of the form (Y_k, X_k, Y_{k-1}), imposed either as a schedule of such constraints versus k or as sums or arithmetic averages over k of functions of said triples, the constrained value of $I_n(Y_0)$ is attained by an input process whose distribution conforms to (3). To see this, for any $\left(X_1^n, Y_1^n\right) \in P\left(X_1^n, Y_0\right)$ satisfying one or more constraints of this type, we construct $\left(\hat{X}_1^n, \hat{Y}_1^n\right)$ as in (9). By Lemma 2, $\left(X_1^n, Y_1^n\right)$ and $\left(\hat{X}_1^n, \hat{Y}_1^n\right)$ are such that for each fixed k, (X_k, Y_k, Y_{k-1}) and $\left(\hat{X}_k, \hat{Y}_k, \hat{Y}_{k-1}\right)$ are identically distributed. This assures us that the expected value of any function of $\left(\hat{X}_k, \hat{Y}_k, \hat{Y}_{k-1}\right)$ is the same as that for (X_k, Y_k, Y_{k-1}), so $\left(\hat{X}_1^n, \hat{Y}_1^n\right)$ also is admissible for the same values of the constraints. Energy constraints on the inputs and outputs are a special case. Thus, processes which communicate information among the brain's neurons in an energy-efficient manner will exhibit the Markovian properties cited in Theorem 1, at least to the degree of accuracy to which the model of Figure 1 reflects reality.

10 Review of Cost-Capacity and Rate-Distortion

In preparation for addressing the topic of decoding, or the lack thereof, let us recall Shannon's formulas characterizing the probability distributions that solve the variational problems of calculating capacity-cost functions of channels and rate-distortion functions of sources.

The cost-capacity variational problem is defined as follows. We are given the transition probabilities $\{p(y \mid x), (x,y) \in \mathcal{X} \times \mathcal{Y}\}$ of a discrete memoryless channel (dmc) and a set of nonnegative numbers $\{c(x) \geq 0, x \in \mathcal{X}\}$, where $c(x)$ is the cost incurred each time the symbol x is inserted into the channel. We seek the probability distribution $\{p(x), x \in \mathcal{X}\}$ that maximizes the mutual information subject to the constraint that the average input cost does not exceed S. We denote this maximum by

$$C(S) := \max_{\{p(x)\} \in \mathcal{S}} \sum_x \sum_y p(x) p\left(y \middle| x\right) \log \left(p\left(y \middle| x\right) / \tilde{p}\left(y\right)\right) \tag{17}$$

where $\mathcal{S} = \{\{p(x)\} : \Sigma_x\, p(x)c(x) \leq S\}$ and

$$\tilde{p}\left(y\right) = \sum_x p\left(x\right) p\left(y \middle| x\right). \tag{18}$$

$C(S)$ is a concave function that usually satisfies $C(0) = 0$ and $\lim_{S\to\infty} C(S) = C$. The constant C, called either the unconstrained capacity or simply the capacity of the channel, is finite if $\mathcal{X}$ and/or $\mathcal{Y}$ have finite cardinality but may be infinite otherwise.

The rate-distortion variational problem is defined as follows. We are given the letter probabilities $\{p(u), u \in \mathcal{U}\}$ of a discrete memoryless source (dms) and a set of nonnegative numbers $\{d(u, v) \geq 0, (u, v) \in \mathcal{U} \times \mathcal{V}\}$. Here, $d(u, v)$ measures the distortion that occurs whenever the dms produces the letter $u \in \mathcal{U}$ and the communication system delivers to a user located at its output the letter $v \in \mathcal{V}$ as its approximation of said u. The alphabets $\mathcal{U}$ and $\mathcal{V}$ may or may not be identical. In fact, the appropriate $\mathcal{V}$ and distortion measure vary from user to user. Alternatively, and more apropos of application to a living organism, they vary over the different uses a single organism has for the information. In what follows we therefore speak of (source,use)-pairs instead of the usual terminology of (source,user)-pairs. In rate distortion theory we seek the transition probability assignment $\{q(v \mid u), (u, v) \in \mathcal{U} \times \mathcal{V}\}$ that minimizes the average mutual information subject to the constraint that the average distortion does not exceed D. We denote this minimum by

$$R\left(D\right) := \min_{\left\{q\left(v|u\right)\right\} \in \mathcal{D}} \sum_u \sum_v p\left(u\right) q\left(v \middle| u\right) \log \left(q\left(v \middle| u\right) / q\left(v\right)\right), \tag{19}$$

where $\mathcal{D} = \{\{q(v \mid u) : \Sigma_u \Sigma_v\, p(u)\, q(v \mid u)\, d(u, v) \leq D\}$ and

$$q\left(v\right) = \sum_u p\left(u\right) q\left(v \middle| u\right). \tag{20}$$

Viewed as a function of D, $R(D)$ is called the rate-distortion function. It is convex on the range $[D_{\min}, D_{\max}]$, where $D_{\min} = \Sigma_u \min_v d(u, v)$ and $D_{\max} = \min_v \Sigma_u\, p(u) d(u, v)$. $R(D) = 0$ for $D \geq D_{\max}$ and is undefined for $D < D_{\min}$. $R(D_{\min})$ equals the source entropy $H = -\Sigma_u\, p(u) \log p(u)$ if for each $u \in \mathcal{U}$ there is a unique $v \in \mathcal{V}$, call it $v(u)$, that minimizes $d(u, v)$ and $v(u) \neq v(u')$ if $u \neq u'$; otherwise, $R(D_{\min}) < H$.

In each of these variational problems, Lagrange optimization yields a necessary condition that relates the extremizing distribution to the constraint function. For the cost-capacity problem this condition, displayed as an expression for $c(x)$ in terms of the given channel transition probabilities and the information-maximizing $\{p(x)\}$, reads

$$c(x) = c_1 \sum_y p(y|x) \log\left(p(y|x) / \tilde{p}(y) \right) + c_2, \tag{21}$$

where $\tilde{p}(y)$ is given linearly in terms of said $\{p(x)\}$ by (18). The constant c_2 represents a fixed cost per channel use that simply translates the $C(S)$ curve vertically, so no loss in generality results from setting $c_2 = 0$. The constant c_1 is interchangeable with the choice of the logarithm base, so we may set $c_1 = 1$, again without any essential loss of generality. It follows that in order for optimality to prevail the cost function must equal the Kullback-Leibler distance[10] between the conditional distribution $\{p(y \mid x), y \in \mathcal{Y}\}$ of the channel's output when the channel input equals letter x and the unconditional output distribution $\{\tilde{p}(y), y \in \mathcal{Y}\}$ obtained by averaging as in (18) over the information-maximizing $\{p(x), x \in \mathcal{X}\}$. Since the expectation over X of the K-L distance between $\{p(y \mid X)\}$ and $\{\tilde{p}(y)\}$ is the mutual information between the channel's input and output, we see that applying the constraint $\Sigma_x p(x)c(x) \le S$ is equivalent, *when the $C(S)$-achieving input distribution is in force*, to maximizing the average mutual information subject to the average mutual information not exceeding a specified amount, call it I. Obviously, this results in $C(I) = I$, a capacity-cost function that is simply a straight line at 45°.

This perhaps confusing state of affairs requires further explanation, since information theorists are justifiably not accustomed to the capacity-cost curve being a straight line.[11] Since studying a well-known example often sheds light on the general case, let us consider again Shannon's famous formula $C(S) = (1/2)\log(1+S/N)$ for the capacity-cost function of the time-discrete, average-power-limited memoryless AWGN channel; clearly, it is a strictly concave function of S. Of course, in this formula S is a constraint on the average power expended to transmit information across the channel, not on the average mutual information between the channel input and the output. Next, recall that for this AWGN, whose transition probabilities are given by $p(y \mid x) = \exp\{-(y-x)^2/2N\}/\sqrt{2\pi N}$, *the optimum power constrained input distribution is known to be Gaussian, namely* $p(x) = \exp\{-x^2/2S\}/\sqrt{2\pi S}$. When $D(p(y \mid x) \| \tilde{p}(y))$ is evaluated in the case of this optimum input, it indeed turns out to be proportional to x^2 plus a constant. Hence, there is no difference in this example between considering the constraint to be imposed on the expected value of X^2 or considering it to be imposed on the expected value of $D(p(y \mid X) \| p(y))$.

The physical significance of an average power constraint is evident, but what, if anything, is the physical meaning of an average K-L distance constraint? First observe that, if for some x it were to be the case that $\{p(y \mid x), y \in \mathcal{Y}\}$ is the same as $\{p(y), y \in \mathcal{Y}\}$, then one would be utterly unable to distinguish on the basis of the channel output between transmission and non-transmission of the symbol x. Little if any resources would need to be expended to build and operate a channel in which $\{p(y \mid x), y \in \mathcal{Y}\}$ does not change with x, since the output of such a channel is independent of its input. In order to assure that the conditional distributions on the channel output space given various input values are well separated from one another, resources must be expended. We have seen that in the case of an AWGN the ability to perform such discriminations is constrained by the average transmission power available; in a non-Gaussian world, the physically operative quantity to constrain likely would be something other than power. In this light I believe (21) is telling us, among others things, that if one is not sure *a priori* to what use(s) the information conveyed through the channel output will be put, one should adopt the viewpoint that the task is to keep the various inputs as easy to discriminate from one another as possible subject to whatever physical constraint(s) are in force. We adopt this viewpoint below in our treatment of the 'decoding' problem.

For the rate-distortion problem Shannon [4] observed that Lagrange minimization over $\{q(v \mid u)\}$ leads to the necessary condition

$$q_s(v \mid u) = \lambda_s(u) q_s(v) \exp(sd(u, v)), \tag{22}$$

where $s \in [-\infty, 0]$ is a parameter that equals the slope $R'(D_s)$ of the rate-distortion function at the point $(D_s, R(D_s))$ that it generates, and $\{q_s(v), v \in \mathcal{V}\}$ is an appropriately selected probability distribution over the space $\mathcal{V}$ of source approximations. Since $q(v \mid u)$ must sum to 1 over v for each fixed $u \in \mathcal{U}$, we have

$$\lambda_s(x) = \left[\sum_v q_s(v) \exp(sd(u, v)) \right]^{-1}. \tag{23}$$

Except in some special but important examples, given a parameter value s it is difficult to find the optimum $\{q_s(v)\}$, and hence the optimum $\{q_s(v \mid u)\}$ from (22).[12] We can recast (22) as an expression for $d(u, v)$ in terms of $q_s(v \mid u)$, namely[13]

$$d(u, v) = (-1/|s|) \log\left(q_s(v|u) / q_s(v) \right) + (1/|s|) \log \lambda_s(u). \tag{24}$$

[10]The K-L distance, or relative entropy, of two probability distributions $\{p(w), w \in \mathcal{W}\}$ and $\{q(w), w \in \mathcal{W}\}$ is given by $D(p\|q) := \sum_w p(w) \log(p(w)/q(w))$. It is nonnegative and equals 0 if and only if $q(w) = p(w)$ for all $w \in \mathcal{W}$.

[11]This does happen for an AWGN channel for all $S << N_0 W$, and hence for all practical values of S in the case of a time-continuous AWGN of extremely broad bandwidth.

[12]Kuhn-Tucker theory tells us that the necessary and sufficient condition for $\{q_s(v)\}$ to generate, via (22), a point on the rate-distortion curve at which the slope is s is $c_s(v) := \Sigma_u \lambda_s(u)p(u) \exp\{sd(u, v)\} \le 1$ for all v, where $\lambda_s(u)$ is given by equation (23) and equality prevails for every v for which $q_s(v) > 0$. Recursive algorithms developed by Blahut [19] and by Rose [20] allow rate-distortion functions to be calculated numerically with great accuracy at moderate computational intensity.

[13]Equations (21) and (24) perhaps first appeared together in the paper by Gastpar et al. [18]. Motivated by exposure to my Examples 1 and 2, they derived conditions for double matching of more general sources and channels, confining attention to the special case of deterministic $r(x \mid u)$ and $w(v \mid y)$.

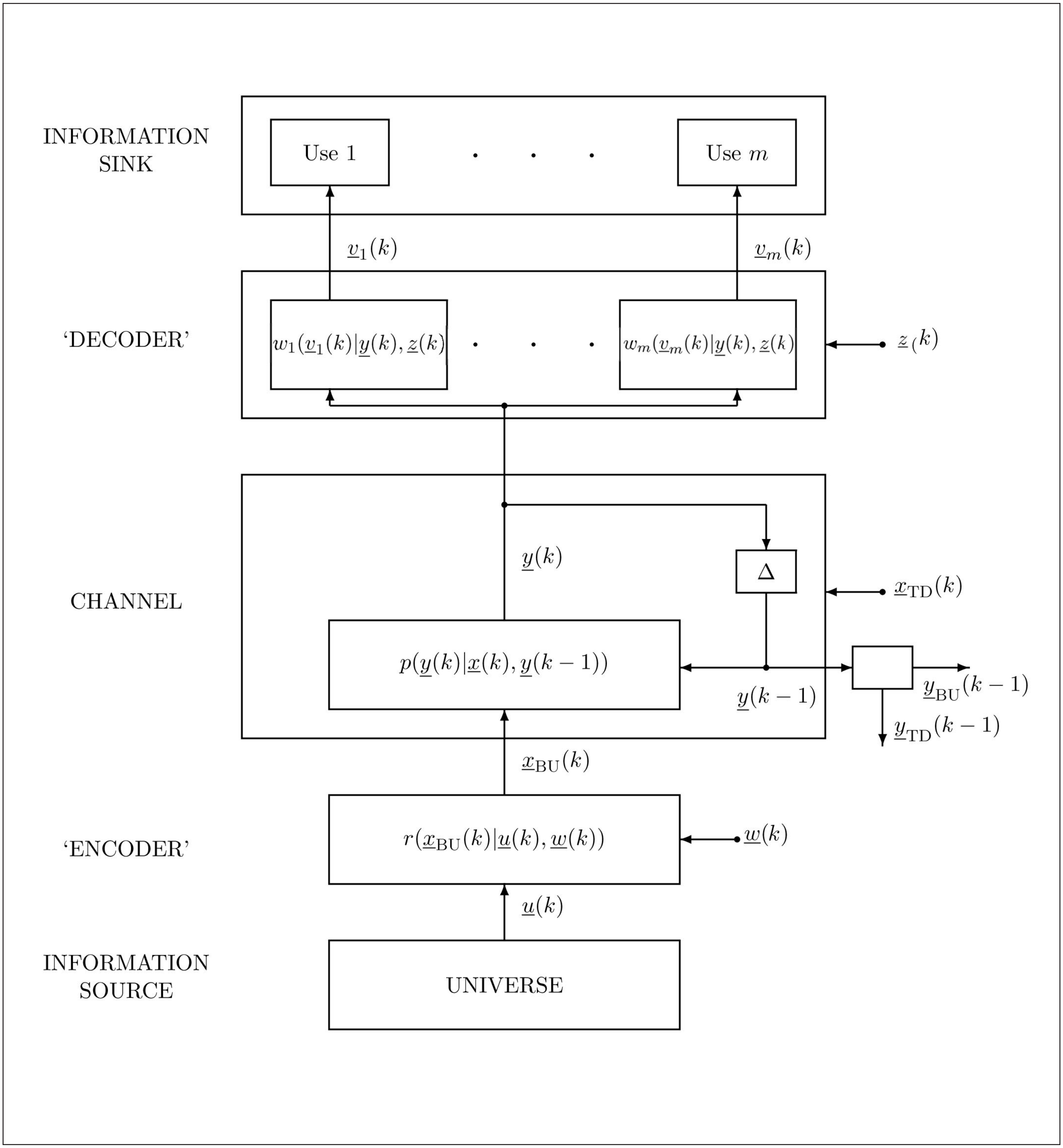


Figure 2: Block Diagram of Neural Sensory Processing.

Signals: {$\underline{u}(k)$} - Sensory Data From Universe; {$\underline{w}(k)$} - Side Information to 'Encoder';
{$\underline{x}_{BU}(k)$} - Channel Bottom-Up Input; {$\underline{x}_{TD}(k)$} - Channel Top-Down Input;
{$\underline{y}(k)$} - Channel Output; {$\underline{z}(k)$} - Side Information to 'Decoder';
{$\underline{v}_i(k)$} - Extracted Data for Use i, $1 \le i \le m$.

Fidelity Constraints: $Ed_i(\underline{u}(k), \underline{v}_i(k + \tau_i)) \le D_i$, $1 \le i \le m$.

Since v does not appear in the second term on the right-hand side of (24), that term reflects only indirectly on the way in which a system that is end-to-end optimum in the sense of achieving the point $(D_s, R(D_s))$ on the rate-distortion function probabilistically reconstructs the source letter u as the various letters $v \in \mathcal{V}$. Recalling that $\log(q_s(v \mid u)/q_s(v))$ is the mutual information $i_s(u;v)$ between symbols u and v for the end-to-end (i.e., source-to-use) optimum system, we see that the right way to build said system is to make the mutual information between the letters of pairs (u, v) decrease as the distortion between them increases. Averaging over the joint distribution $p(u)q_s(v \mid u)$ that is optimum for parameter value s affirms the inverse relation between average distortion and average mutual information that pertains in rate-distortion. This relationship is analogous to the directly-varying relation between average cost and average mutual information in the channel variational problem.

11 Low-Latency Multiuse Decoding

For purposes of the present exposition, a key consequence of the preceding paragraph is that it helps explain how a channel $p(y \mid x)$ can be matched to many (source,use)-pairs at once. Specifically, under our definition channel $p(y \mid x)$ is matched to source $p(u)$ and distortion measure $d(u, v)$ at slope s on their rate-distortion function if, and only if, there exists a pair of conditional probability distributions $\{r_s(x \mid u)\}$ and $\{w_s(v \mid y)\}$ such that the optimum end-to-end system transition probabilities $\{q_s(v \mid u)\}$ in the rate-distortion problem can be written in the form

$$q_s\left(v|u\right) = \sum_x \sum_y r_s\left(x|u\right) p\left(y|x\right) w_s\left(v|y\right). \tag{25}$$

It should be clear that (25) often can hold for many (source,use) pairs that are of interest to an organism. In such instances it will be significantly more efficient computationally for the organism to share the $p(y \mid x)$ part of the construction of the desired transition probability assignments for these (source,use) pairs rather than to have to in effect build and then operate in parallel a separate version of it for each of said applications. This will be all the more so the case if it is not known until after $p(y \mid x)$ has been exercised just which potential uses appear to be intriguing enough to bother computing their $w(v \mid y)$-parts and which do not.

I conjecture that the previous paragraph has much to say about why neural coalitions are constructed and interconnected the way they are. Namely, the coalitionwise transition probabilities effected are common to numerous potential applications only some of which actually get explored. The situation is sketched schematically in Figure 2, from which the reader can see how a given neural coalition, viewed as a channel, might both be matched by the source that drives it and at the same time could help match that source to many potential uses via interconnection with other coalitions and subcoalitions.

Whether or not use i, associated with some i^{th} neural subcoalition described by transition probabilities $\{w_{s,i}(v_i \mid y), v_i \in \mathcal{V}_i\}$, gets actively explored at a given instant depends on what side information in addition to part of $\{\underline{Y}_k\}$ gets presented to it. That side information depends, in turn, on Bayesian-style prior probabilities that are continually being recursively updated as the bottom-up and top-down processing of data from stimuli proceeds.[14] When said side information is relatively inhibitory rather than excitatory, the subregion does not "ramp up". Then energy is saved but of course less information is conveyed.[15]

Acknowledgement

Among the several people who were kind enough to read and/or proofread this manuscript, Ram Zamir deserves particular mention.

I am deeply indebted to Professors William B. "Chip" Levy and James W. Mandell of the University of Virginia Medical School's neurosurgery and neuropathology departments, respectively, for frequent, lengthy, and far-ranging discussions of neuroscience over the past seven years. In his role as my neuroscience mentor, Chip has graciously enlightened me regarding his own and others' theoretical approaches to computational neuroscience. This presentation is in substantial measure a reflection of Chip Levy's unrelenting pursuit of neuroscientific theory rooted in biological reality. This work was supported in part by NIH Grant RR15205 for which Professor Levy is the Principal Investigator.

[14] Recursive estimation is the name of the game in sensory signal processing by neurons. A Kalman formalism [21] is appropriate for this, subject to certain provisos. One is that it seems best to adopt the conditional error entropy minimization criterion [22] [23] [24] as opposed to, say, a minimum MSE criterion; this is in keeping with our view that an information criterion is to be preferred for as long as possible before specializing to a more physical criterion associated with a particular use. Another is that the full Kalman solution requires inverting matrices of the form $I + MM^T$ in order to update the conditional covariance matrix. Matrix inversion is not believed to be in the repertoire of mathematical operations readily amenable to realization via neurons unless the effective rank of the matrix M is quite low.

[15] We remark that neurons are remarkably sensitive in this respect. They idle at a mean PSP level that is one or two standard deviations below their spiking threshold. In this mode they spike only occasionally when the random fluctuations in their effectively Poisson synaptic bombardment happen to bunch together in time to build the PSP up above threshold. However, a small percentage change in the bombardment (e.g., a slightly increased overall intensity of bombardment and/or in a shift toward a higher excitatory-to-inhibitory ratio of the synapses being bombarded) can significantly increase the spiking frequency. See [25] [26]. Given the predominantly positive feedback among the members of a coalition, many of its members can be made to ramp up their spiking intensities nearly simultaneously. This helps explain why coalitions of neural cortex exhibit dramatic variations on a time scale of several tens of milliseconds in their rates of spiking and hence in their rates of information transmission and of energy depletion.

References

[1] C. E. Shannon, A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, vol. 27, 379-423, 623-656, July and October, 1948. (Also in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, eds., IEEE Press, Piscataway, NJ, 1993, 5-83.)

[2] C. E. Shannon, Channels with Side Information at the Transmitter. *IBM J. Research and Development*, 2, 289-293, 1958.

[3] C. E. Shannon, An Algebra for Theoretical Genetics, Ph.D. Dissertation, Department of Mathematics, MIT, Cambridge, Massachusetts, April 15, 1940.

[4] C. E. Shannon, Coding Theorems for a Discrete Source with a Fidelity Criterion. *IRE Convention Record, Vol. 7*. 142-163, 1959. (Also in *Information and Decision Processes*. R. E. Machol, ed. McGraw-Hill, Inc. New York, 1960, 93-126, and in *Claude Elwood Shannon: Collected Papers*, N. J. A. Sloane and A. D. Wyner, eds., IEEE Press, Piscataway, NJ, 1993, 325-350.)

[5] H. B. Barlow, Sensory Mechanisms, the Reduction of Redundancy, and Intelligence. In: The Mechanization of Thought Processes. National Physical Laboratory Symposium No. 10, Her Majesty's Stationery Office, London, pp. 537-559, 1958.

[6] D.H. Hubel and T.N. Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex, *J. Physiol.*, **160** 154, 1962.

[7] T. Berger and Y. Ying, Characterizing Optimum (Input,Output) Processes for Finite-State Channels with Feedback, to be presented at ISIT 2003, Yokohama, Japan, June 30 - July 4, 2003.

[8] C. B. Woods and J. H. Krantz, Lecture notes at Lemoyne University based on Chapter 9 of *Human Sensory Perception: A Window into the Brain*, 2001.

[9] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[10] F. Jelinek, *Probabilistic Information Theory*, McGraw-Hill, New York, 1968.

[11] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[13] W. B Levy and R. A Baxter, Energy Efficient Neural Codes. Neural Comp., **8**, 531-543, 1996. (Also in: *Neural Codes and Distributed Representations*, L. Abbott and T. J. Sejnowski, Eds., MIT Press, Cambridge, MA, 105-117, 1999).

[14] W. B. Levy and R. A. Baxter, Energy-Efficient Neuronal Computation Via Quantal Synaptic Failures. *J. Neuroscience*, **22**, 4746-4755, 2002.

[15] H. K. Hartline, H. G. Wagner, and F. Ratliff, Inhibition in the Eyes of *Limulus. J. Gen. Physiol.*, **39**, 651-673.

[16] A. H. Burkhalter, Cortical Circuits for Bottom-Up and Top-Down Processing. Presented in Symposium 3, Cortical Feedback in the Visual System, Paper 3.2, *Neuroscience 2002 - Society for Neuroscience 32nd Annual Meeting*, Orlando, FL, November 2-7, 2002.

[17] J. Bullier, The Role of Feedback Cortical Connections: Spatial and Temporal Aspects. Presented in Symposium 3, Cortical Feedback in the Visual System, Paper 3.3, *Neuroscience 2002 - Society for Neuroscience 32nd Annual Meeting*, Orlando, FL, November 2-7, 2002.

[18] M. Gastpar, B. Rimoldi and M. Vetterli, "To Code or Not to Code: On the Optimality of Symbol-by-Symbol Communication." Technical Report IC/2001/26 EPFL, Lausanne, Switzerland, May 2001.

[19] R. E. Blahut, Computation of Channel Capacity and Rate -Distortion Functions. *IEEE Trans. Inform. Theory*, **18**, 460-473, 1972.

[20] K. Rose, A Mapping Approach to Rate-Distortion Computation and Analysis. *IEEE Trans. Inform. Theory*, **42**, 1939-1952, 1996.

[21] T. Kailath, *Lectures on Wiener and Kalman Filtering*, CISM Courses and Lectures No. 140, Springer-Verlag, Vienna and New York, 1981.

[22] H. L. Weidemann and E. B. Stear, Entropy Analysis of Parameter Estimation. *Info. And Control*, **14**, 493-506, 1969.

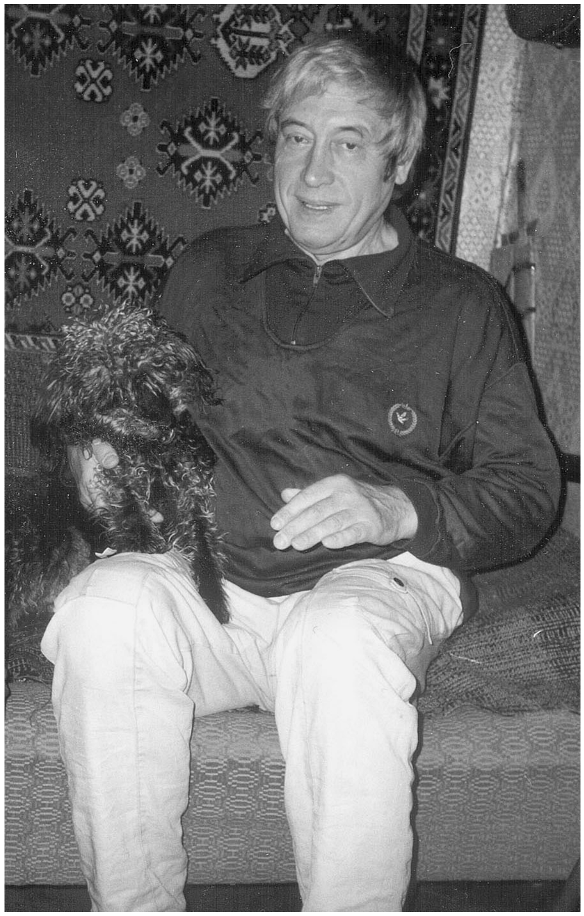
[23] N. Minamide and P. N. Nikiforuk, Conditional Entropy Theorem for Recursive Parameter Estimation and its Application to State Estimation Problems. *Int. J. Systems Sci.*, **24**, 53-63, 1993.

[24] M. Janzura and T. Koski, Minimum Entropy of Error Principle in Estimation. *Info. Sciences*, **79**, 123-144, 1994.

[25] A. Buoncore, V. Giorno, A G. Nobile and L. M. Ricciardi, A Neural Modeling Paradigm in the Presence of Refractoriness. *BioSystems*, **67**, 35-43, 2002.

[26] T. Berger, Interspike Interval Analysis via a PDE. Preprint, Electrical and Computer Engineering Department, Cornell University, Ithaca, New York, 2002.

Valery N. Koshelev (1932 - 2002)



Valery N. Koshelev was born in Moscow, U.S.S.R., on November 11, 1932. He received the Master degree from Moscow State University in 1956 and the "Candidate in physics and mathematics" degree (the U.S.S.R. equivalent of the Ph.D. degree) in 1966.

From 1961 to 1969, he was with the Institute for Problems on Information Transmission, Moscow. Since 1970, he was with the Council for Cybernetics, The Academy of Sciences, Moscow.

The research interests of Valery Koshelev were concentrated on various mathematical problems in probability theory and combinatorics with applications to information and coding theory. He also had a good understanding of practical issues and devoted considerable effort to establishing connections between industry and science. During his work in the Council for Cybernetics, he built a strong group for solving difficult problems related to constructing local computer networks.

The main concern of Valery Koshelev in information theory can possibly be formulated as creating a theory of hierarchical communication schemes. He was one of the pioneers in multi-user information theory who realized the importance of hierarchical structures in information theory for constructing networks. As a result, he introduced the classification of sources of information based on the possibility of attaining the rate-distortion curve in successive steps. He called this property "the divisibility of a source", and this concept became a part of information theory. In particular, it turned out that successive refinement (or divisibility) can fail to hold even when the source letters are equiprobable and the distortion measure is balanced in the sense introduced by Shannon in 1959. An extension of this notion led Valery Koshelev to the classification of channels via the possibility of attaining the capacity-cost curve in successive steps. He called this property "the divisibility of a channel". We are preparing the English translation of his recent monograph and hope that it will appear in the near future. The list of the most important publications of Valery Koshelev is given below.

Valery Koshelev was a sincere and open person who considered all problems of life in a very deep and scientific way. Valery's death was sad news for many people who will miss him very much.

References

[1] V. N. Koshelev, "Direct sequential encoding and decoding for discrete sources," IEEE Trans. Inform. Theory, vol. 17, no. 3, pp. 340-343, 1973.

[2] V. N. Koshelev, "A problem of separate coding of two dependent sources," Probl. Peredachi Informatsii, vol. 13, no. 1, pp. 26-32, 1977. English translation: vol. 13, pp. 18-22, 1977.

[3] V. N. Koshelev, "Hierarchical coding of discrete sources," Probl. Peredachi Informatsii, vol. 16, no. 3, pp. 31-49, 1980. English translation: vol. 16, pp. 186-203, 1980.

[4] V. N. Koshelev, "Estimation of mean error for a discrete successive-approximation scheme," Probl. Peredachi Informatsii, vol. 17, no. 3, pp. 20-33, 1981.
English translation: vol. 17, pp. 161-171, 1980.

[5] V. N. Koshelev, "Computational aspects of the triangle axiom," Cybernetics and Computer Technique, no. 4, Nauka, Moscow, 204-216, 1988.
(In Russian).

[6] V. N. Koshelev, "On the divisibility of discrete sources with an additive single-letter distortion measure," Probl. Peredachi Informatsii, vol. 30, no. 1, pp. 31-50, 1994.
English translation: vol. 30, pp. 27-43, 1994.

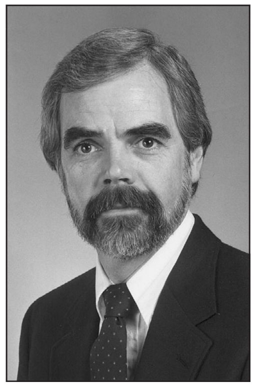
[7] V. N. Koshelev and E. C. van der Meulen, "Two-level coding for input- constraint channels," Proc. ISIT'98, Washington, U.S.A., p. 490, 1998.

Pursley Awarded the IEEE Communications Society Edwin Howard Armstrong Achievement Award

Michael B. Pursley was awarded the 2002 Edwin Howard Armstrong Achievement Award by the IEEE Communication Society. The award is given for "Outstanding contributions over a period of years in the field of the IEEE Communications Society" and consists of a certificate, plaque and an honorarium of $2,000. Dr. Pursley's award reads:

> "For seminal contributions to spread-spectrum communications and adaptive protocols for mobile wireless communication networks."

This award is named in honor of Edwin H. Armstrong, most notably the inventor and father of the complete FM radio system. He is responsible for the Regenerative Circuit, the Superheterodyne Circuit, and the Superregenerative Circuit. His inventions and development form the backbone of Radio Communications as we know it. A contemporary (in writing about his life and achievements) referred to him as the "Man of High Fidelity" and more recently, his life is documented in a book justly called "Empire of the Air."



Michael B. Pursley

Dr. Pursley is currently the Holcombe Professor of Electrical and Computer Engineering at Clemson University, Clemson, South Carolina. His research is in the general area of communications and information theory with emphasis on spread-spectrum communications, communication over fading channels, applications of error-control coding, protocols for packet radio networks, and mobile communications systems and networks.

Dr. Pursley is a member of Phi Eta Sigma, Tau Beta Pi, and the Institute of Mathematical Statistics, and he is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE). He was elected to three-year terms on the Board of Governors of the IEEE Information Theory Society in 1977 and again in 1989. In 1982 he was elected Fellow of the IEEE "for contributions to information theory and spread-spectrum communications." In 1983 he was elected president of the Information Theory Society. Dr. Pursley was a member of the Editorial Board of the Proceedings of the IEEE for the period 1984-1991. He is currently a member of the Editorial Advisory Board for the International Journal of Wireless Information Networks, and he is a Senior Editor of the IEEE Journal of Selected Areas in Communications. He served as Technical Program Chairman for the 1979 IEEE International Symposium on Information Theory which was held in Grignano, Italy, and as Co-Chairman for the 1995 IEEE International Symposium on Information Theory in Whistler, Canada.

Dr. Pursley received Clemson University's McQueen Quattlebaum Faculty Achievement Award in 1995 and Clemson's Board of Trustees Award for Faculty Excellence in 1997 and 2000. He was awarded an IEEE Centennial Medal in 1984, and he is co-recipient (with John M. Shea) of the 1996 Ellersick Award of the IEEE Communications Society for the best paper in the unclassified technical program of the IEEE Military Communications Conference. In 1999 he received the IEEE Military Communications Conference Award for Technical Achievement "for sustained technical contributions to military communications," and he was awarded the IEEE Millennium Medal in 2000. In 2000 he was installed as an honorary member of the Golden Key National Honor Society.

Dr. Pursley is the author of Random Processes in Linear Systems which was published by Prentice Hall in 2002.

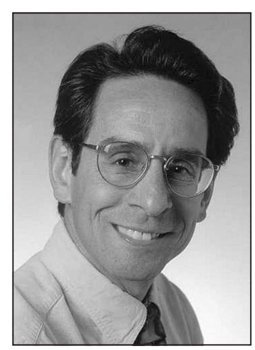
From the Transactions Editor-in-Chief

January 1 marked the midpoint of my term as Editor-in-Chief of the IEEE Transactions on Information Theory. In this report, I want to review the highlights of the first 18 months of my tenure as EiC, discuss several important issues currently facing the Transactions, and offer a few thoughts about the future.

Acknowledgements

The Transactions is surely the pride and joy of our Society. The intellectual excitement that I feel at the arrival of each new issue of our "green magazine" is now augmented by the personal satisfaction of being part of the editorial team that produces it. So, let me begin this report by expressing my sincere gratitude to all of the members of that dedicated team.

Past Editor-in-Chief



I would be remiss if I did not begin by acknowledging my predecessor as Editor-in-Chief of the Transactions, Alex Vardy. During his term, Alex introduced a number of changes that materially improved the Transactions' effectiveness, both operationally and in substance. He also worked very closely with me to ensure continued smooth sailing as I took the editorial helm from him.

Alex restructured the Editorial Board to include the new technical areas of Sequences and Quantum Information Theory, and further adjusted the composition of the board to permit a more balanced assignment of papers to Associate Editors. He ensured that, with few exceptions, the average monthly load per Associate Editor did not exceed 3 papers per month, a practice that I have endeavored to follow.

Alex also spearheaded the effort to "go electronic," establishing electronic submission procedures and encouraging the use of email correspondence among authors, editors and reviewers. These steps, as well as his institution of a zero-backlog policy, have helped to reduce the Transactions' average publication delay. Alex also introduced the practice of including in each issue the biographies of all contributors – authors of correspondence articles and regular papers, alike – and added to each published paper a footnote acknowledging the communicating Associate Editor. Finally, he commissioned the development of a database tool for recording and tracking the review cycle of every paper submitted to the Transactions. (This web database, nicknamed SAGE, is the handiwork of Christopher Barton and Ari Trachtenberg.) Thank you, Alex, for a job truly well done!

Associate Editors

Considering the effort required to maintain the quality and timeliness of our journal, which received approximately 550 papers (excluding special issues) in both 2001 and 2002, it is truly remarkable that a team of only 20 Associate Editors can meet the challenge. All authors and readers of the Transactions owe an enormous debt of gratitude to this dedicated and hard-working group of individuals, who possess, in equal measure, technical expertise, organizational skills, and a commendable spirit of service to the IT Society. Working with the members of this elite group is one of the most rewarding activities associated with the position of Editor-in-Chief.

I was very fortunate to inherit from Alex a talented and responsible group of continuing editors, some of whose terms have expired since I began my editorship. The latter group includes:

Jonathan J. Ashley	Coding Theory
Patrick Solé	Coding Theory
Neal I. Koblitz	Complexity & Cryptography
Gábor Lugosi	Nonparametric Estimation
Peter W. Shor	Quantum Information Theory
Andrew M. Klapper	Sequences
Prakash Narayan	Shannon Theory
Marcelo Weinberger	Source Coding

as well as Richard (Dick) E. Blahut who served as Associate Editor for Book Reviews.

I am grateful to them for their past efforts, as well as for their continued service to the Transactions as they complete the processing of their "active" papers.

Table 1 contains the roster of the current team of Associate Editors, as well as their IEEE Region affiliations and the end dates of their 3-year editorial appointments. I am truly delighted with the outstanding new members of the Editorial Board, whose names are marked with an asterisk in the table.

Table 1: Current members of the Transactions Editorial Board

Associate Editor	*Region*	*Editorial Area*	*End date*
Gérard Battail*	8	At Large	7/1/04
Rüdiger Urbanke	8	Coding Techniques	10/1/03
Jørn Justesen Ralf Koetter Simon Litsyn Khaled Abdel-Ghaffar* Claude Carlet*	8 4 8 6 8	Coding Theory	7/1/03 9/1/03 9/1/03 3/1/05 3/1/05
Giuseppe Caire David N.C. Tse*	8 6	Communications	10/1/03 4/1/04
Leandros Tassiulas*	2	Communication Networks	7/1/04
Thomas Johansson*	8	Complexity and Cryptography	9/1/05
Venugopal V. Veeravalli Aleksandar Kavèiæ*	4 1	Detection and Estimation	1/1/04 7/1/04
Andrew B. Nobel*	3	Nonparametric Estimation	9/1/05
Emanuel H. Knill*	6	Quantum Information Theory	10/1/05
Kenny Paterson*	8	Sequences	10/1/05
Emre Telatar Raymond Yeung*	8 10	Shannon Theory	1/1/04 1/1/06
Ram Zamir Serap Savari*	8 2	Source Coding	7/1/03 7/1/05
Sergio Verdú*	2	Book Reviews	3/1/05

Publications Editors and Staff

The successful production of each issue of the Transactions depends critically upon the attention and care of the Publications Editors, who also hold 3-year appointments. So, it is a real pleasure to thank the two former Publications Editors, Ramesh R. Rao (end date 5/1/02) and Erik Agrell (end date 7/1/02), whose meticulous editing and overall commitment to the Transactions went far beyond the call of volunteer duty. The handoff to their equally dedicated and able successors, Bruce E. Moision (end date 1/1/04) and Kevin Quirk (end date 7/1/05), has gone without the slightest hitch, by my reckoning.

I am also indebted to our Publications Coordinator, Katherine Perry, who has managed the office of the Editor-in-Chief here at UC San Diego since September 2000, first assisting Alex Vardy, and now me. Every author and Associate Editor has benefited from Katherine's administrative skills, cooperative spirit, and gracious manner.

Finally, I wish to express my very sincere thanks to IEEE Senior Editor Nela Rybowicz, who, for the past eight years, has labored diligently and tirelessly to ensure that every published issue of the Transactions meets the highest standards of accuracy and quality. Her devotion to our journal is worthy of our deepest respect and gratitude.

IEEE TAB 5-year Transactions review

In 2001, for the first time, the Periodicals Committee of the IEEE Technical Activities Board (TAB) conducted a 5-year review of the Transactions (and Newsletter) in parallel with the corresponding review of the IT Society.

The primary objectives of TAB periodical reviews are to examine the timeliness and quality of IEEE TAB publications, provide suggestions for improvement, and share best practices with other IEEE Societies and Councils.

As Editor-in-Chief, my responsibility was to complete the questionnaire provided by TAB, discuss the Transactions with the Review Committee during the TAB series meeting (held in Mexico City, November 2001), examine and comment upon the Committee Report (which incorporated the completed questionnaire) before its submission to TAB, and, finally, respond formally to the Committee's comments and recommendations as presented in their report. The results of this process – the full Committee Report and my response – are available at the new IT Society website.

It is my belief that anyone who reads the report will recognize what we all know quite well – that the Transactions is a shining jewel among IEEE publications. And even though the citation statistics, which were requested as part of the questionnaire, cannot measure information theory's intrinsic beauty or the sweep of its transforming effect upon modern society, they can confirm the relevance of the Transactions to its readership and its stature among scientific publications.

Journal Citation Reports©, the source of the data, uses three quantities to evaluate a publication. They are defined as follows.

The **Cited Half-Life** is the number of publication years from the current year that account for 50% of current citations received.

The **Immediacy Index** measures how quickly the average article in a journal is cited. The immediacy index tells how often articles published in a journal are cited within the same year. It is calculated by dividing the number of citations to articles published in a given year by the number of articles published in that same year.

The **Impact Factor** measures how frequently an average article in a journal has been cited in a particular year. It is calculated by dividing the number of current citations to items published in the two previous years by the total number of articles and reviews published in the two previous years.

JCR© ranks periodicals within subject categories according to the Impact Factor; for the Transactions, the category is Electrical and Electronic Engineering.

The Transactions' figures for the years 1996-2001 are shown in Table 2 below.

Table 2: IT Transactions citation indices for the years 1996 - 2001.

Year	*Cited Half- Life*	*Immediacy Index*	*Impact Factor*	*Rank/Total (Impact Factor)*
2001	9.8	0.495	2.077	12/199
2000	>10.0	0.363	1.825	13/202
1999	9.7	0.352	2.009	11/205
1998	9.9	0.521	2.083	3/205
1997	9.8	0.213	1.354	17/192
1996	9.9	0.359	1.698	7/172

For purposes of comparison, I'll mention that the Impact Factor rankings of the IEEE Journal on Selected Areas in Communications and the IEEE Transactions on Communications during the 1996 - 2001 period were: 15, 19, 9, 10, 15, 19 and 52, 57, 37, 46, 28, 25, respectively.

There is an interesting story behind the Impact Factor figure for the year 2000. The published JCR© indicated 5944 citations in 2000 and an Impact Factor of 0.654, corresponding to the rank 77/202 - substantially lower than in prior years. At our request, JCR© looked into this aberration and quickly discovered that the data used to compute the 2000 Impact Factor were incomplete. The adjusted figure, based upon 6715 citations, restored the Transactions to a more appropriate rank of 13/202.

As was to be expected, the Committee's assessment of the Transactions was extremely positive:

In general this Transaction is well managed, financially sound, and meets the quality standards set by IEEE. This is amply demonstrated by the impact factor (1.825) and ranking (13 out of 202 electrical engineering journals).

We found that a good practice in the peer review process was to select three reviewers who agreed to review the papers in a specified time period of typically six weeks. We also think the three-year staggered terms of office for all Editorial Board members to be a good practice. The "no backlog" policy is commendable and a practice we would like to see other journals employ.

Two other exemplary practices we would like to comment on are:

1. The addition of expository papers that may be longer in length to fully cover a topic and the fact that one of your expository issues became a book.
2. The practice of taking a broad view of IT allowing you to be flexible in your inclusions to be able to capture newly emerging trends and technical directions.

The Committee's evaluation of the Newsletter was equally favorable.

The newsletter issued four times a year appears to be well written and covers the span of topical news items to more magazine like expository articles. It meets all the guidelines for copyright, timeliness and quality.

(Thanks go to Lance Pérez for his stewardship of the IT Society Newsletter, particularly for the restoration of its regular publication schedule and for the consistently interesting material that he assembles for each issue.)

The Committee also made several recommendations. They encouraged continued efforts to reduce the average time from submission to publication, with an ambitious target of nine months. It was also felt that the Transactions should adopt a more automated manuscript tracking and data reporting system, such as ScholarOne's Management Central™, a web-based tool used by several IEEE journals. Finally, the Committee suggested that the Society consider the imposition of overlength page charges in response to the continued increase in the average length of both regular papers and correspondence articles. As will be discussed later in this report, the Editorial Board is taking steps to address these recommendations.

Recent Changes

During the past 18 months, the Transactions has seen several significant changes. Perhaps the most obvious was the move to a monthly publication schedule, beginning in January 2002. I am very happy to report that the anticipated benefits of this change - please refer to my editorial note in the January 2002 issue - have indeed been realized.

The introduction of web-based posting of galley proofs in PDF format is another change for the better, eliminating the expense and delay associated with express mailing of hard copies.

In addition, during the past year the IEEE took several actions with respect to all of its publications that affected the "look and feel" of the Transactions. You may have noticed the new Transactions cover design that had its debut in January 2002. The IEEE "Master Brand," comprising the "right-hand-rule" logo and bold-font initials "IEEE," was relocated from the northwest corner to the southeast corner, and the IT Society logo was shifted from a position to the right of the journal title to a location above the issue date. A more subtle change was the reduced size of our Society logo relative to the IEEE Master Brand.

As Editor-in-Chief, I was asked by the IEEE for comments on the new cover design in late November 2001 - only one month before its planned introduction. I voiced several concerns involving aesthetics and poor use of cover "real-estate," as well as my opinion that, in view of IEEE's recent fiscal actions, this seemed like an inopportune time for IEEE to be, apparently, asserting its dominance graphically. My criticisms were seconded by Joachim Hagenauer, our Society President at the time.

The response we received stated that "the modifications in progress for all Transactions/Journals were undertaken in order to comply with the new IEEE Identity Standards, the text of which can be found at http://www.ieee.org/about/documentation/IDStandards.pdf.

The standards call for a specific Master Brand minimum size and discuss the prominence of the IEEE Master Brand in relation to other logos on a cover. The IT logo was purposely made smaller so that the IEEE Master Brand could be more prominent. This in no way diminishes the IT Society, but serves to strengthen the identity of the IEEE as a whole."

We were also told that any proposal to revisit the new cover design - that apparently had been approved by an earlier vote of IEEE Society Presidents - would lead to severe publication delays, a situation that we considered unacceptable. Fortunately, the Transactions' move to a monthly publication schedule was perfectly timed, and my concerns about the new design's encroachment upon precious Table of Contents space proved to be unwarranted.

The January 2003 issue reflects two further changes mandated by the IEEE. The "trim size" of the journal - meaning the actual page dimensions - has been cropped from the historical 8 ¼" x 11" aspect to a slimmer 8" x 10". Perhaps less perceptible is the reduced weight of the paper, from 40- to 36-pound stock. Both of these alterations were part of a suite of cost-saving measures adopted by the IEEE on behalf of its Societies. The net savings to the IT Society in publishing costs are, in fact, quite nominal, and our Society, along with a number of others, asked to be exempted from the trim-size reduction. There were several reasons, not the least being the effect of the reduction on the width of the "gutter" - the page margin adjacent to the binding. After consideration of our objections, however, IEEE decided to proceed with the reduction for all journals.

Another cost-saving measure, which was voluntarily adopted by the IT Society, proposed the use of a remailer (rather than the U.S. Post Office) to reduce expenses associated with delivery of print copies of the Transactions to Regions 8 through 10. This change would benefit most journals financially; moreover, delivery times were expected to be reduced from 30-60 days to 15-20 days. However, the financial savings would be realized only when the weight of the journal issue fell below a threshold in the range of 18-20 ounces. Ironically, even with the reduction in weight stemming from the move to monthly publication, the IEEE estimated that the Transactions would incur a financial penalty of approximately $2900 per year due to the excess weight of a typical monthly issue. Underscoring one of our Society's top priorities, namely service to the Transactions readership, the Board of Governors admirably voted to use the remailer in order to expedite delivery to Regions 8 through 10, despite the possibility of additional mailing costs.

Highlights

Special issues:
The June 2002 issue, entitled "Shannon Theory: Perspective, Trends, and Applications," was masterfully edited by Henry Landau, Jim Mazo, Shlomo Shamai, and Jacob Ziv.

Dedicated to the memory of Dr. Aaron Wyner, the issue contained a fine collection of papers that served to emphasize the enormous legacy that Dr. Wyner bequeathed to us. A commemorative plaque featuring the cover of the issue was presented by Shlomo Shamai and Jacob Ziv to Dr. Wyner's wife, Nusha Wyner, at the ISIT 2002 banquet in Lausanne.

There are two special issues in the works. Bertrand Hochwald, Thomas Marzetta, Babak Hassibi, and Giuseppe Caire are the guest editors for an issue on "Space-Time Transmission, Reception, Coding and Signal Design." The initial call for papers appeared in the June 2002 Transactions and Newsletter, and the issue closed on October 30, 2002 with - brace yourselves - 72 submissions! The target publication date is October 2003.

The next special issue in the queue, entitled "Problems on Sequences: Information Theory and Computer Science Interface," is being guest edited by John Kieffer, Wojciech Szpankowski, and Eh-hui Yang. The call for papers first appeared in November 2002, and the submission deadline is March 15, 2003. The issue is scheduled to appear in mid-2004. I expect that it will further expand the interplay of ideas between these two disciplines, from whose interface a number of exciting results in complexity theory and coding theory have recently emerged.

Paper Awards:
After the Claude E. Shannon Award, the IT Society Paper Award is the highest form of technical recognition bestowed by the IT Society. The Paper Awards for 2001 and 2002 were given to three papers whose wide-ranging influence was perceived well before their official publication.

The 2001 prize paper entitled " Capacity of Multi-Antenna Gaussian Channels," by Emre Telatar, was published in the European Transactions on Telecommunications in 1999. It was among those pioneering studies that triggered the explosion of research in space-time techniques, the impact of which is clearly seen in the number of submissions to the planned 2003 special issue.

Similarly, the two Transactions papers sharing the 2002 award - "The Capacity of Low-Density Parity Check Codes Under Message-Passing Decoding," by Thomas J. Richardson and Rüdiger L. Urbanke, and "Improved Low-Density Parity Check Codes Using Irregular Graphs," by Michael G. Luby, Michael Mitzenmacher, M. Amin Shokrollahi, and Daniel A. Spielman - have already inspired innumerable research efforts within the coding community.

It is interesting to note that the Swiss Federal Institute of Technology Lausanne can now boast of having *three* faculty members who authored or co-authored *three* papers that won the IT Society Paper Awards in *two* consecutive years. By the same token, Bell Laboratories can be congratulated for having supported the related research of *four* such individuals while they were in its employ. These will be tough records to beat!

In 2001, a new ComSoc/ITSoc Joint Paper Award was introduced. The 2001 award was shared by two papers, "Linear Multiuser Receivers: Effective Interference, Effective Bandwidth, and User Capacity," by David N.C. Tse and Stephen Hanley, and "Iterative (Turbo) Soft Interference Cancellation and Decoding for Coded CDMA," by Xiaodong Wang and H. Vincent Poor. These appeared in the March 1999 Transactions and the July 1999 Transactions on Communications, respectively. The 2002 award went to Michael Honig and Weimin Xioa for "Performance of Reduced-Rank Linear Interference Suppresion," which appeared in the July 2001 Transactions.

Another Transactions paper, "The Effect Upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," by Muriel Médard, received the very prestigious 2002 IEEE Leon K. Kirchmayer Prize Paper Award, for the most outstanding paper by an author under the age of 30 years, at the date of submission, published in any IEEE publication during the preceding calendar year.

Congratulations to all of the authors of these outstanding papers.

Shannon Retrospective:
The November 2001 issue featured an invited paper by Robert M. Gallager, entitled "Claude E. Shannon: A Retrospective on His Life, Work, and Impact." This wonderful contribution, written by a true giant of information theory and appearing in the same year as Shannon's death, allowed

the Transactions to pay special tribute to the man whose genius and humanity continue to inspire all those who work in our field. The care, effort, insight, and sense of urgency that Prof. Gallager brought to the writing of this article was, and continues to be, greatly appreciated.

Looking to the Future

The future of the Transactions promises to be as bright as its past. However, like many fine institutions, the Transactions, including the complex process that produces each issue, represents a work in progress. I, along with members of the Editorial Board and the Society Board of Governors, will be carefully examining several matters that bear upon the future operations of the Transactions. Here are my thoughts on some of them.

Editorial Loads and Terms: It is not proven that there is a direct relationship between Associate Editorial loads and publication delays. However, the comments I receive from even the most diligent Associate Editors suggest that there may be a link. In any case, if there is a sense that Associate Editors would be more effective handling fewer papers per month and/or having a shorter term, it behooves the Society to consider these options.

Reducing the load would imply expansion of the Editorial Board. With this in mind, I am considering the addition of new Associate Editors in two "high-volume" areas, specifically Coding Techniques and Communications. As far back as Alex Vardy's "state-of-the-Transactions" report in June 2000, it was clear that increasing attention was being paid to research topics falling into these two categories, particularly iterative decoding of codes on graphs and space-time signal processing and coding. (Not coincidentally, the most recent IT Society Paper Awards represent seminal works relating to these topics.)

I am also considering the option of appointing Associate Editors to 2-year terms. In view of the fact that new papers are assigned to Associate Editors throughout their "official" terms, and recognizing that review cycles can last for more than a year, this still translates into a "real" term of 3 to 4 years.

Transactions Website: The time has come for the Transactions to establish a flexible, user-friendly, and comprehensive web-based system for managing manuscript submissions and tracking the editorial review and decision process. Our current web database, SAGE, has many attractive features and is an invaluable aid to the editorial team. However, in order to extend the scope of the system to include the submission process and to more fully automate the tracking of papers throughout the review cycle, additional capabilities are required. Katherine Perry has experimented with Manuscript Central™ on a trial basis, but has found it to be somewhat rigid (and expensive). On the other hand, Publications Editor Kevin Quirk has been designing a software package, using widely accepted web-authoring tools and database design techniques, that incorporates the features of SAGE into a system that could more fully meet the future needs of the Transactions. We will be further evaluating these and possibly other candidate systems in the coming months.

Overlength page charges: As mentioned above, the average size of Transactions articles is following the same trend as the average size of American waistlines. The average lengths of regular papers and correspondence articles published in 2001 were 14.9 pages and 5.5 pages, respectively, and in 2002 these figures were 14.7 and 5.8, respectively. The number of pages published in 2000 was 2880; it grew to 3208 in 2001, and then to 3280 in 2002. I am a firm opponent of page limitations, but it is conceivable that lengthy papers take longer to review, and they certainly cost more to publish. Overlength page charges can be considered as one mechanism to encourage more concise exposition, shorten the review cycle, and help defray the additional expense of publication, without imposing rigid constraints on paper length. This is a delicate issue, requiring careful consideration by the Board of Governors.

Alternative Publishing: Several factors have driven the IT Society to explore publication alternatives for the Transactions. One is the very real prospect that the IEEE's Society infrastructure tax, along with IEEE's increased shares of Society investment and intellectual property package income, could bankrupt the IT Society within a few years. In that scenario, the Transactions will need to identify an alternative to IEEE Publishing Services. Even if this does not come to pass, we may want to follow the lead of some IEEE Societies that have found other avenues to publication. Several organizations are promoting new paths for scholarly communication, such as SPARC, the Scholarly Publishing and Academic Resources Coalition, which publishes the IEEE Sensors Journal. Also, a number of venues for on-line peer-reviewed journals have recently been established. Parting ways with IEEE Publications is certainly a disturbing prospect, but the Society must be prepared for that possible eventuality.

Concluding Remarks

The position of Editor-in-Chief of the Transactions is time-consuming and demanding, but it is also enormously interesting and gratifying. During the second half of my term, I will continue striving to maintain and, where possible, improve upon the tradition of excellence that has been the hallmark of the Transactions throughout its history. And I welcome your suggestions on how to more effectively achieve that goal.

The talented and diverse members of the information theory community, with whom I interact daily in my editorial capacity, are bound together by a shared passion for the conceptual elegance and practical impact of our discipline, in all of its manifestations. I feel privileged to belong to that community, and I am grateful for having this opportunity to serve it.

Historian's Column

A. Ephremides

A fundamental necessary condition for a historian to earn respect (and self-respect, at that) is to have the facts right. Failing that, the next best thing to do for restitution and redemption is to admit the errors and to reconstruct the truth (without necessarily … self-flagellating!).

I am in the sad position to admit that my facts were not entirely correct in my column of two issues ago where I made reference to the Swedish-USSR workshops. I spoke the truth, as I (mostly) do, but not the whole truth. So, I am taking the opportunity to make amends for humbling public retribution.

A. Ephremides

I said in that column that the first Swedish-USSR workshop took place in Gränna, Sweden, in 1985 and the last one in Moscow in 1991. Fortunately, there are alert readers out there and, on this occasion, Rolf Johannesson raised the flag. In fact, the first workshop, he points out, took place in 1983 in Sochi on the Black Sea. It was followed by the one in Gränna in 1985 and then by one again in Sochi in 1987, which was followed by the Gotland one in 1989, the Moscow one in 1991, another one in Sweden (in Molle) in 1993, and, finally by a workshop in St. Petersburg in 1995, which was the last one of that remarkable series of seven legendary meetings. What makes my earlier report unbelievably ridiculous and … shameful is that I attended the one in Molle in 1993 and managed to, how shall I put it, … forget about it! In full realization that such a bangled-up report of truth may have shaken my credibility irreparably I proceed, nonetheless, to add some more details about the Molle workshop. It took place in the small village by the sea in a charming hotel that was run by an elderly gentleman (who sadly has passed away) who led the participants on a typical Scandinavian summer evening to a forest walk that culminated in a cheerful salmon cookout by the shore. It was in Molle that I discovered a great Scandinavian delicacy called by the Swedes "lojrom". It is the roe of a fish from the local waters and is comparable in flavor and delicateness to fine caviars at a fraction of the cost. Before leaving Molle I made sure to pack two pounds of it in my bag.

I also recall Levenshtein discovering a beautiful gigantic mushroom in the woods that he then prepared with onions in a frying pan. By the way, some of us went back to Molle for the 1997 ISIT Technical Program Committee meeting in December of 1996. The howling winds and pounding rain contrasted starkly with the long days of August of 1993. We spent all the time indoors and one evening, while sipping brandy by the fireplace, I heard for the first time Jim Massey recite the famous poem "Casey at the Bat". As many of you recall, it was this rendition of the poem that inspired me to compose a song that I … "inflicted" upon you at the closing farewell party of the 1998 Jubilee ISIT in Cambridge. So much for forgetting I was in Molle!

Well, Rolf went on to reveal additional interesting details and highlights from the Swedish-USSR workshops. He noted that attendance of the Gränna workshop by Sergei Gelfand and of the Gotland workshop by Mark Pinsker marked the start (or the "re-start" for the latter) of their travels outside the Iron Curtain. He also reminded us that in Sochi in 1987 a basketball game took place between the members of IPPI (the Soviet Institute for Problems of Information Transmission) and the "rest-of-the-world". The "rest-of-the-world" was ahead by one point moments before the end of the game. The IPPI team demanded then that the game be prolonged by an additional (third) "half" (it was still the time that the game was structured in two "halves"). The demand was granted but no one scored and IPPI remained narrowly defeated.

Rolf also reminded us that in 1991 in Moscow, when a jubilant "pre-revolutionary" atmosphere prevailed (despite the darkness of the Moscow January weather), Nikita Vvedenskaya had the courage to stand at the podium and recognize the brutality of the Soviet regime in the Baltic States at the time and to ask for a moment of silence in memory of those killed during those confrontations. She was subsequently spotted on television near the tank on which Boris Yeltsin stood to halt symbolically the attempted coup against Gorbachev. What momentous times and what brave people indeed! Well, I hope that this helps in setting the record straight and in … repairing my shattered credentials! I thank Rolf for prompting this and I hereby bestow upon him the coveted title of … "historian par-excellence", which only two other people have earned so far (Toby Berger and Jim Massey), while some more individuals are standing in the wings. Rolf pointed out also that Valery Koshelev, whom I mentioned in my earlier column, passed away suddenly last August. We will miss him as we mentally lower the IT flag at half-mast.

The events I recounted constitute a small sample of the drama and the human interactions that are prompted by tensions created by global conflicts such as the Cold War. I am sure there will be more to come.

GOLOMB'S PUZZLE COLUMN™

FACTS ABOUT $\binom{2n}{n}$

Solomon W. Golomb

This time we will look at properties of the "central binomial coefficient", $\binom{2n}{n} = \frac{(2n)!}{(n!)^2}$. (Most of these are well-known.)

1. *Prove these approximations:*

 a. $2^n < \binom{2n}{n} < 4^n$, all $n > 1$.

 b. $\binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} 4^n$ as $n \to \infty$.

2. *Prove these identities, for all $n \geq 1$.*

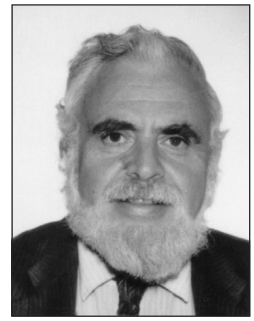
 a. $\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$.

 b. $\binom{2n}{n} = (-1)^n \sum_{k=0}^{2n} (-1)^k \binom{2n}{k}^2$.

 c. $\binom{2n}{n} = 2(-1)^n \sum_{k=1}^{2n-1} (-1)^k \binom{2n-1}{k}\binom{2n-1}{k-1}$.

 d. $\binom{2n}{n} = \prod_{j=1}^{n} \left(4 - \frac{2}{j}\right)$

3. *Prove that each of the following must be an integer, for all $n \geq 1$.*



 a. $\frac{1}{n+1}\binom{2n}{n}$.

 b. $\frac{1}{p}\binom{2n}{n}$, for each prime p, n, $p \leq 2n$.

 c. $\frac{1}{R}\binom{2n}{n}$, where $R = 2 \prod_{n<p_j \leq 2n} p_j$, where p_j runs through all primes in $(n,2n]$.

4. *Let $L(n)$ = l.c.m. $\{1,2,3, \ldots, n\}$, let $L(x) = L(\lfloor x \rfloor)$ for real x, and set $L(x) = 1$ for $0 < x < 2$.*

 a. Show that $\binom{2n}{n}$ divides $L(2n)$.

 b. Show that $\binom{2n}{n} = \prod_{k=0}^{n} \left\{ L\left(\frac{2n}{k}\right) \right\}^{(-1)^k}$, for all $n \geq 1$.

Distinguished Lectures by new IEEE Fellows in Hong Kong

Torleiv Kløve (currently visiting professor at the Hong Kong University of Science and Technology) and Raymond Yeung, both IT members, are among the 10 newly elected IEEE Fellows in Hong Kong this year.

A half-day event was organized at the Hong Kong University of Science and Technology on December 12, which included two distinguished lectures and a dinner. The title of Torleiv Kløve's lecture was *Error-Correction Capability of Codes Beyond Half of the Minimum Distance*, and that of Raymond Yeung's lecture was *Entropy, Information Inequalities, and Groups*. The lectures were followed by a farewell dinner for Torleiv and his wife who are returning to Norway in spring 2003.

Photo taken in front of the HKUST Library, (from left to right) Wai Ho Mow, Raymond Yeung, Torleiv Kløve, Marit Kløve, Mordecai Golin, Eliza Wing-Yee Lee, Weijuan Shan, Wende Chen.

New Senior Members

The following Information Theory Society members were elected to the rank of Senior Member in the IEEE in 2002.

Erik Agrell
Lisa Anneberg
Helmut Boelcskei
Angel M. Bravo
Jiangnan J. Chen
Marco Chiani
Jinho Choi
Wu Chou
Laurie K. Cihlar
Michael P. Clark
Iain Collins
Josep Rifa Coma
Carlos Enrique D'Attellis
Rolf Ernst
Robin John Evans
Josep Domingo-Ferrer
Michael P. Fitz
James E. Fowler
Andrew M. Fraser
Saul B. Gelfand
Anders Host-Madsen
Nils Holte
Johannes B. Huber
Ragnar Hlynur Jonsson
Ryuki Kohno
Jozef Korbicz
Teng-Joon Lim
Angel Lozano
Zhi-Quan Luo
Ranjan K. Mallik
Vladimir Marbukh
Gianluca Mazzini
Muriel Medard
Thomas Mittelholzer
Dharmendra S. Modha
Carl R. Nassar
Daniel L. Noneaker
Geoffrey Orsak
Patric Ostergard
Mohit Kishore Prasad
Thomas J. Richardson
Tony J. Rouphael
Brian M. Sadler
John S. Sadowsky
Eugenio Sansosti
Yoichi Sato
Akbar M. Sayeed
Gianluca Setti
Hussein Sherief
Leif Sornmo
Vahid Tarokh
Alexander Tartakovsky
Shahrokh Valaee
Bane Vasic
Tan F. Wong
Zixiang Xiong
Lie-Liang Yang
Yari Zafrir
Xiao-Ping Zhang

CALL FOR NOMINATIONS

IEEE Medals, Service Awards, and Prize Papers

IEEE has many awards, ranging from prizes for technical achievement to recognition of service to IEEE. The Information Theory Society has many distinguished members who would be strong candidates for IEEE awards. In the past, when the Society has submitted completed nominations, they have been very successful in winning. Your help is needed to identify candidates and, equally importantly, help us to find people who know the candidates and their work, so that nomination forms can be completed in a substantial way.

All of the awards have a NOMINATION DEADLINE of JULY 1, 2003. We strongly encourage suggestions and or nominations, which can be directed to Hideki Imai at imai@iis.u-tokyo.ac.jp

More information can be found on the Web at http://www.ieee.org/awards/.

IEEE Information Theory Society Paper Award

The Information Theory Society Paper Award shall be given annually for an outstanding publication in the fields of interest to the Society appearing anywhere during the preceding two calendar years (2001-2002).

The purpose of this Award is to recognize exceptional publications in the field and to stimulate interest in and encourage contributions to fields of interest of the Society. The award consists of an appropriately worded certificate(s) and an honorarium of $1000 for single author papers or $2000 split equally among the authors of multiply authored papers.

Nomination Procedure (from the bylaws):

The Awards Subcommittee shall take into account

(a) all nominations submitted in response to the open call for nominations in the last two years;

(b) the nominations supplied by the Publications Committee in the last two years;

(c) any nomination that its members may want to submit for consideration.

The Awards Subcommittee shall submit to the Board a list of up to three selected nominations for the Information Theory Society Paper Award at least 3 weeks in advance of the first Board meeting following June 1st of the award year, and shall enclose a rationale for each nominated paper explaining its contribution to the field.

The Board shall then vote for the nominees by ballot, conducted by the Society President or designee at the first Board Meeting following June 1st of the award year. The paper receiving the highest total number of votes in the balloting shall be declared the winner of the Information Theory Society Paper Award.

Please send a brief rationale (limited to 300 words) for each nominated paper explaining its contribution to the field by May 2, 2003 to the Society's First Vice President: Professor Hideki Imai via e-mail (imai@iis.u-tokyo.ac.jp) or by post addressed as: Hideki Imai, Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan.

GOLOMB'S PUZZLE COLUMN™

Early Bird Numbers — Solutions

1. The 45 two-digit *early bird numbers* (e.b. nos.) can be described as follows. Let $n = ab$ (where the standard decimal notation ab stands for $10 \cdot a+b$). If $0<b<a<9$, then $ba<ab$, and the sequence (∗) of all positive integers in natural order will contain $(ba)(ba+1)$ and will exhibit ab in the overlap. (E.g. if n=53=ab, then we see n in the overlap of (35)(36).) There are $\binom{8}{2} = 28$ numbers of this type.

   Next, if n=ab where b=a+1, and $0<a<9$, then we see n early in the sequence (∗) where the single-digit number a is followed by a+1. (Thus, n=23 occurs in 123456….) There are 8 such values of n. Finally, the 9 numbers from 91 to 99 appear in the overlaps of 9-10, 19-20, 29-30, …, 89-90. Altogether, this gives 28+8+9=45 two-digit e.b. nos., exactly half of the numbers from 10 through 99. (None of the others are e.b. nos.)

2. If n is a k-digit positive integer (k>1) such that there is another number n' consisting of a cyclic permutation of the digits of n, with $n'<n$, and the left-most digit of n' being from 1 to 9 inclusive, and the right-most digit of n' is other than 9, then n is an e.b. no. because it appears in the overlap of the consecutive integers n' and n'+1. (For example, if n=215, we may take n'=152, and then in (n') $(n'+1)$ we see 152-153 with the original n in the overlap.)

3. If, in the previous problem, there is a cyclic permutation n' of the digits of n, with $n'<n$, but where the right-most digit of n' is 9, the conclusion that n is an e.b. no. is still true, but the proof is more complicated. Here are the typical situations.

   a. If n=291, we take n'=129 and see n in the overlap of n' and n'+1: (129)(130), as in the previous solution.

   b. If n=9193, we cannot use n'=1939, since the overlap of n' and n'+1, (1939)(1940), does not contain n in its overlap. However, we *can* use n'' = 3919, since now $(n'')(n'' + 1)$ = (3919)(3920) has n in its overlap, and we still have $n''<n$.

   c. If n=919, we cannot use n' =199, since $(n')(n'+1)$ = (199)(200) does not have n as an overlap. However, n already appears in the overlap of (91)(92).

   d. If n=9199, we cannot use n'=1999; but n already occurs in the overlap of (919)(920).

4. a. We already saw that every integer from 91 to 99 inclusive is an e.b. no. in the solution to problem 1.

   b. For the numbers from 901 to 999, problems 2 and 3 show that all are e.b. nos. with the possible exceptions of 909 and 999. (The others have cyclic permutations $n'<n$ with the required characteristics.) But 909 appears in the overlap of (90)(91); and 999 is found in the overlap of (899)(900).

   c. The generalization to all numbers from $9 \cdot 10^d + 1$ to $10^{d+1} - 1$ (inclusive) being e.b. nos. for all is false. As counter-examples, consider n=9090, n=900900, and more generally, $n=0.9(10^{2c}+10^c)$ for all $c \geq 2$. None of these is an e.b. no.

5. The 5-digit number n=11121 is an e.b. number for (at least) the following six overlap representations: a. (11)(12)(13), b. (111)(112)(113), c. (1112)(1113), d. (1211)(1212), e. (2111)(2112), f. (11112)(11113).

6. It is true that, asymptotically, 100% of all positive integers are e.b. nos. That is, if $e(x)$ denotes the number of e.b. nos. ≤ x, then $\lim_{x\to\infty} \frac{e(x)}{x} = 1$ (There are infinitely many non-e.b. nos. also, but they become increasingly infrequent.)

   To see this, observe that the "typical" positive integer has a huge number of digits. (Paradoxically, although any specific integer has a finite number of digits, the *expected* number of digits in a "random" integer is infinite!) With so many digits in the typical integer n, it is overwhelmingly likely that there is a cyclic permutation of these digits satisfying the sufficient condition in Problem 2 (or Problem 3) for n to be an e.b. no.

   An open question is to determine how many of the 10^k-10^{k-1} k-digit integers are e.b. nos., for each k. (For k=1 it is 0 for 9, and for k=2 it is 45 out of 90.) It is very likely to be easier to answer this question if we only count those e.b. nos. n that appear in the overlap of *two* consecutive integers less than n.

CALL FOR PAPERS

Problems on Sequences: Information Theory & Computer Science Interface

2004 Special Issue of IEEE Transactions on Information Theory

The IEEE Transactions on Information Theory will publish a Special Issue on "*Problems on Sequences: Information Theory and Computer Science Interface*".

Recent years have seen a proliferation of research in "Problems on Sequences" which has benefited from the interplay between information theory and computer science – each of these two fields has had an impact upon the other in providing design paradigms and in providing ways of obtaining performance bounds. Because of the continued expansion of this research interest, it is an opportune time for a special issue to be devoted to this area.

Below are examples of topics illustrating some of the ways in which information theory (IT) and computer science (CS) impact each other.

- Analysis of algorithms applied to analyze algorithms for data compression, prediction, pseudorandom number generation, and classification.
- Computational or descriptive complexity applied to examine the complexity/performance tradeoff in lossless/lossy data compression.
- CS data structures (grammars, trees, and graphs) applied to design codes for sequences.
- Exact and approximate pattern matching techniques applied to context modeling/coding.
- Entropy, mutual information and information distance concepts applied to problems of computational biology.
- IT concepts applied to provide bounds in computational learning theory.
- Applications of IT to computational linguistics.
- Applications of IT to information retrieval (especially data mining).

The litmus test regarding the suitability of a paper submitted to the special issue will be whether it addresses a problem on sequences in a manner in which IT and CS interface in an essential way. It is expected that the paper submissions to the special issue will predominately concern the compression/prediction/modeling subarea of IT and the theory of algorithms subarea of CS. The list of topics above reflects that expectation, but should not be considered to be all-inclusive: there could be paper submissions meeting the litmus test which address other subareas of IT and CS, or which address new problems at the interface.

Prospective authors are invited to submit their papers electronically as postscript files to any of the following guest editors. All submissions should follow the guidelines of the IT Transactions as to format, and will undergo a rigorous review. The deadline for submission is **March 15, 2003**.

John Kieffer	Wojciech Szpankowski	En-hui Yang
Dept. of Electrical&Comp. Engr.	Dept. of Computer Science	Dept. of Electrical&Comp. Engr.
University of Minnesota	Purdue University	University of Waterloo
Minneapolis, MN 55455	W. Lafayette, IN 47907	Waterloo, Ontario N2L 3G1
U.S.A.	U.S.A.	CANADA
kieffer@ece.umn.edu	spa@cs.purdue.edu	E.Yang@ece.uwaterloo.ca

Call for Papers

2003 IEEE Information Theory Workshop

Hong Kong, July 6 - 10, 2003

General Co-Chairs

Victor K. Wei
kwwei@ie.cuhk.edu.hk

Raymond W. Yeung
whyeung@ie.cuhk.edu.hk

Technical Co-Chairs

Wai Ho Mow
w.mow@ieee.org

Li Ping
eeliping@cityu.edu.hk

Local Arrangements

Tat Ming Lok
tmlok@ie.cuhk.edu.hk

Cunsheng Ding
cding@cs.ust.hk

Treasurer

C. W. Kok
eekok@ust.hk

Publicity

Mordecai Golin
golin@cs.ust.hk

Publication

Morris M. Z. Wang
enmzwang@polyu.edu.hk

International Liaison

Zhen Zhang
zzhang@milly.usc.edu

Workshop Announcement

The 2003 IEEE Information Theory Workshop will be held at the New World Renaissance Hotel, Hong Kong, China, from July 6 (Sunday) through July 10 (Thursday), 2003. Detailed information including submission guidelines, contact links, technical program, registration, travel, accommodation, getting around, and social events will be available at the workshop web site: http://itwhk03.cs.ust.hk/

Program Information

The workshop has a three and half-day technical program featuring plenary talks, as well as invited and contributed paper presentations. These sessions will be organized around the topics below:

- Shannon theory and statistics
- Modulation and coding
- Communication systems and networking
- Sequences and cryptography
- Source coding, data compression and signal processing
- Quantum information theory
- Wireless systems, multiuser detection and space-time processing

Paper Submission and Important Dates

Papers presenting new results in the above areas are hereby solicited. Only electronic submissions sent to **submit-itw@ee.ust.hk** in PDF or PS formats are accepted. Please mention "Contributed Paper" in the subject line of your email. Each submission must be at most 2 pages in length using no smaller than 10-point fonts. The Latex template for the manuscripts can be downloaded here.

LaTeX style file Example(TEX) Example(PDF) Example(PS) Example(DOC)

Submission deadline:	**March 15, 2003**
Notification of acceptance:	**April 15, 2003**
Camera-ready deadline:	**May 15, 2003**

Submissions that cannot be accommodated in contributed paper sessions may be considered for poster sessions. All papers accepted for presentation will be published in the Workshop Proceedings.

Further Inquiries

Please send your questions to: **info-itw@ie.cuhk.edu.hk**

3rd International Symposium on Turbo Codes & Related Topics

BREST, FRANCE

1 - 5 September, 2003

General Chair:
R. Pyndiah, *ENST Bretagne, France*

Vice General Chair:
M. Jézéquel, *ENST Bretagne, France*

Technical Program Chair:
C. Berrou, *ENST Bretagne, France*

Members of the technical program:
P. Adde, *ENST Bretagne, France*
T. Aulin, *Chalmers University, Göteborg, Sweden*
G. Battail, *ENST Paris, France*
J. C. Belfiore, *ENST Paris, France*
S. Benedetto, *Politecnico di Torino, Italy*
A. G. Burr, *University of York, Great Britain*
G. P. Calzolari, *ESA, Germany*
D. J. Costello Jr., *University of Notre Dame, Indiana, USA*
C. Douillard, *ENST Bretagne, France*
M. Fossorier, *University of Hawaii, USA*
J. Hagenauer, *UTM, München, Germany*
L. Hanzo, *University of Southampton, Great Britain*
J. Huber, *Universität Erlangen-Nürnberg, Germany*
M. Jézéquel, *ENST Bretagne, France*
R. Johannesson, *Lund University, Sweden*
F. R. Kschischang, *University of Toronto, Canada*
D. Lacroix-Penther, *France Telecom R&D, France*
J. Lodge, *CRC, Ottawa, Canada*
H. A. Loeliger, *ETH, Zürich, Switzerland*
T. Maseng, *Lund University, Sweden*
A. Picart, *ENST Bretagne, France*
V. Poor, *Princeton University, USA*
R. Pyndiah, *ENST Bretagne, France*
P. Robertson, *DLR, Wessling, Germany*
H. Sari, *Pacific Broadband Communications, France*
C. Schlegel, *University of Alberta, Canada*
S. Shamai (Shitz), *Technion, Israel*
H. Suda, *NTT DoCoMo, Yokosuka, Japan*
R. M. Tanner, *University of Illinois, USA*
G. Ungerboeck, *Broadcom Corporation, USA*
R. Urbanke, *EPFL, Lausanne, Switzerland*
L. Vandendorpe, *Université Catholique de Louvain, Belgium*
G. Zemor, *ENST Paris, France*

Brest, at the westernmost tip of France, is renowned for its deep well-protected natural harbour and its famous navigators.

First Call for Papers

The 3rd International Symposium on Turbo Codes & Related Topics, organized by ENST Bretagne, will be held on Monday 1st - Friday 5th September, 2003, at the Quartz Congress Center, Brest, France.
The Symposium will be the opportunity to provide a broad overview of the current status and advanced research in **iterative methods and their application to information theory,** especially for digital communications. The symposium will include regular papers for oral and poster sessions as well as some invited papers.
All contributions based on the "turbo" or "probabilistic message passing" principle will be considered, in both the theoretical and the application field. The non-exhaustive list below gives possible topics for the papers submitted:

- error correction coding, turbo codes and turbo-like codes (parallel, serial, convolutional, block, LDPC, ...)
- coded and turbo coded modulation
- detection and turbo detection
- equalization and turbo equalization
- synchronization and turbo synchronization
- multi-user detection
- any combination of the above functions
- bounds and performance
- component algorithms (SOVA, MAP, ...)
- interleaving and graphs
- convergence properties
- circuits and software
- current applications and standards
- new applications of the "message passing" method.

Like for the previous Symposium, some papers will be selected for publication of an extended version, in a special issue of the *"Annals of Telecommunications"* journal.

Submissions
Authors are invited to submit a full 4-page paper before **March 31, 2003.** Only electronic submissions will be accepted via the Symposium web page at:

www-turbo.enst-bretagne.fr

where instructions to authors will be available.
At least one author of the paper **must be registered** for the Symposium by **June 15, 2003** in order to be published in the proceedings.

Key dates
Information regarding registration, accommodation and transport will be given later. Please check on the Symposium web site.

Submission of papers deadline: **March 31, 2003**
Notification of acceptance: **May 15, 2003**
Final versions of papers and preferential rate registration deadline: **June15, 2003**

ENST Bretagne
Secretary for the International Symposium on Turbo Codes
BP 832 - 29285 BREST Cedex - FRANCE
Phone: +33 2 29 00 10 28 Fax: +33 2 29 00 11 84
email: turbo.symposium@enst-bretagne.fr

Conference Calendar

DATE	CONFERENCE	LOCATION	CONTACT/INFORMATION	DUE DATE
March 31- April 4, 2003	**2003 IEEE Information Theory Workshop**	Louis Liard Room of La Sorbonne Paris, France	See CFP in this issue. http://www.comelec.enst.fr/ itw2003/index.html	Oct. 30, 2003
May 18 – 21, 2003	**2003 Canadian Workshop on Information Theory**	Waterloo, Ontario Canada	http://www.multicom.uwaterloo. ca/cwit2003	Jan. 7, 2003
June 29 - July 4, 2003	**2003 IEEE International Symposium on Information Theory (ISIT)**	Pacifico Yokohama, Yokohama, Japan	Prof. Ryuji Kohno Yokohama National University Graduate School of Engineering Division of Physics, Electrical and Computer Engineering 79-5 Tokiwadai, Hodogaya-ku Yokohama, 240-8501 JAPAN +81-45-339-4116 +81-45-338-1157 (fax) isit2003@kohnolab.dnj.ynu.ac.jp http://www.isit2003.org	Nov. 1, 2002
July 6-10, 2003	**2003 IEEE Information Theory Workshop**	New World Renaissance Hotel Hong Kong, China	Victor Keh-wei Wei & Raymond Wai-ho Yeung The Chinese University of Hong Kong {whyeung,kwwei}@ie.cuhk.edu.hk http://itwhk03.cs.ust.hk	Mar. 15, 2003
August 27-29, 2003	**13th IFAC Symposium on System Identification**	Rotterdam, The Netherlands	Prof. Paul Van den Hof Delft University of Technology The Netherlands p.m.j.vandenhof@tnw.tudelft.nl www.sysid2003.nl	Nov. 20, 2002
September 1-5, 2003	**3rd International Symposium on Turbo Codes and Related Topics**	Brest, France	http://www-turbo.enst-bretagne.fr/	March 31, 2003
September 24-25, 2003	**InOWo'03 - 8th International OFDM Workshop**	Hamburg, Germany	Prof. Hermann Rohling Department of Telecommunications TU Hamburg-Harburg, Ei§endorfer Str. 40 D-21073 Hamburg, Germany Tel: +49 (0)40 42878 3228 Fax: +49 (0)40 42878 2881 email:rohling@tu-harburg.de http://ofdm.tu-harburg.de	TBA
December 1-5, 2003	**GLOBECOM 2003**	San Francisco Marriott San Francisco, CA	Ms. Patricia Dyett IEEE Communications Society 305 E. 47th St., 9th Floor New York, NY 10017 +1 212 705 8999 (Fax) +1 212 705 8943 GLO2003C@comsoc.org	February 15, 2003

June 20 - 24, 2004	**2004 ICC**	Paris, France	http://www.icc2004.org	TBA
June 27 - July 2, 2004	**2004 IEEE International Symposium on Information Theory (ISIT)**	Chicago, Illinois, USA	See CFP in this issue http://www.isit2004.org	Dec. 1, 2003
July 19-24, 2004	**2004 Stochastic Networks Conference**	Centre de Recherches Mathematiques Universite de Montreal Montreal, Canada	http://www.stanford.edu/group/stochnetconf/	
TBA	**2005 IEEE International Symposium on Information Theory (ISIT)**	Adelaide, AUSTRALIA		TBA

IEEE IEEE Information Theory Society Newsletter

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