IEEE Information Theory Society Newsletter

Vol. 57, No. 4, Dec. 2007 Editor: Daniela Tuninetti ISSN 1059-2362

Teaching it

XXVIII Shannon Lecture, presented on Thursday, June 28, 2007 at the 2007 IEEE International Symposium on Information Theory, Nice, France

Sergio Verdú

Having attended all Shannon Lectures since Bob Gallager's in 1983, I have enjoyed the variety of ways in which Shannon lecturers have risen to the challenge, ranging from surveys to open problems, from tutorials to unveiling of new research results. My favorite Shannon lectures managed to find a unifying theme with a strong autobiographical component mixed with a sense of historical perspective.

As I started to ponder a topic for my lecture, it became apparent that identifying a unifying theme in my information theory research would be challenging. Over the years, I have moved from problem to problem with scarcely more sense of recognizable direction than a random walk. I suspect that in most research careers what we find along the way often turns out to be more seductive than whatever we set out to pursue. Certain traits to be avoided in most walks of life I find to be surprisingly beneficial to the researcher. A certain dose of attention deficit disorder can be salutary: not only it prevents you from turning the crank for too long but forces you to approach problems from the clean-slate perspective of the beginner; curbing one's industriousness enjoys a longstanding reputation as an engine of good mathematics; and nothing better than a touch of claustrophobia to avoid crowded bandwagons.

One of the great assets in our field is the wide variety of excellent textbooks, written from drastically different perspectives that reflect the wonderful diversity of our community. As a graduate student, I had taken a graduate information theory course from Bob McEliece using his concise textbook. The first time I taught the subject, Dick Blahut's book had just come out, and then, in the early nineties Cover-Thomas arrived to the delight of students and teachers alike, quickly becoming the established canon, just as a quarter of a century earlier, Gallager had become the undisputed leading textbook. Although not exactly suitable for teaching engineers, special mention must be made of Csiszár-Körner, whose influence on contemporary Shannon theory research is hard to overestimate.

Soon after I began doing research in information theory, I started teaching it, and since then, one has influenced and (hopefully) benefited from the other. I structured my Shannon lecture so as to highlight several points of departure of my teaching from the traditional approaches, while giving myself a chance to showcase some of my own research results. I reckon that for any sixty-year old discipline one would be hard-pressed to argue that there is anything radically new we can come up with while teaching the basics. Furthermore, the presumption is that after so many excellent and influential textbooks, the established approaches should have found a pretty good approximation to the most enlightening path. The endless quest for simplicity and elegance is a powerful force in both teaching and research. However, simplicity is very much in the eye of the beholder, and likely to be gauged differently by the novice and the seasoned researcher.

Over the years, the prevailing outlook on the fundamentals in information theory has evolved greatly. Consider the proof of the positive part of the channel coding theorem: Feinstein's non-random coding proof [1] was hailed in its time as the first rigorous approach, and even the legitimacy of Shannon's random coding was challenged by someone as influential as Wolfowitz. The emergence of Gallager's random-coding analysis [2] of the maximum-likelihood decoder put random coding back at the forefront of the standard teaching approach. At present, the overwhelmingly favorite proof is that of jointly typical sequences, which can be viewed as a formalization of the reasoning in *A Mathematical Theory of Communication*. And like in 1948, currently the prevailing wind comes from probability, rather than combinatorics.

In "teaching it" I went over the major topics covered in an information theory course: coding theorems for lossless compression, channel coding and lossy compression, information measures

continued on page 8

From the Editor

Daniela Tuninetti

Dear IT society members,

I trust you all had a good start of the fall semester. By the time this issue arrives on your desk, you will already be preparing finals and planning winter vacations. Although summer may appear long gone, I hope this issue of the newsletter will bring back warm memories of the sunny Cote D'Azur of late June 2007. For those who attended ISIT 2007 in Nice, France, and even more so for those who could not attend it, this issue summarizes most of the major events of the conference.

Before I give you a taste of what you will find in this issue, I would like to thank our president Bixio Rimoldi, whose term will end in December 2007, for his excellent work and steering of the IT society during this year. It was a pleasure to work with Bixio and to be able to learn from him.

Let me start the review of ISIT by thanking the two co-chairs, Giuseppe Caire and Marc Fossorier, and the team that worked with them, for the formidable organization of ISIT 2007. The highlight of every ISIT is probably the Shannon Lecture. This issue opens with a summary of the memorable XXVIII Shannon Lecture "teaching it" given by Sergio Verdú. In addition to the summery of the lecture, you can find copies of the slides and the recorded video of the presentation at http:// www.isit2007.org/ (follow the link "Shannon Lecture".)

In this issue you will find our *regular* columns by our president Bixio Rimoldi, our historian Anthony Ephremides, our creative puzzle maker Sol Golomb, and NSF program manager Sirin Tekinay. Those "regulars" are followed by a report of the main ISIT 2007 events by Marc Fossorier and Giuseppe Caire. I usually place conference reports at the back of the issue. However, this time, the conference report precedes the other articles since this December 2007 issue is dedicated in large part to all that happened at ISIT 2007 in Nice. The conference report gives you a succinct timeline description of the hectic ISIT week. The detailed summary of the major events follow.

I am proud to offer you the summary of the tutorial on *Network Coding*, by Christina Fragouli and Emina Soljianin. In case the article made you feel the urge to learn more on Network Coding, the same authors have recently published a monograph entitled "Network coding fundamentals" in Foundation and Trends in Networking, Now Publishers, June 2007, soon to be followed by "Network coding advanced topics and applications" by the same publisher.

You will also find the summary of all plenary lectures: Michelle Effros: *Network Source Coding: Pipe Dream or Promise?*; Shlomo Shamai (Shitz): *Reflections on the Gaussian Broadcast Channel: Progress and Challenges* (actually, this lecture is partly summarized in the reflection the same author provided for the 2007 IT Paper Award, also in this issue); Vincent Poor: *Competition and Collaboration in Wireless Network;* and Emery N. Brown: *Signal Processing Algorithms to Decipher Brain Function*. As for the Shannon Lecture, you can find copy of the slides and the video of the presentations at http://www.isit2007.org/ (follow "Plenary Speakers".)

continued on page 4

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

President's Column

Bixio Rimoldi

On Sept. 26, at Allerton Park, the IT Society had its last Board of Governors (BoG) meeting of the year. The highlight of the meeting was the proposal of the Online Committee, chaired by Nick Laneman, for a new forward-looking multi-purpose web site. Nick presented a very sound project and the BoG approved its realization. By the time you read this column we may already have a new web site. Beyond the look, the essential difference of the new site will reside in its "engine", which will be based on a content management system (CMS). This approach will enable a more dynamic website with state-of-the-art extension possibilities. The development will be supervised by the Online Committee, with certain aspects outsourced to professional developers. To minimize risk and attract volunteers, the project will proceed in phases. The goals of the initial phase are to choose a CMS and build an interface that focuses on ease of use while providing advanced functionality with respect to the current site. Subsequent phases will target customizations that will help Society volunteers in their various roles and add value to Society membership. The Online Committee welcomes suggestions for such customizations and encourages feedback on the web site as it evolves. You may send them to webdiscuss@lists.itsoc.org

Of the other decisions and announcements that were made at the meeting, I would like to mention a few that are of general interest:

(i) The BoG passed a proposal by Aylin Yener and Gerhard Kramer to organize a School in Information Theory inspired by the European model of Winter/Summer Schools. Aylin is the new Chair of the Student Committee;

(ii) The ad-hoc Editorial Committee that Ezio Biglieri was charged to form and chair is now in place. The other committee members are Helmut Bölcskei, Tony Ephremides, Dave Forney, Vince Poor, Anant Sahai, and Daniela Tuninetti. The committee has already reached a consensus on one of the various issues that it was charged to consider (see the President's Column in the September issue). It recommends eliminating the Correspondence category from the IEEE Transactions on Information Theory. If approved by the BoG, all future IT Transactions papers will belong to the same category;

(iii) The BoG approved a proposal by Alexander Kuleshov, Vladimir Blinovsky, and Anthony Ephremides to organize the 2011 IEEE International Symposium on Information Theory in St. Petersburg, Russia;

(iv) The ISIT'07 Student Paper Award was given to two papers, namely to "Minimum Expected Distortion in Gaussian Layered Broadcast Coding with Successive Refinement" by Chris T.K. Ng, Deniz Gunduz, Andrea Goldsmith, and Elza Erkip, and to "Uplink Macro Diversity with Limited Backhaul Capacity" by Amichai Sanderovich, Oren Somekh, and Shlomo Shamai. Congratulations to the student authors and their advisors;

(v) A motion was passed to make the Student Paper Award a regular ISIT event.

Next, I would like to briefly summarize the highlights of the feedback from the IEEE committee that reviewed the Society this year. As a general comment, the committee found the Information Theory Society to be "well managed and well run" as well as "financially sound, with a successful conference and a prestigious publication." The section dedicated to the transactions made a particularly glowing comment, namely "The IEEE Transactions on Information Theory is an exemplary periodical among all IEEE publications, with a very strong tradition in quality. The impact factor is among the highest of all IEEE periodicals, and has been for several years. The Editorial staff is a dedicated group of volunteers who are all focused on the single goal of maintaining the highest standards of the Transactions. While past problems with respect to timeliness have marred this exemplary performance, over the last five years since the last review the Information Theory Society has worked to steadily reduce the submission-to-publication time to the point that it is now within target guidelines for IEEE Periodicals." The report welcomes the creation of the Student Committee as well as the experiment of including, as a membership benefit, the online access to conference proceedings. The report also makes a number of recommendations including the following: the society is encouraged to increase the representation of BoG members from Regions outside of 1-6; the formation of Technical Committees to provide a platform to discuss and develop technical subjects is recommended; the paper acceptance rate at ISITs (60-70%) is found to be on the high side; It is recommended that more be done to ensure participation from the industry and to highlight the relevance of our work to the practitioner; it advocates that the support for Chapters should be reinforced and formalized; a similar comment is made for the Distinguished Speakers program and it is suggested that it could be made a strategic vehicle in reinforcing the Society's ties with its chapters. Once again, I would like to thank Marc Fossorier, Andrea Goldsmith, Nick Laneman, Steve McLaughlin, Muriel Medard, Dave Neuhoff, Vince Poor, and Daniela Tuninetti for their participation during the preparation of the review material and/or during the review itself. The highest praise goes to those that have contributed to keeping the quality of our IT Transactions so high, i.e., the past Editors in Chief, the Associate Editors, the reviewers, and of course the authors. The plan is to discuss the review feedback at the next officers meeting in January. As a side note, I'd like to mention that the reduction of the average submission-to-publication time is one of Ezio Biglieri's priorities.

This is my last President's Column. I would like to thank the officers, the Board of Governors members, and all the volunteers for their tireless dedication. In particular, I would like to thank Steve McLaughlin, outgoing Senior Past President, who will leave the BoG after twelve straight years of service.

Serving as the President of the IT Society has been a very interesting and rewarding experience for me. Certainly this is mainly due to the quality of the interactions I have had with many dedicated society members. It also gave me the opportunity to interact with the rest of IEEE. I have found it puzzling how little influence societies have on the IEEE organization. Part of the reason is that technical societies participate in the decision-making process through the Division Directors who are outnumbered by the directors of the non-technical areas. The fact that, in most cases, Division Directors have to represent many societies that have little interaction among one another further weakens the position of individual societies. In fact, societies talk to one another on IEEE matters through their presidents which meet three times a year during the Technical Activity Board (TAB) meetings. It is also at these meetings that society presidents talk to their Division Director and interact with some of the IEEE organizational units. The three meetings that the President of the IT Society attends are far to few to impact the system. Serving two years, as it is the case for many IEEE societies, does not solve the problem. A solution could be that a society has a sort of ambassador, ideally a past president, who serves for many years and accompanies the society president at all IEEE meetings. At next TAB meeting I will check if this is a possibility.

From the Editor

continued from page 2

The summaries of the plenary talks will be followed by the reflections on the paper awards presented or announced at the Award Luncheon in Nice: a) The 2006 IEEE Information Theory Society Paper Award "*Universal Compression of Memoryless Sources Over Unknown Alphabets*,", co-authored by A. Orlitsky, N. P. Santhanam, and J. Zhang; b) The 2007 IEEE Information Theory Society Paper "*The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel,*" co-authored by by H. Weingarten, Y. Steinberg and S. Shamai (Shitz); c) The 2007 IEEE Eric E. Sumner Award conferred to M. Luby and A. Shokrollahi for their work on Fountain Codes. Congratulations to all the authors on their paper award.

In Nice, I had the pleasure to meet again with Kees Schouhamer Immink (foreign associate of the US National Academy of Engineering since 2007 "for pioneering and advancing the era of digital audio, video, and data recording"). Kees contributed to this newsletter with an extremely enjoyable article on the Compact Disc's silver (25 years) jubilee. The article gives some background related to the coding technology options within a historical context. Do not miss it!

The 2006 IEEE Information Theory Society Chapter of the year was announced in Nice to go to the Hong Kong Chapter. Jong-Seon No has provided us with a description of the activities and initiatives of the Chapter. Congratulations to the Hong Kong Chapter members on this award.

I hope you will also enjoy the update on the activities of the student committee, on the work of the Ad-hoc committee for on-line content and Web, the minutes of the Board of Governors in Nice, the report on the 2007 Information Theory Workshops that took place in Bergen, Norway, in July and Lake Tahoe, CA USA, in September, as well as the report on the Redundancy 2007 workshop, that took place in Saint-Petersburg, Russia.

If you made it all the way to the end of the issue, do not miss out our latest call for papers, the conference calendar for incoming deadlines, and the call for nominations for the 2008 IT awards (whose deadlines are fast approaching.)

In late July, our society was shocked by the news of the tragic loss of Prof. Sergio Servetto from Cornell. Sergio was a dear friend, a passionate scholar and an active member of our society. The student committee, that Sergio chaired, organized a memorial to celebrate the life and work of Sergio in September at Allerton. A special session to commemorate Sergio's contributions to the field will be organized by Toby Berger at ISIT 2008. In this issue, you will read about the BoG's suggestion on how to show appreciation for Sergio's service to the IT society. The obituary will appear in the next issue of the newsletter.

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

Issue	**Deadline**
March 2008	January 10, 2008
June 2008	April 10, 2008
September 2008	July 10, 2008
December 2008	October 10, 2008

Electronic submission in Ascii, LaTeX and Word formats is encouraged. Potential authors should not worry about layout and fonts of their contributions. Our IEEE professionals take care of formatting the source files according to the IEEE Newsletter style. Electronic photos and graphs should be in high resolution and sent in as separate file.

I may be reached at the following address:

Daniela Tuninetti
Department of Electrical and Computer Engineering
University of Illinois at Chicago,
M/C 154
851 S. Morgan St.,
Chicago, IL, 60607-7053, USA
E-mail: daniela@ece.uic.edu

I wish everyone happy winter holidays and all the best for the New Year,

Daniela Tuninetti

[1] The reflection on the 2006 Joint ComSoc/IT Society Paper Award "*Universal Discrete Denoising: Known Channel,*" co-authored by T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. Weinberger, appeared in the previous issue of the Newsletter.

The Historian's Column

Anthony Ephremides

Recently, out of the blue, a book found its way into my hands, courtesy of David Middleton, who needs no introduction to our community. Well in his eighties, David continues to work along the lines that defined his scientific quest over the last several decades, namely the exploration of how physics interacts with the process of communication. So he brought to my attention a book of unique historical interest for our field. It is a book that was published (in Russian) late last year by Technosphera publishers in Moscow. Its title is "Pioneers of the Information Century" and the author is Mark Bykhovsky, Professor and author of 245 scientific articles that span several areas from electromagnetism and propagation to radio systems analysis and history aspects of science and technology.

The book has a Foreword note by Academician Y.V. Gulyaev and consists of the technical biographical sketches of a formidable selection of giant figures who have essentially defined and developed our field in the broadest sense. What sparked my interest is the perspective of a "fellow"-historian from Russia in the selection of scientists and in the description of their contributions. Before I mention the names of these scientists I must emphasize that my reading of this book is still a "work-in-progress", since it is written in Russian with only a table of contents (and Gulyaev's brief introduction) in English. My Russian is limited to what I was able to "pick" after one semester course at Princeton in 1968 and what I managed to add from my reading of the libretti of operas like Boris Godunov, Queen of Spaces, etc. However, I was lucky to have some wonderful colleagues with Russian as their native language. I enlisted their help and they obliged gracefully. They are Sasha Barg (well known to our community) and Isaac Mayergoyz, an Aristotelian scientist for whom any description of his "specialty" will not do him justice. Let us say that he rides the border between physics and electrical engineering and that his "home" area in IEEE is the Magnetics Society. David, Sasha, and Isaac are hereby recognized as "honorary" Historians and thereby join the "league" of other luminary scientists in our field who have earned their "history" credentials through contributions to this column.

The list of the "pioneers" that Mark Bykhovsky has selected, in alphabetical order, includes the following: Dimitry Ageev, Vladimir Bunimovich, Harry Van Trees, Norbert Wiener, Andrew Viterbi, Roland Dobrushin, Andrei Kolmogorov, Vladimir Kotelnikov, Boris Levin, Victor Melnikov, David Middleton[1], Harry Nyquist, Stephan Rice, Vladimir Siforov, Ruslan Stratonovich, Vasily Tikhonov, Lev Fink, Alexander Kharkevich, Alexander Khinchin, Richard Hamming, Claude Shannon, and Agner Erland.

After catching his/her breath from reading this list, the reader may have various thoughts in reaction to this list. First of all, it does include all the "major" giants like Shannon, Kolmogorov, Kotelnikov, and Wiener. Secondly, it includes many Russian contributors who are not as broadly known in our community (this is expected, of course). Thirdly, there are significant omissions (especially from outside Russia). Before engaging in a perilous debate about who is and who is not a pioneer of the Information age, it is helpful to review the second part of this book which outlines (over 4 chapters) the author's view of the History of Communication. His perspective in choosing the names of those he included is thus revealed, by means of his "view" of what constitutes a proper account of Statistical Communication Theory and its evolution.

The first chapter discusses essentially random process theory and, in particular, the creation and evolution of linear and non-linear transformations of random processes. The second (very brief) chapter focuses on optimal linear filtering. The third chapter studies the theory of noise and the optimization of noise elimination in the transmission of signals. This is an extensive chapter and it highlights the contributions of additional scientists, like Robert Price, Robert Lucky, Don Snyder, and others. The fourth (and last) chapter is devoted to the creation and evolution of Information Theory. It is also quite extensive and recognizes additional contributors to the field like Fano, Pinsker, Gallager, Huffman, Ziv, Slepian, Berlekamp, Elias, Forney, and others.

The book concludes with a list of references that in the author's view represent a good compromise between completeness and adequacy in outlining the landmarks in the development of our field.

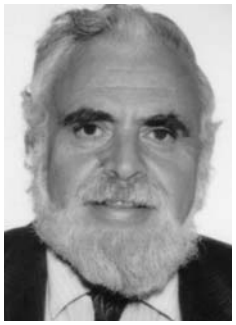
One may react to this book in a variety of ways. It is possible to dismiss the book as amateurish, incomplete, and biased. It is also possible to hail it as a landmark in chronicling the stages of development of our field and in recognizing those who made it possible. It is always risky to establish a "pantheon" of any sort; and it is doubly risky when the field is still evolving. It is also a monumental undertaking to try to encapsulate the technical benchmarks of the field in brief and epigrammatic outline. Assigning degree of "significance" to scientific contributions is a task that sparks debate and, possibly, discord. Nonetheless, it is a worthwhile undertaking. Until I am able to read the content in full I am inclined to withhold final judgment. However, I am prepared to recognize this book as a bold initiative in filling the relative vacuum that exists in the recording of the history of our field. The author must be commended for a valiant and careful effort that does valuable service to our field. Let it also be noted that the book is quite up-to-date since it even includes the development of turbo-codes. There will be more commentary to come in future issues. Writing the history of our field deserves recognition and attention. Take it from a "fellow historian."

[1] The English translation of the section dedicated to David Middleton appeared in the June 2007 issue of the newsletter. The translation was a courtesy of Mark Bykhovsky's daughter Julia Bykhovskaia.

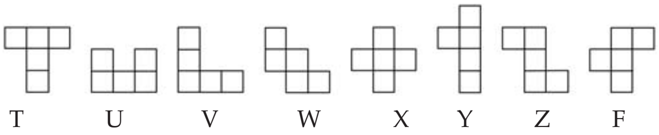
GOLOMB'S PUZZLE COLUMN™

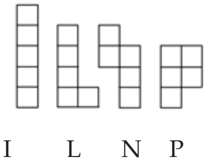
PENTOMINO EXCLUSION

Solomon W. Golomb



The 12 pentominoes are the figures made of 5 connected squares, identified by letter names as follows:

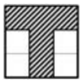
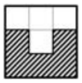




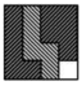
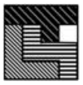
(These figures may be rotated and flipped over at will.)

The "Pentomino Game" consists of two players taking turns placing a not-yet-used pentomino on an initially empty rectangular board (following the grid lines on the board), with the first player unable to fit another pentomino on the board being the loser.

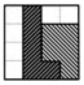
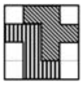
In this column, we will only consider the shortest possible pentomino games on $n \times n$ boards --- i.e., the smallest number, k, of pentominoes which can be placed on an $n \times n$ board so that none of the other $12 - k$ pentominoes will fit. On the 3×3 board, clearly $k = 1$, since any one of T, U, V, W, X, Z, F, or P will fit, and only four empty squares will be left, too few to hold a pentomino. E.g.:

 and  .

On the 4×4 board, it is *possible* to fit three pentominoes, e.g.

 or  ;

but it is very easy to place only two pentominoes so that none of the other 10 will fit, e.g.:

 or  .

(It is *not* possible to place a single pentomino on the 4×4 board in such a way that none of the other 11 will fit.)

Here are your problems.

1. Place *two* pentominoes on the 5×5 board so that none of the other ten will fit. (This has a unique solution!)

2. Place 3 pentominoes on the 6×6 board to keep off the other 9. (There are several solutions. Most of them use the L, the V, and one other.)

3. Place 4 pentominoes on the 7×7 board so that none of the other 8 will fit. (There are many solutions, e.g. using the I, L, V and *any* one other pentomino. Only 4 solutions are known using neither I nor V, and one of these is the unique solution that does not use the L. Extra credit if you can find these four, and especially the *last* one!)

4. Place 5 pentominoes on the classic 8×8 board so that none of the other 7 will fit. (Fifteen solutions are known. Of these, 14 use I, L, U, V, and Y. Find at least one of these. The remaining solution does not use U or Y. Extra credit if you can find it!)

5. Place 6 pentominoes on the 9×9 board to keep off the other 6. (I have several solutions, but they all use the same six pentominoes.)

Note that in Problems 1 through 5, the minimum number of k pentominoes on the $n \times n$ board needed to keep out the other $12 - k$ has $k = n - 3$. This works for $5 \leq n \leq 9$, and trivially for $n = 15$ (since after you have placed $n - 3 = 12$ pentominoes on the 15×15 board, there are none left to place!). Here is the best I've found for $n = 10, 11, 12$, and 13. Perhaps you can do better.

6. Place 7 pentominoes and 1 monomino (i.e., a single square) on the 10×10 board so that none of the other 5 pentominoes will fit.

7. Place 8 pentominoes and 1 monomino on the11 $\times$ 11 board so that none of the other 4 pentominoes will fit.

(Let me know if you find a solution to Problem 6 or Problem 7 without using that extra square!)

8. Place 10 pentominoes on the 12×12 board to keep off the other two. (My solution excludes the W and the X. Can you use 9 pentominoes and an extra square or two to exclude the other 3 pentominoes?)

9. Place 11 pentominoes on the 13×13 board that will keep off the remaining pentomino. (Hint: The X-pentomino is the easiest one to exclude.)

Tragic Loss of Sergio Servetto

The IEEE Information Theory Society mourns the loss of Professor Sergio D. Servetto, who passed away on July 24th 2007. He was the sole victim of an accident in a private plane, which he was flying from Michigan to Ithaca, NY.

Remembered by many as a passionate scientist with great intelligence, energy, daring, and personal warmth, Sergio was well known for his research and his involvement in the society, in which he recently served as chair of the Student Committee. Dr. Servetto is survived by his wife Viviana and their two small boys, Luciano and Alejandro. A condolence book for the family is available online: its link can be found on the society website.

Until his tragic accident, Sergio was the sole provider for the entire income of his family. Given the very minimal amount of life insurance that he had, Sergio's family now faces extremely difficult times, as they are left without a secure source of income. A trust fund aimed at helping the Servetto family, not only in their immediate hours of need but also in the difficult years to come, has been established by the IT Society and Cornell University. The IT Society Board of Governors suggests that individual members consider making a donation to the trust fund as a token of appreciation for Sergio's service and as a way to support fellow members in the worst of times.

Donations can be made using the following information:

Name of Fund: The Servetto Family Fund;

Through: CFCU Community Credit Union

Contributions: at any CFCU branch or mail to the main office. CFCU Community Credit Union, 1030 Craft Road, Ithaca, NY 14850

Checks payable to: Servetto Family Fund

Donations can also be made via international bank transfer to a European account using the following codes:

IBAN: PT50 0035 0578 00004602500 54;

BIC SWIFT: CGDIPTPL

Name: João Barros; Address: Rua do Campo Alegre, 1266, 4 DIR, 4150-174 Porto, Portugal.

When using this option, please email jbarros@fc.up.pt with the name, location, and donated amount for bookkeeping purposes. Inquiries may be sent to the same email.

Teaching it

continued from page 1

and their operational roles, as well as the extremization of mutual information.

Summarizing the main contentions I put forward in the lecture:

1. The Asymptotic Equipartition Property (AEP) does not lead to the simplest proof of the data compression theorem. The AEP itself is equivalent to a simplified version, which states that the probability of the too likely sequences vanishes. However, the role of the AEP is vindicated by the fact that it is a necessary and sufficient condition for the validity of the strong coding theorem (even in abstract non-serial settings) [3].

2. Most of the action in information theory takes place in the non-asymptotic regime. It is conceptually enlightening to separate the proof of coding theorems into the non-asymptotic and the asymptotic stages. The non-asymptotic stage deals with the core achievability and converse bounds on error probability. The asymptotic stage usually invokes a standard tool such as the law of large numbers (LLN), or the Shannon-MacMillan ergodic theorem in order to show that the bounds are asymptotically tight. The *information spectrum* approach carries this program to the point where the fundamental information theoretic limits are expressed in terms of the source and/or channel without ergodicity assumptions [4], [5].

3. The conceptually simplest bounds show that, modulo epsilons/deltas, the error probability behaves as the cdf of the information random variables: log probabilities (often referred to as "ideal codelenghts") in lossless data compression; log likelihood ratios of conditional to unconditional probabilities ("information densities") in channel coding and lossy data compression.

4. Confining the teaching of lossless data compression to memoryless sources incurs a big loss, both conceptually and in real-world relevance. The memoryless assumption (of channels and sources) does not simplify the non-asymptotic analysis conceptually, although it certainly simplifies the evaluation of the gap from the asymptotic fundamental limit at a given blocklength.

5. For the most important results it is enlightening to teach more than one approach. For example, conventionally, the channel-coding converse is done via Fano's inequality. However, to show the (weak) converse in the ergodic case, it suffices to invoke the mutual information data processing theorem and the fact that block error rate cannot be smaller than bit error rate. Furthermore, Fano's inequality is not powerful enough to deal with nonergodic channels or to lead to the strong converse (see [5], [6]).

6. The conventional way to refine the asymptotic analysis leading to capacity and minimum source coding rate (based on LLN and generalizations) is to perform a large-deviations error exponent analysis. In contrast to most other applications of probability/statistics, the intermediate central-limit-theorem analysis has been largely and unjustifiably neglected in information theory. From a practical viewpoint, it is very relevant to investigate the tradeoff between performance and blocklength at rates in the vicinity of the fundamental limit. In contrast, the venerable large-deviations information theory school seems to have made little headway in convincing engineers to compress at, say, twice the entropy or to transmit at half of capacity for the sake of astronomically low block error probability in the asymptotic regime.

7. Based on the average-length analysis of prefix symbol codes, the standard approach to teaching variable-length lossless data compression misses the more fundamental performance limits of the optimal non-prefix code (also known as *one-to-one* code) applied not symbol-by-symbol but to the whole input. In addition to the average of the minimal length (e.g. [7]) it is of interest to analyze its distribution. It is easy to show that that is precisely accomplished by the analysis of the optimal almost-lossless fixed-block-length code.

8. Most inequalities among information measures boil down to the nonnegativity of (conditional) divergence; in those cases, it is less elegant and insightful to invoke Jensen's inequality or the log-sum inequality.

9. I highly recommend cutting down on the use of differential entropy (not the prettiest object in information theory).

10. Among the operational roles of the divergence $D(P||Q)$ is the large-deviations measure of the degree of difficulty of impersonating P by Q. Also, in information theory, besides the key role it plays in the method of types [8], it quantifies the excess compression rate incurred when the optimum code for Q is used in lieu of that of P. Less well-known is an LLN characterization showing that the log ratio of the a posteriori probabilities converges to the divergence. $D(P||Q)$ is also the maximum number of bits per $ that can be sent through a binary input channel in which sending 1 costs $1 and produces output distribution P, and sending 0 costs nothing and produces output distribution Q [9].

11. As is well known, in the minimization of mutual information to obtain the rate-distortion function, the key unknown is the optimal conditional distribution of the source given the reproduction. Likewise, in channel capacity, the key goal is to obtain the optimal output distribution.

12. In addition to channel capacity, the maximal mutual information finds operational interpretations as the minimax compression redundancy [10], [11]; identification capacity [12]; and channel resolvability for minimal-randomness simulation [4]. Another operational role for the maximal mutual information is the reliability of ergodicity testing: fix K probability distributions $P_1, \ldots P_K$; iid observations are generated in either ergodic mode (the distribution is the mixture $\pi_1 P_1 + \ldots + \pi_K P_K$) or in nonergodic mode (P_i is chosen initially with probability π_i). Based on n observations, we need to decide which mode is in effect. Fixing the probability of erroneously deciding ergodic, and assuming that the probabilities π_i are chosen in the most favorable way, the probability of erroneously deciding that the mode is nonergodic vanishes exponentially with n, with an exponent equal to the capacity of a memoryless channel whose input alphabet is $\{1, \ldots K\}$ and i^{th} conditional output distribution is equal to P_i.

But most of the above observations deal with proofs, tools and approaches. Even more important while teaching information theory is to instill the great allure of its fundamental formulas. If we have learned anything since 1948, it is that there is nothing more practical than a beautiful formula.

I ended the lecture with a salute to the previous winners of the Claude E. Shannon Award, accompanied by the overture of "Die Meistersinger von Nürnberg," an opera that extols the idea that progress is the finest tribute to tradition. Guided by the spirit of *A Mathematical Theory of Communication*, our continued pursuit of truth, simplicity and beauty in information theory is the best tribute to Claude Shannon's legacy.

References

[1] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. on Information Theory*, vol. PGIT-4, pp. 2–22, 1954.

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

[3] S. Verdú and T. Han, "The Role of the Asymptotic Equipartition Property in Noiseless Source Coding", *IEEE Trans. on Information Theory,* vol. 43, no. 3, pp. 847-857, May 1997.

[4] T. S. Han and S. Verdú, "Approximation Theory of Output Statistics," *IEEE Trans. on Information Theory*, vol. IT-39, no. 3, pp. 752-772, May 1993.

[5] S. Verdú and T. S. Han, "A General Formula for Channel Capacity," *IEEE Trans. on Information Theory*, vol. 40, no. 4, pp. 1147-1157, July 1994.

[6] J. Wolfowitz, *Coding Theorems of Information Theory,* 3rd Edition, Springer, New York, 1978

[7] N. Alon and A. Orlitsky, "A Lower Bound on the Expected Length of One-to-One Codes," *IEEE Trans. on Information Theory,* vol. 40, no. 5, pp. 1670-1672, Sept. 1994.

[8] I. Csiszár, "The method of types," *IEEE Trans. on Information Theory*, vol. 44, pp. 2505–2523, Oct. 1998.

[9] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Information Theory,* vol. 36, pp. 1019–1030, Sept. 1990.

[10] B. Ryabko, "Encoding a Source with Unknown but Ordered Probabilities," *Problems Information Transmission*, vol. 15, pp. 134–138, 1979.

[11] R. G. Gallager, "Source Coding with Side Information and Universal Coding," Tech. Rep. LIDS-P-937, MIT, Cambridge, MA, 1979.

[12] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. on Information Theory*, vol. 35, pp. 15–29, Jan. 1989.

Conference Report: The 2007 IEEE International Symposium on Information Theory, June 24-29, 2007, Nice France

Giuseppe Caire and Marc Fossorier

The 2007 International Symposium on Information Theory was held at the Acropolis Convention Center in Nice, France on June 24-29, 2007. It attracted 776 participants, including 275 students.

As usual, pre-conference tutorials followed by 295 registrants were presented on the first day of the event:

1. Statistical Physics Tools in Information Science, by M. Mezard and A. Montanari
2. Feedback and Side-Information in Information Theory, by A. Sahai and S. Tatikonda
3. Network Coding: Theory and Practice, by C. Fragouli and E. Soljanin
4. Recent Trends in Denoising, by D. Donoho and T. Weissman
5. Theory and Strategies for Cooperative Communications, by G. Kramer and N. Laneman

In an effort to value Society student membership, these tutorials were heavily discounted for students belonging to the IEEE Information Theory Society.

The following five days were dedicated to the presentation of the

Sergio Verdu "teaching it."

VIP table at the banquet: (from left to right) Robert Gallager, Giuseppe Caire, Mercedes Paratje, Sergio Verdu, Isabella Caire, Bixio Rimoldi, Muriel Medard, Tom Cover and Karen Cover.

The TCP co-chairs (from left to right) Andrea Goldsmith, Muriel Medard with little Jack, Amin Shokrollahi and Ram Zamir.

Bixio Rimoli and Andrea Goldsmith dancing at the banquet.

603 accepted papers in 8 parallel sessions. In addition, four excellent plenary lectures started four days of the event:

- On Monday, June 25, ``Network Source Coding: Pipe Dream or Promise?'' by M. Effros.

- On Tuesday, June 26, ``Reflections on the Gaussian Broadcast Channel: Progress and Challenges'' by S. Shamai

- On Wednesday, June 27, "Competition and Collaboration in Wireless Network" by H.V. Poor

- On Friday, June 29, ``Signal Processing Algorithms to Decipher Brain Function'' by E.N. Brown

The highlight of the event was the XXVIII Shannon Lecture by Sergio Verdu on Thursday, June 28: ``teaching it.'' This memorable lecture, as well as the four plenary talks, are summarized in this issue of the newsletter and are available on the conference website.

For the first time at an ISIT, a Student Paper Award was given and two papers shared this first award:

- "Minimum Expected Distortion in Gaussian Layered Broadcast Coding with Successive Refinement'' by C.T.K. Ng, D. Gunduz, A. Goldsmith and E. Erkip

- "Uplink Macro Diversity with Limited Backhaul Capacity'' by A. Sanderovich, O. Somekh and S. Shamai

Thanks to the National Science Foundation, our generous sponsors, as well as the IEEE Information Theory Society via a special initiative (in the amount of USD 27 K), free registrations were awarded to 87 students from 12 different countries as well as 11 researchers from 6 disfavored countries.

At the end of the banquet, the next symposium co-chairs Frank Kschischang and En-Hui Yang welcome everyone to meet again in Toronto, Canada for ISIT'08.

Overall, ISIT'07 was a great success under all viewpoints: in terms of scientific quality, attendance, venue and social events. As general co-chairs, we could count on a fantastic organizing team, without the help of which the success of the conference would not have been possible. Thanking individually each member of the team would be just too long, since the people involved with the organization were many. However, we would like to mention explicitly the individuals whose involvement was far beyond the normal volunteer level, and whose help was particularly appreciated. In particular, we would like to mention:

Prof. Merouane Debbah and Mr. Marios Kountouris (Ph.D. student) took care of the local organization and coordinated the legion of student volunteers who served at the conference center and at the registration desk for technical assistance.

Prof. Ulrich Speidel, who served as the conference webmaster and hosted the conference website.

Prof. Jean-Claude Belfiore, who edited and produced the program, abstract book and CDROM for the proceedings.

The technical program committee (TPC), and the four TPC co-chairs in particular, Profs.

Andrea Goldsmith, Muriel Medard, Amin Shokrollahi and Ram Zamir, accomplished the hard task of selecting 603 out of 995 submitted papers and formed an outstanding technical program. In particular, they managed to reach the initial goal of obtaining three reviews for almost every paper submitted.

Finally, we are also particularly grateful to Prof. Ubli Mitra for organizing and coordinating the tutorials and to Prof. Mehul Motani for the publicity (by inundating several mailing lists with ISIT'07 invitations and call for papers).

We are also very pleased for the large involvement of students in the conference organization, which contributed to create the familiar and friendly "academic" environment that, in our view, ISIT should have, and that makes it distinctively different from other large conferences held by other societies in the broad field of telecommunications.

In conclusions, we may say that one of the most significant achievements of this edition of ISIT was a conclusive proof of the French Paradox: "keeping clear minds to discuss information theory problems is possible even in the presence of Camembert cheese and Burgundy wines."

Network Coding Ten Years after the Butterfly Emerged

Tutorial presented on Sunday June 24, 2007 at the 2007 IEEE International Symposium on Information Theory, Nice France

Christina Fragouli and Emina Soljanin

Ten summers have passed since that of 1997, when Raymond Yeung visited Rudi Ahlswede at Bielefeld right before the ISIT at Ulm. There for the first time he met Ning Cai, and only a few weeks later the two had a manuscript containing, among other results, a claim that if there are only two sink nodes in a single-source network, store-and-forward routers are sufficient for both sinks to achieve the min cut rate.

Back home at CUHK, Raymond gave a copy of the manuscript to his colleague Robert Li, motivated by Bob's expertise in switching theory. In spite of not being able to make a connection with his own work, Bob was intrigued by the material enough to immediately spot the incorrect result. He went on to explain his argument to Raymond, and after thinking in front of his office white board for a little while, drew a counterexample, now widely known as the butterfly network shown in Fig. 1.

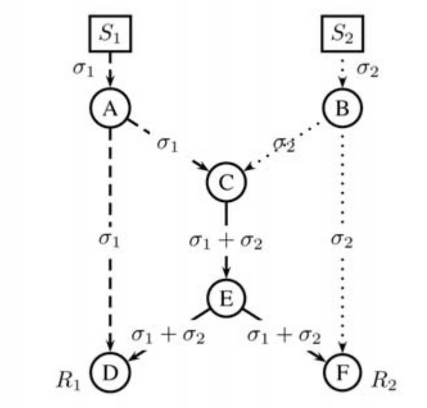


Fig. 1. The butterfly network: an example of linear network coding.

It took the group a few more years to get the original results in print [1] and subsequently those on linear network coding [2], but the cocoon had been broken and other researchers almost immediately started looking beyond routing. The publication presenting an algebraic approach to network coding by Ralf Kötter and Muriel Médard, in particular, galvanized the field [3]. Today there are two sets of monographs [4]–[7], an upcoming book [8], and even a popular science article [9] on the subject.

Network coding is attracting significant interest from engineers, computer scientists, and mathematicians at the wold's leading universities and research centers as well as in established companies such as Microsoft and Alcatel-Lucent, who see a financial benefit in making use of such methods. The widespread research on network coding is reflected on the number of papers on the subject and research grants awarded to support such research efforts. What makes the area particularly attractive is that ideas from network coding promise to advance at the same time both the theory and practice of networking, and also bring together researchers from diverse communities, thus creating opportunities for exciting interactions.

I. STATE OF THE ART

The emergence of the butterfly and network coding has indeed brought about a metamorphosis in thinking about network communication, with its simple but important premise that in communication networks, we can allow nodes to not only forward but also process the incoming independent information flows. At the network layer, for example, intermediate nodes can perform binary addition of independent bitstreams, whereas, at the physical layer of optical networks, intermediate nodes can superimpose incoming optical signals. In other words, data streams that are separately produced and consumed do not necessarily need to be kept disjoint while they are transported throughout the network: there are ways to combine and later extract independent information. Combining independent data streams allows to better tailor the information flow to the network environment and accommodate the demands of specific traffic patterns. This shift in paradigm is expected to revolutionize the way we manage, operate, and understand organization in networks, as well as to have a deep impact on a wide range of areas such as reliable delivery, resource sharing, efficient flow control, network monitoring, and security.

Today, ten years after the emergence of the butterfly network, a lot is already known about network coding, in particular for the case of network multicast. Network multicast refers to simultaneously transmitting the same information to multiple receivers in the network. The fascinating fact that the original network coding theorem brought was that the conditions necessary and sufficient for unicast at a certain rate to each of these receivers are also necessary and sufficient for multicast at the same rate, provided the intermediate network nodes are allowed to combine and process different information streams [1]. Soon after this initial finding, we learned that the multicast rate can be achieved by allowing the nodes to linearly combine their inputs over a sufficiently large finite field [2], and about a bound on the probability of successful decoding by all receivers when the nodes make random linear combinations over some finite field in terms of the field size and the number of receivers [10].

Algorithms to design linear network codes followed: deterministically in polynomial time [11] as well as deterministically and in a decentralized and scalable fashion [12]. Today we also know the field size that is sufficient to design network multicast codes in all networks with a fixed number of receivers [3], [12], and we understand the throughput benefits of coding in both directed and undirected multicast networks [13]–[15]. We are also beginning to understand the complexity of network coding [16], [12], [17], and how to code for networks with cycles and delay [12], [18]. At the same time, implementations such as Microsoft's Avalanche for content distribution [19], [20] and MIT's COPE [21] for wireless networks have demonstrated the practical effectiveness of network coding.

Network coding can and has been studied within a number of different theoretical frameworks, in several research communities, most notably Information Theory and Computer Science. The choice of framework a researcher makes most frequently depends on his background and preferences. However, one may also argue that each network coding issue (e.g., code design, throughput benefits, complexity) should be put in the framework in which it can be studied the most naturally and efficiently.

The original approach to network coding as well as the original proof of the main theorem was information theoretic [1]. Subsequently, both information theorists and computer scientists used information theoretic tools to establish bounds on achievable rates for the cases where the main theorem does not hold (e.g., to derive bounds for multiple unicast sessions traffic [22], [23]). One of the first approaches to network coding was algebraic [3]. Because the approach is very natural and requires little background, it is often adopted for introducing the subject of network coding. Randomized coding, believed to be of paramount importance for practice, has been developed within the algebraic framework [6], [7].

There is also the combinatorial framework of network coding that is mainly based on graph theory and combinatorics. This approach has been useful for understanding computational complexity of network coding in terms of the required code alphabet size and the number of routers required to code [16], [12], [17]. Combinatorial framework, in particular, enabled us to spot and exploit connections between network coding and well-studied discrete math problems, such as Latin squares [24], graph coloring [16], [12], coupon collection and other occupancy models [25], [15], and gossip algorithms [26].

Finally, there is a very rich history in using linear and integer programming for studying flows through networks. Network coding has also been studied within this framework, which turned out to be the most natural for addressing networks with costs [27]. Another notable achievement of this approach is the characterization of throughput benefits of network coding for multicasting [13]–[15].

II. OPEN PROBLEMS AND NEW APPLICATIONS

Network coding continues to be a very active field. More and more researchers and engineers ask what network coding is, what its benefits are, and how much it costs to design and operate networks implementing network coding. At this point, we do not have complete answers to these questions even in the case of network multicast. For example, the minimum network operating costs required to achieve maximum throughput are not known in general in terms of the code alphabet size and the number of routers required to code. (Multicast in networks with two sources and arbitrary number of receivers is almost completely understood.)

Even less is known in the arguably practically more important case of multiple unicasts, where not even throughput benefits of coding have been completely characterized. Although there are directed graph instances where network coding throughput increase is proportional to the number of nodes in the graph, we are yet to find an undirected graph instance where network coding offers any benefits. Another transmission scenario for which benefits of coding are not fully understood are networks with non-uniform demands [28], [29]. General traffic patterns are often very difficult to study, and optimal solutions may require exponential alphabet sizes [30] and non-linear operations [24], [31]. Finally, research on disseminating correlated information and the related problem of distributed source coding over network coded systems has just begun [32].

The Microsoft's Avalanche system has sparked the interest in using network coding for content distribution. Various tests and measurements have been carried out on experimental P2P systems, and results together with numerous observed advantages of using network coding were reported in [19], [20], [33]. However, fine-tuning this approach for specific applications, such as video on demand, and developing a theory that would completely support the experimental evidence are still missing.

Network coding allows us to exploit the broadcasting of wireless nodes through the shared wireless medium to provide benefits in terms of bandwidth, transmission power, and delay [34], [35]. Clearly, to warranty the deployment of such techniques, the required processing of data within the network needs to have low complexity and power consumption. MIT's COPE demonstrated that even when coding operations are confined to simple binary additions obeying some additional constraints, there are still gains to be had in terms of throughput and efficiency of MAC layer protocols [21].

Early work on wireless network coding ignored the interference of multiple broadcast transmissions at a receiver. Physical layer network coding was an initial attempt to move beyond this assumption [36]. Very recently, a linear deterministic model was developed that captures the interactions between the signals in a wireless network, and corresponding information-theoretic max-flow, min-cut results were obtained [37], [38]. Wireless and sensor networks provide vast opportunities for applications of network coding, and numerous and diverse problems are beginning to receive attention, ranging from techniques such as cross layer design [39] over issues such as fairness and delay [40], to untuned radios [41] and distributed storage [42]–[44].

Another line of recent work deals with networks in which some edges are in a certain way compromised [45]–[60]. The information carried by such edges may be deleted, altered by channel errors or malicious adversaries, or observed by someone whose information gain we would like to limit. Usually, no assumption is made on the choice of these edges, but their number is limited. Network codes can be designed for such networks, although some throughput has to be sacrificed to accommodate for compromised edges. The maximum achievable throughput is known for many of such scenarios, and, naturally, it depends on the size of the affected edge set. Classical information theoretic techniques have been useful to characterize achievable rates for networks with lossy and wiretapped links [49], [50], [58], but a number of results still have to be fully generalized. Algorithms for designing error-correcting, attack resilient, and secure network codes have also been proposed. For some cases, however, the codes we have today require huge alphabet size and are in general too complex to implement [52], [53]. Thus design of practical codes is of interest. A recent novel approach to coding for networks with errors seems particularly promising here [60]. In general, addressing security and reliability problems in networks implementing network coding is an area with many interesting still unanswered questions [53]–[59].

These days we are beginning to see network coding ideas being put to use in problems other than increasing throughput in networks with multiple users. There is evidence that network coding may be beneficial for active network monitoring [61], [62] as well as passive inference of link loss rates [63]. Interestingly, in a system employing randomized network coding, the randomly created linear combinations implicitly carry information about the network topology, that we can exploit towards diverse applications [64], [33]. Use of network coding techniques can help to increase rates of multicast switches [65]–[67], make data acquired by sensors more persistent and available throughout the network [42]–[44], and reduce on-chip wiring [68].

The network coding butterfly has even reached quantum information theorists. If we recall that in multicast communications networks, large throughput gains are possible with respect to their (physical) transportation or fluid counterparts because classical information can be processed in a way that physical entities can not, an interesting question to ask is whether anything can be gained by allowing processing of quantum information at nodes in quantum networks. Although physical carriers of quantum information can be processed in certain ways determined by the laws of quantum mechanics, two operations essential in classical information networking, replication (cloning) and broadcasting, are not possible. However, approximate and probabilistic cloning as well as different types of compression of quantum states are possible, and have been used in attempts to find a quantum counterpart of network coding [69]–[71].

REFERENCES

[1] R. Ahlswede, N. Cai, S-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204– 1216, July 2000.

[2] S-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," I*EEE Trans. Inform. Theory*, vol. 49, pp. 371–381, Feb. 2003.

[3] R. Kötter and M. Médard, "Beyond routing: an algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, pp. 782–796, October 2003.

[4] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang, *Network Coding Theory: Single Sources*, Foundations and Trends in Communications and Information Theory . Hanover, MA: now Publishers Inc., Volume 2, Issue 4, 2005.

[5] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang, *Network Coding Theory: Multiple Sources*, Foundations and Trends in Communications and Information Theory . Hanover, MA: now Publishers Inc., Volume 2, Issue 5, 2005.

[6] C. Fragouli and E. Soljanin, *Network Coding Fundamentals. Foundations and Trends in Networking*. Hanover, MA: now Publishers Inc., Volume 2, Issue 2, 2007.

[7] C. Fragouli and E. Soljanin, *Network Coding Applications. Foundations and Trends in Networking.* Hanover, MA: now Publishers Inc., to appear.

[8] T. Ho and D. S. Lun, *Network Coding: An Introduction*. Cambridge University Press, Cambridge, UK. Forthcoming, 2008.

[9] M. Effros, R. Kötter, and M. M´edard, "Breaking Network Logjams," *Scientific American*, June 2007, Vol. 296, Issue 6.

[10] T. Ho, R. K¨otter, M. M´edard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast," *IEEE Trans. Imform. Theory*, vol. 52, pp. 4413-4430, October 2006.

[11] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain and L. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Imform. Theory*, vol. 51, no. 6, pp. 1973–1982, 2005.

[12] C. Fragouli and E. Soljanin, "Information flow decomposition for network coding," *IEEE Trans. Information Theory*, vol. 52, pp. 829–848, March 2006.

[13] Z. Li, B. Li, L. C. Lau, "On achieving optimal multicast throughput in undirected networks," In *Joint Special Issue on Networking and Information Theory, IEEE Trans. Imform. Theory (IT) and IEEE/ACM Trans. Networking* (TON), vol. 52, June 2006.

[14] A. Agarwal and M. Charikar, "On the advantage of network coding for improving network throughput," *IEEE Information Theory Workshop*, San Antonio, Texas, 2004.

[15] C. Chekuri, C. Fragouli and E. Soljanin, "On average throughput benefits and alphabet size for network coding," *joint special issue of the IEEE Trans. Imform. Theory and the IEEE/ACM Trans. Networking*, vol. 52, pp. 2410–2424, June 2006.

[16] A. Lehman and E. Lehman, "Complexity classification of network information flow problems", in *Proc. of 15th Symposium on Discrete Algorithms* (SODA'04), New Orleans, LA, January 2004.

[17] M. Langberg, A. Sprintson, J. Bruck, "The encoding complexity of network coding," *Joint special issue of the IEEE Trans. Imform. Theory and the IEEE/ACM Trans. Networking*, vol. 52, pp. 2386–2397, 2006.

[18] ´A. M. Barbero, and Ø. Ytrehus, "Cycle-logical treatment for cyclopathic networks," *Joint special issue of the IEEE Trans. Imform. Theory and the IEEE/ACM Trans. Networking*, vol. 52, pp. 2795–2804, June 2006.

[19] C. Gkantsidis and P. Rodriguez, "Network coding for large scale content distribution," in *Proceedings of Infocom*. IEEE, 2005.

[20] C. Gkantsidis, J. Miller, and P. Rodriguez, "Comprehensive view of a live network coding p2p system," in IMC '06: *Proceedings of the 6th ACM SIGCOMM on Internet Measurement*. New York, NY, USA: ACM Press, 2006, pp. 177–188.

[21] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: practical wireless network coding," in *SIGCOMM*. Pisa, Italy: ACM, September 2006, pp. 243–254.

[22] M. Adler, N. J. A. Harvey, K. Jain, R. Kleinberg, and A. R. Lehman, "On the capacity of information networks," *Proc. of 17th Symposium on Discrete Algorithms* (SODA'06), Miami, FL, January 2006.

[23] G. Kramer and S. A. Savari, “Edge-cut bounds on network coding rates,” *Journal of Network and Systems Management*, vol. 14, 2006.

[24] S. Riis, “Linear versus non-linear boolean functions in network flow,” *CISS*, 2004.

[25] S. Deb, M. Medard, and C. Choute, “Algebraic gossip: a network coding approach to optimal multiple rumor mongering,” *IEEE/ACM Transactions on Networking*, vol. 14, pp. 2486 –2507, June 2006.

[26] D. Mosk-Aoyamam and D. Shah, “Information dissemination via network coding,” *ISIT*, pp. 1748–1752, 2006.

[27] D. S. Lun, N. Ratnakar, M. M´edard, R. K¨otter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, “Minimum-cost multicast over coded packet networks,” I*EEE Trans. Imform. Theory*, vol. 52, pp. 2608-2623, June 2006.

[28] Y. Cassuto and J. Bruck, “Network coding for nonuniform demands,” in *Proc. 2007 IEEE Int. Symp. Inform. Theory* (ISIT’07), Adelaide, Australia, Sept. 2005.

[29] C. Fragouli, C. Chekuri and E. Soljanin, “Achievable information rates in single-source non-uniform demand networks,” in P*roc. 2006 IEEE Int. Symp. Inform. Theory* (ISIT’06), Seattle, WA, USA., July 2006.

[30] A. R. Lehman and E. Lehman, “Network coding: does the model need tuning?” in *Proc. of 16th Symposium on Discrete Algorithms* (SODA’05), Vancouver, Canada, January 2005.

[31] R. Dougherty, C. Freiling, and K. Zeger, “Unachievability of network coding capacity,” *IEEE Transactions on Information Theory and IEEE/ACM Transactions on Networking*, vol. 52, pp. 2365–2372, June 2006.

[32] T. Ho, M. Medard, M. Effros, and R. Koetter, “Network coding for correlated sources,” *38th Annual Conference on Information Sciences and Systems* (CISS’04), Princeton, NJ, March 2004.

[33] M. Jafarisiavoshani, C. Fragouli, and S. Diggavi, “ Bottleneck discovery and overlay management in network coded Peer-to-Peer systems”, *ACM SigComm Workshop on Internet Network Management* (INM), 2007.

[34] Y. Wu, P. A. Chou, and S.-Y. Kung, “Minimum-energy multicast in mobile ad hoc networks using network coding,” *IEEE Trans. on Communications*, vol. 53, no. 11, pp. 1906–1918, November 2005.

[35] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, “A network coding approach to energy efficient broadcasting: from theory to practice,” in *IEEE Infocom*, Barcelona, Spain, April 2006.

[36] S. Zhang, S. Liew, and P. Lam, “Physical layer network coding,” *ACM MobiCom 2006*, pp. 24–29, September 2006.

[37] S. Avestimehr, S. Diggavi and D. Tse, “A d eterministic model for wireless relay networks and its capacity”, in *Proc. 2007 IEEE Int. Workshop Inform. Theory* (ITW’07), Bergen, Norway, July 2007.

[38] S. Avestimehr, S. Diggavi and D. Tse, “Wireless network information flow,” in *Proc. 45th Annual Allerton Conference*, Monticello, IL, Sept. 2007.

[39] Y. E. Sagduyu and A. Ephremides, “Joint scheduling and wireless network coding,” *Network Coding Workshop*, 2005.

[40] A. Erylmaz, A. Ozdaglar, and M. M´edard, “On delay performance gains from network coding,” *CISS*, 2006.

[41] D. Petrovi´c, K. Ramchandran, and J. Rabaey, “Overcoming untuned radios in wireless networks with network coding,” *IEEE/ACM Transactions on Networking*, vol. 14, pp. 2649–2657, June 2006.

[42] A. Kamra, V. Misra, J. Feldman and D. Rubenstein, “Growth codes: maximizing sensor network data persistence”, *SigComm* 2006.

[43] Y. Lin, B. Liang, and B. Li, “Data persistence in large-scale sensor networks with decentralized fountain codes,” in the *Proceedings of the 26th IEEE INFOCOM 2007*, Anchorage, Alaska, May 6-12, 2007.

[44] A. G. Dimakis, P. B. Godfrey, M. Wainwright, and K. Ramchandran, “Network coding for peer-to-peer storage,” in the *Proceedings of the 26th IEEE INFOCOM 2007*, Anchorage, Alaska, May 6-12, 2007.

[45] D. Lun, M. Médard, and M. Effros, “On coding for reliable communication over packet networks,” *Proc. Allerton Conference on Communication, Control, and Computing*, September-October 2004.

[46] R. W. Yeung and N. Cai, “Network error correction, I: basic concepts and upper bounds,” *Commun. Inf. Syst.*, vol. 6, pp. 19–35, 2006.

[47] N. Cai and R. W. Yeung, “Network error correction, II: lower bounds,” *Commun. Inf. Syst.*, vol. 6, pp. 37–54, 2006.

[48] D. S. Lun, P. Pakzad, C. Fragouli, M. Mdard, and R. Koetter, “An analysis of finite-memory random linear coding on packet streams,” *WiOpt ’06*, April 2006.

[49] U. Niesen, C. Fragouli and D. Tunineti, “On capacity of line networks”, *IEEE Transactions on Information Theory*, November 2007.

[50] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, “Capacity of wireless erasure networks”, IEEE Transactions on Information Theory 2004. In review.

[51] S. El Rouayheb, A. Sprintson, and C. N. Georghiades “Simple network codes for instantaneous recovery from edge failures in unicast connections,” in *2006 Workshop on Information Theory and its Applications* (ITA’06) San Diego, California, February, 2006.

[52] Z. Zhang, “Network error correction coding in packetized networks,” in *Proc. 2006 IEEE Int. Workshop Inform. Theory* (ITW’06), Chengdu, China, October 2006.

[53] R. W. Yeung and N. Cai, “Secure network coding,” in P*roc. 2002 IEEE Internat. Symp. Inform. Th.* (ISIT’02), Lausanne, Switzerland, June 2002.

[54] J. Feldman, T. Malkin, C. Stein, and R. A. Servedio, “On the capacity of secure network coding,” in *Proc. 42nd Annual Allerton Conference on Commun., Control, and Comput.*, September 2004.

[55] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, and D. Karger, “Byzantine modification detection in multicast networks using randomized network coding,” in *Proc. 2004 IEEE Internat. Symp. Inform. Th.* (ISIT’04), June 2004.

[56] K. Bhattad and K. R. Narayanan, “Weakly secure network coding,” in *Proc. First Workshop on Network Coding, Theory, and Applications* (NetCod’05), April 2005.

[57] D. Charles, K. Jain, and K. Lauter, Signatures for network coding. In *Proceedings of Conference on Information Sciences and Systems*, invited paper, 2006.

[58] S. El Rouayheb and E. Soljanin, “On wiretap networks II,” *2007 IEEE Int. Symp. Inform. Theory* (ISIT’07), Nice, France, June, 2007.

[59] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard, “Resilient network coding in the presence of byzantine adversaries,” *Infocom*, pp. 616–624, 2007.

[60] R. Koetter and F. Kschischang, “Coding for errors and erasures in random network coding,” *ISIT*, June 2007.

[61] C. Fragouli and A. Markopoulou, “A network coding approach to network monitoring,” *Allerton*, 2005.

[62] C. Fragouli, A. Markopoulou, and S. Diggavi, “Active topology inference using network coding,” *Allerton*, 2006.

[63] T. Ho, B. Leong, Y. Chang, Y. Wen, and R. Koetter, “Network monitoring in multicast networks using network coding,” *International Symposium on Information Theory* (ISIT), June 2005.

[64] M. Jafarisiavoshani, C. Fragouli, and S. Diggavi, “On subspace properties for randomized network coding,” in *Proc. 2007 IEEE Int. Workshop Inform. Theory* (ITW’07), Bergen, Norway, July 2007.

[65] J. Sundararajan, S. Deb, and M. M´edard, “Extending the birkhoff-von neumann switching strategy to multicast switches,” *Proc. of the IFIP Networking 2005 Conference*, May 2005.

[66] M. Kim, J. Sundararajan, and M. M´edard, “Network coding for speedup in switches,” *ISIT*, June 2007.

[67] J. Sundararajan, M. M´edard, M. Kim, A. Eryilmaz, D. Shah, and R. Koetter, “Network coding in a multicast switch,” *Infocom*, March 2007.

[68] N. Jayakumar, K. Gulati, and S. K. A. Sprintson, “Network coding for routability improvement in VLSI,” *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2006.

[69] K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita, “Quantum network coding for general graphs,” 2006. [Online]. Available: http://arxiv.org/abs/quant-ph/ 0611039

[70] Y. Shi and E. Soljanin, “On multicast in quantum networks,” in *Proc. 40th Annual Conference on Information Sciences and Systems (CISS’06)*, Princeton, NJ, March 2006.

[71] D. Leung, J. Oppenheim, and A. Winter, “Quantum network communication - the butterfly and beyond,” 2006. [Online]. Available: http://arxiv.org/abs/quant-ph/0608223

Network Source Coding: A Perspective

Plenary talk presented on Monday June 25, 2007 at the 2007 IEEE International Symposium on Information Theory, Nice France

Michelle Effros

Abstract—**In my ISIT 2007 plenary talk [1], I gave a brief introduction to the field of network source coding and then shared a bit of personal perspective on this area of research. Network source coding is a field at once marked by pitfalls and promise. The field’s promise arises from its obvious potential for enormous performance gains and the abundance of applications for these advances. The potential pitfalls are made evident by the monumental gap between what we understand today and what we must come to master if we hope to realize the field’s full promise. This paper gives a brief overview of some of the ideas raised in that talk. Here like there, I begin with a brief introduction to the field of network source coding, defining the problem, noting the role played by the network, and briefly commenting on the issue of separation. The history of the field is broad and deep and worthy of far more attention than I was able to give it in the scope of my talk; once again, I give only the briefest hint of the history of the field—attempting to impart a bit of the area’s flavor without attempting to catalog the many beautiful contributions that fill decades’ worth of literature on this fascinating topic. Finally, I suggest that the central questions on which we, as a community, have historically focused may actually be impeding our progress, and I suggest a few alternative avenues for future advance.**

I. Introduction

Source coding is the science of efficient information representation. In the source coding problem defined by Shannon [2] (see Figure 1), an encoder observes samples $X_1, X_2, \ldots$ from some random source X and describes them at an expected rate of R bits per sample to the system decoder. The decoder uses the source description to build reconstructions $\hat{X}_1, \hat{X}_2, \ldots$ of the original source sequence. In lossless source coding, the code designer hopes to achieve the minimal rate R for which the code’s error probability $\Pr(\hat{X}^n \neq X^n)$ can be made arbitrarily small as n grows without bound. In lossy source coding, the objective is to achieve the minimal rate R for which the per-symbol expected distortion $\frac{1}{n}Ed(X^n, \hat{X}^n)$ asymptotically meets some distortion constraint D as n grows without bound.

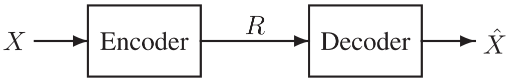


Figure 1. Shannon’s source coding problem.

The field of network source coding treats the data compression problem for networks beyond the point-to-point network introduced by Shannon. These include networks with multiple transmitters, multiple receivers, decoder side information, intermediate nodes, and any combination of these features, as illustrated in Figures 2–5. Thus the field of network source coding attempts to bridge the gap between the point-to-point network studied by Shannon and today’s more complicated network environments.

The potential performance benefits of network source codes arise from a number of factors. Like point-to-point source

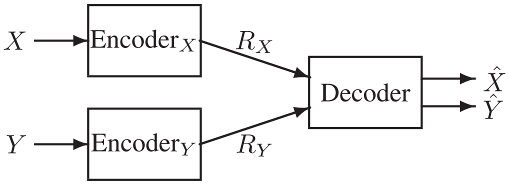


Figure 2. A source coding problem for a network with multiple transmitters.

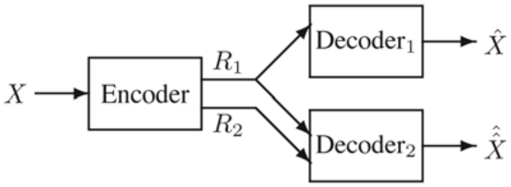


Figure 3. A source coding problem for a network with multiple receivers.

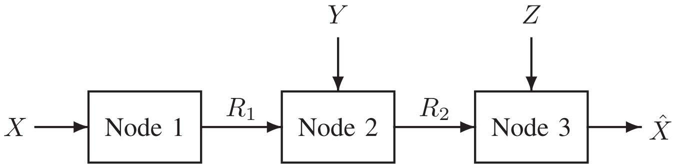


Figure 4. A source coding problem for a network with intermediate nodes.

codes, network source codes can achieve some rate savings by exploiting features such as non-uniformity of source statistics, memory in the source distribution, and economies of scale. Other potential savings arising in network source coding arise only when the network is generalized beyond Shannon's original point-to-point network model. Three examples of such factors follow.

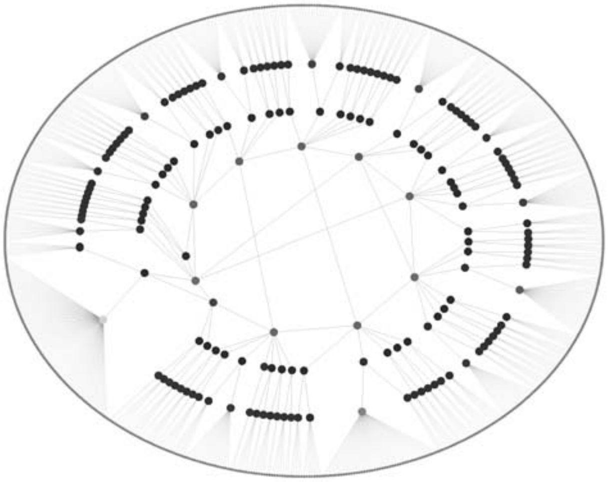


Figure 5. A network with multiple transmitters, multiple receivers, and intermediate nodes. This illustration of the internet from [3] appears by permission of Lun Li.

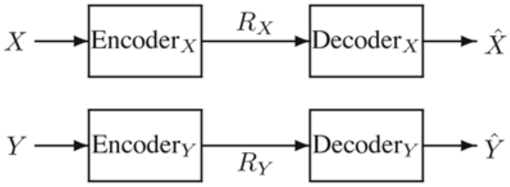


Figure 6. The rate achievable by a pair of independent source codes for *X* and *Y* provides an upper bound on the achievable rate region for the network in Figure 2.

A. The Network Source Coding Advantage

Source Dependence

Consider the network shown in Figure 2. The two encoders observe sources X and Y, respectively, and describe them to a shared decoder. Sources X and Y are dependent random variables, and their joint distribution is known. There is no communication between the two encoders. If we treat the problems of describing X and describing Y independently, then we can achieve the given demands using a pair of conventional (point-to-point) source codes, as shown in Figure 6. The resulting achievable rate region is $\{(R_X, R_Y) : R_X \geq H(X), R_Y \geq H(Y)\}$, giving the independent coding upper bound shown in Figure 7. In contrast, joint coding of sources X and Y using a single encoder and a single decoder, as shown in Figure 8, gives the joint coding lower bound $R_X + R_Y \geq H(X, Y)$ in Figure 7. Finally, the optimal achievable region for the given multiple access source coding problem is the region

$$\begin{aligned}\{(R_X, R_Y) : \quad & R_X \geq H(X|Y), \quad R_Y \geq H(Y|X)\\ & R_X + R_Y \geq H(X, Y)\}\end{aligned}$$

derived by Slepian and Wolf in [4]. This example demonstrates that network source codes can take advantage of the statistical

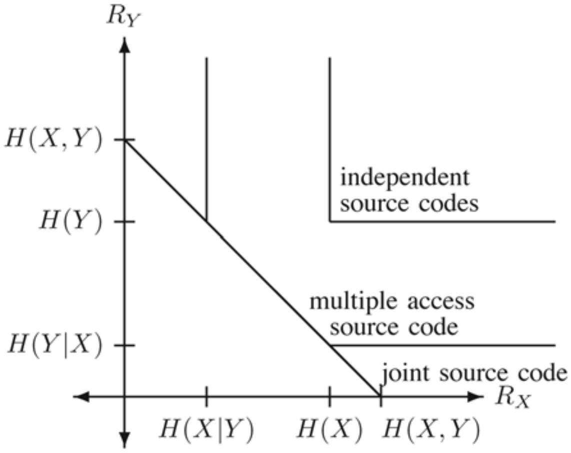


Figure 7. The optimal achievable rate region for the multiple access network shown in Figure 2 and its independent coding upper bound and joint coding lower bound.

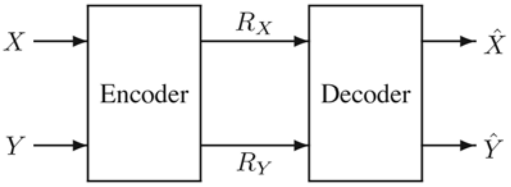


Figure 8. The rate achievable by a single encoder provides a lower bound on the achievable rate region for the network in Figure 2.

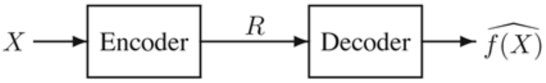


Figure 9. A functional source coding problem for Shannon's point-to-point network.

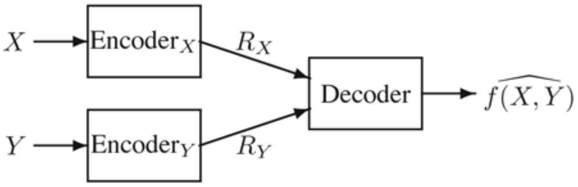


Figure 10. A functional source coding problem for a network with multiple transmitters.

dependence of sources available to distinct nodes in a network. The gap between the optimal network source code and the optimal application of conventional codes can be arbitrarily large.

The ability to exploit source dependence may be particularly useful for applications such as sensor networks, video conferencing, and shared gaming environments where statistical dependence among distributed source observations is likely to occur.

Functional Demands

The functional source coding problem arises when the system decoder is interested not in the sources observed by other nodes in the network but instead in some function of those sources. Functional source coding does not really arise in Shannon's point-to-point network, since meeting functional demands is trivial in this case. For example, Figure 9 shows a point-to-point source coding problem with a decoder that wishes to reconstruct a function $f(X)$ of the source X observed at the encoder. The solution to this problem is immediate since the encoder can simply calculate $f(X)$ and describe it to the decoder using conventional tools.

The problem becomes more interesting in network systems. For example, consider the network shown in Figure 10. Two encoders observe sources X and Y, respectively. The decoder wishes to reconstruct some function $f(X, Y)$ of the source observations. Unlike in the point-to-point case, neither encoder can calculate the desired function. For simplicity, we reduce the given example to a single rate of interest by setting rate R_Y to $H(Y)$. In this case, the network effectively becomes the side information case shown in Figure 11, where a single encoder observes source X and describes it to a decoder with side information Y. The decoder wishes to reconstruct a function $f(X, Y)$ of both sources.

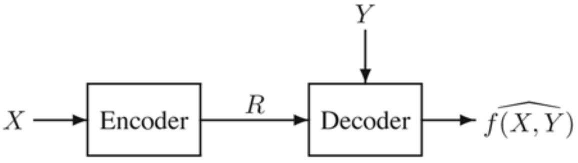


Figure 11. A functional source coding problem for a source coding network with decoder side information.

Results for the functional source coding problem for the point-to-point network with decoder side information include bounds for the lossy [5], [6] and lossless [7] cases. I use an example from [6] to illustrate the potential gains achievable with functional source coding. Let X and Y be independent and uniformly distributed on $\{0, \ldots, N-1\}$ and $\{0, 1\}$, respectively. The desired function is $f(x, y) = (x + y) \mod 2$. We wish to calculate the rate-distortion bound for the given source using the Hamming distortion measure $d_H(z, \hat{z})$ which equals 0 when $\hat{z} = z$ and 1 otherwise. The choice of independent sources for this example is made deliberately to separate the effect of source dependence illustrated above from the effect of functional coding. The potential benefit of functional source coding on dependent sources is even more pronounced.

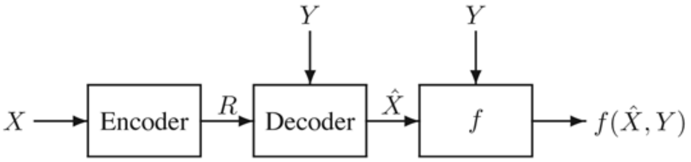


Figure 12. The rate-distortion region for a network with functional demands achieved through conventional source coding techniques provides an upper bound on the functional rate-distortion bound.

Current systems typically meet functional source coding demands using conventional source coding methods. As shown in Figure 12, the encoder first describes source X to the decoder, which builds source reconstruction $\hat{X}$. Applying the desired function to the source reconstruction and available side information gives functional reconstruction $f(\hat{X}, Y)$. (The use of decoder side information is typically characterized as network source coding rather than conventional source coding, so even the given system is more sophisticated than a true conventional coding approach. In this example, however, they are identical since X and Y are independent by assumption.) The resulting rate-distortion region is

$$R_{\hat{X}|(Y)}(f, D) = \log N - \alpha_N D \log(N-1) - H(\alpha_N D)\,,$$

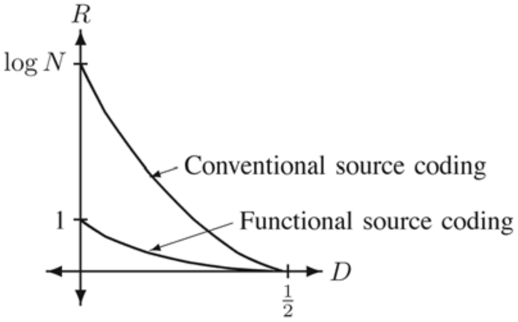


Figure 13. The rate-distortion region for a functional source coding problem with independent sources X uniformly distributed on $\{0, \ldots, N-1\}$, and Y uniformly distributed on $\{0, 1\}$, and function $f(x, y) = (x + y)_{\text{ mod } 2}$ and the rate-distortion region achievable without functional coding.

where $\alpha_N = 2N/(N+1)$ if N is odd, $\alpha_N = 2(N-1)/N$ if N is even, and $H(p) = -p \log p - (1-p)\log(1-p)$ is the binary entropy function. This result gives an upper bound on the desired functional rate-distortion region, as shown in Figure 13. The functional rate-distortion function $R_{X|(Y)}(f, D)$ is

$$R_{X|(Y)}(f, D) = \begin{cases} H((N-1)/(2N)) - H(D) & \text{if } N \text{ is odd} \\ 1 - H(D) & \text{if } N \text{ is even.} \end{cases}$$

When $R = 0$, both codes achieve an expected distortion of 1/2. When $D = 0$, the gap between the two curves is large since the conventional code must describe X losslessly at rate $\log N$ while the functional code need only to send a lossless description of whether X is even or odd. Again, the performance gap between the two strategies can be made arbitrarily large.

The performance benefit arising from functional demands is likely to be useful in many measurement scenarios where we are interested in gathered data for the functions that they allow us to compute rather than for their own sake.

Resource Sharing

A third benefit of network source codes arises from the ability for distinct users to share resources in a network environment. The network source coding example of Figure 14 shows one example of a network where this phenomenon arises. In this example, a single encoder observes sources X and Y and broadcasts a single description of the pair to one decoder who observes Y and wishes to reconstruct X and another decoder who observes X and wishes to reconstruct Y. For the sake of example, let X and Y be independent and uniformly distributed on the alphabet $\{0, \ldots, N-1\}$. Once again, the choice of independent sources and non-functional demands is made to separate the benefits of careful resource sharing from the earlier phenomena.

The optimal conventional lossless source coding strategy for this scenario describes source X to the first decoder at rate $R_X \geq H(X) = \log N$ and source Y to the second decoder at rate $R_Y \geq H(Y) = \log N$. (Recall that the sources are independent by assumption, so the side information at the receivers is not useful for this example.) The resulting rate for applying this pair of codes is

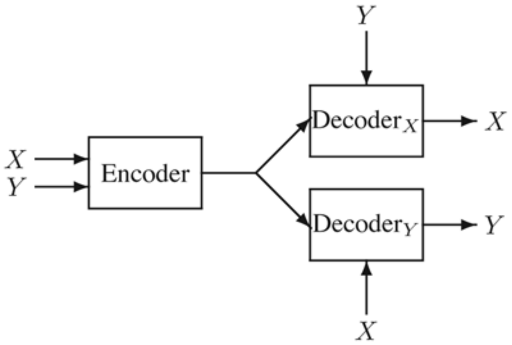


Figure 14. A broadcast source coding problem for a network with two receivers.

$$R_X + R_Y \geq 2 \log N\,.$$

In contrast, applying a network source code to describe the pair of sources simultaneously requires only rate

$$R \geq \log N.$$

This result may be derived from prior results in a number of different ways. For example, the problem can be viewed as a multicast problem and solved using [8]. The given bound can be implemented by an encoder that describes $(X + Y)_{\text{ mod } N}$ at rate $\log N$. Each receiver can then recover its desired source using its available side information.

Again, the rate benefit for applying a network source code rather than a pair of independent source codes can be made very large even when the sources are independent. This benefit becomes even more pronounced when combined with the other benefits described above. For example, when X and Y are dependent, viewing this example as a multicast problem and applying [9] gives the optimal rate region

$$R \geq \max\{H(X|Y), H(Y|X)\}.$$

This contrasts with the rate bounds

$$R_X + R_Y \geq H(X|Y) + H(Y|X)$$

achieved by a pair of independent Slepian-Wolf codes and

$$R_X + R_Y \geq H(X) + H(Y)$$

achieved by a pair of independent conventional source codes.

B. Separation

While the preceding examples illustrate the potential benefit of network source codes, it is important to note that even optimal network source codes achieve suboptimal communication performance in some networks. The problem, of course, is separation. Shannon's 1948 proof that optimal source coding followed by optimal channel coding achieves the optimal end-to-end communication performance in a point-to-point network separated the fields of source and channel coding from their very inception. Yet separation does not necessarily hold in network systems, as the following example from [10] illustrates.

Consider a pair of sources $(U_1, U_2) \in \{0, 1\}^2$ with probability mass function $p(u_1, u_2) = 1/3$ for all $(u_1, u_2) \in \{(0, 0), (0, 1), (1, 1)\}$ and $p(1, 0) = 0$. We wish to losslessly describe the given source across an additive multiple access channel with channel inputs $X_1, X_2 \in \{0, 1\}$ and channel output $Y = X_1 + X_2 \in \{0, 1, 2\}$. Figure 15 shows the Slepian-Wolf rate region and multiple access capacity region for the given source and channel. Since the regions do not overlap, the given source cannot be reliably transmitted across the given channel using an optimal multiple access source code followed by an optimal multiple access channel code. Yet transmitting the source directly across the channel—with no source or channel coding—does achieve reliable communication. As a result, this example demonstrates the failure of separation in the given network.

In some sense, the observation that Shannon's separation of source and channel coding fails to generalize to networks calls into question the fields of network source and channel coding.

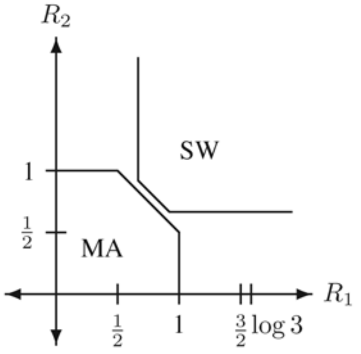


Figure 15. The Slepian-Wolf rate region and multiple access capacity region for sending a pair of sources (U_1, U_2) uniformly distributed across alphabet $\{(0, 0), (0, 1), (1, 1)\}$ across an additive multiple access channel with inputs $X_1, X_2 \in \{0, 1\}$ and output $Y = X_1 + X_2 \in \{0, 1, 2\}$. Since the regions do not overlap, the given source cannot be reliably transmitted across the given channel using an optimal multiple access source code followed by an optimal multiple access channel code. Since transmitting the source directly across the channel—with no source or channel coding—achieves reliable communication, this example illustrates the failure of separation in the given network.

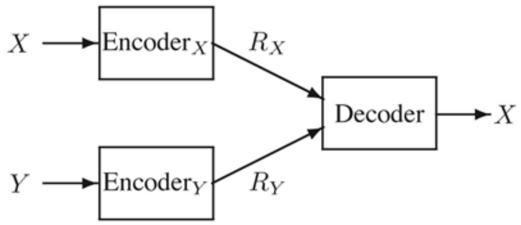


Figure 16. The coded side information problem of Ahlswede and Körner.

Why do we pursue rate regions and optimal codes for network source codes and network channel codes if together the optimal codes do not guarantee the optimal performance across our networks? While a complete understanding of the question of separation in network systems is not yet available, several observations motivate the continued investigation of these separate disciplines in addition to the study of optimal joint codes.

First, the failure of separation in one network does not imply the failure of separation in all networks. Even a simple change in the above example to replace the channel's integer sum with a modulo two sum yields a channel for which separation holds no matter what the source statistic [11], [12]. In fact, [11] demonstrates that it is easy to construct both channels for which separation holds for all sources, sources for which separation holds for all channels, and source-channel pairs for which neither of the above statements holds and yet separately designed source and channel codes achieve the optimal performance. The same paper also investigates the penalty associated with using separately designed source and channel codes in some examples where separation fails.

Second, rate has long served as the central point of negotiation between communication network providers and the user applications that we hope to transmit across these networks. While true joint codes and their performance are an important topic for our understanding of the limits of communication system performance, it seems unrealistic to expect very complicated exchanges of complete source and channel statistics and specialized code design for each new source we hope to transmit across an existing network. Until other similarly simple characterizations become available, rate seems a likely point of negotiation between the application and the provider for practical network applications.

C. A Hint of History

While a thorough overview of the history of network source coding is beyond the scope of this paper, a brief comment on the flavor of that work is useful for the discussion that follows. While there are notable exceptions, the treatment of network source coding in the information theory literature has focused primarily on finding rate regions and has often advanced one network at a time. For example, in the lossless network source coding literature, Slepian and Wolf's seminal work on the rate region of the multiple access network [4] was followed closely by Ahlswede and Körner's achievable rate region for the coded sided information problem (see Figure 16) [13] and Wyner's rate region for a generalization on that theme (see Figure 17) [14]. Progress in the

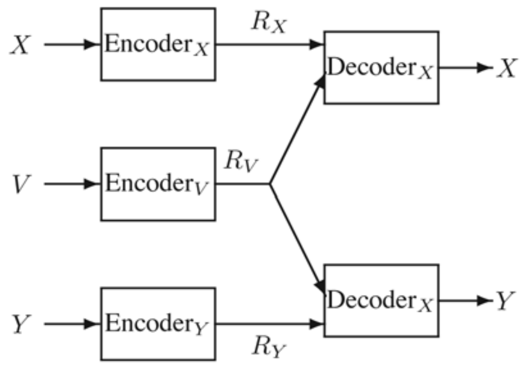


Figure 17. The coded side information problem of Wyner.

lossy source coding literature has likewise occurred through the study of individual examples. While the literature contains some attempts at understanding the source coding problem for more general classes of networks (see, for example, [15]), even three-node networks remain incompletely characterized to this day. Comparing our rate or progress with the growth of common networks like the internet (see Figure 18) reveals a disturbing but inescapable trend: the problem is getting away from us.

II. Alternative Questions

As the gap between the networks that we consider solved and the networks that we use in practice grows, so too does the urgency for reexamining our approach. While many possible questions may arise, I touch upon only three. First, the difficulty of developing systematic techniques for deriving rate regions for arbitrary networks makes it worth considering whether rate regions are really the direction most worthy of our attention. Second, if we continue the pursuit of this traditional goal, it may be helpful to consider whether new rate regions are the only form of progress in this pursuit. Finally, if deriving new rate regions remains the objective, it may be instructive to investigate whether precise calculations are the only provably good solutions. The following sections touch on each of these questions in turn.

A. Rate Region Alternatives

In the information theory literature, a large fraction of the research devoted to network source coding focuses on the derivation of rate regions. (I here use "rate regions" to refer to both optimal achievable rate regions for lossless codes and optimal achievable rate-distortion regions for lossy codes.) These bounds present a number of challenges. First, they are extremely difficult to derive. While some basic lessons have been learned through the study of special cases, solving even small new networks often requires the derivation of fundamentally new techniques specially designed for the current network scenario. A systematic approach to deriving rate regions for large networks remains elusive.

Second, even "solved" rate regions are difficult to characterize. Information theoretic characterizations of rate regions are often given in the form of optimization problems. For example, Shannon's rate-distortion region involves a minimization of mutual information over all conditional distributions that meet a given distortion constraint. Similarly, Ahlswede and Körner's characterization of the rate region for the coded side information problem illustrated in Figure 16 involves finding the optimal tradeoff between rates $R_X = H(X|U)$ and $R_Y = I(Y;U)$ over all auxiliary random variables U for which $X \rightarrow Y \rightarrow U$ forms a Markov chain [13]. The optimization problems for different rate regions are different, and finding their numerical characterizations even for simple source distributions can be surprisingly difficult (see, for example, [17]–[19]). Yet without such numerical results even "solved" rate regions are difficult to apply in practice.

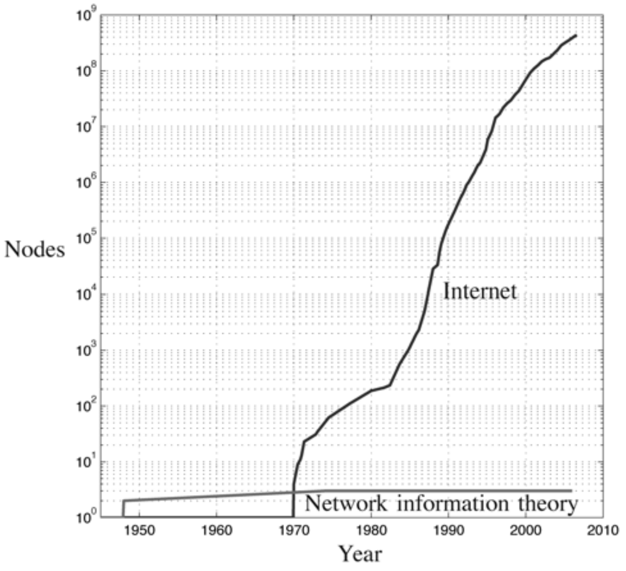


Figure 18. The given graph shows estimates of the number of nodes in the internet [16] and (a generous estimate of) the size of networks that are well understood in network information theory both as a function of time. While the figure gives network information theory credit for mastering the three-node network, several key three-node networks remain incompletely understood in the network information theory literature.

Finally, even if we could derive and calculate achievable rate regions for the networks of interest, the feasibility of working with such solutions for large networks is questionable. For example, the rate region for the network shown in Figure 5 would be an undoubtedly complex subspace of a very high-dimensional space. (The dimension of the space equals the number of edges in the graph.) Even if we could fully characterize such a region, working with that characterization in practice would be enormously difficult.

Faced with these difficulties, it is useful to consider alternatives to the traditional rate-region objective. One such alternative has seen spectacular success in the field of network coding introduced in [8]. In network coding, we typically begin with a graph of lossless, capacitated links. Some nodes are labeled as source nodes, and the sources available to those nodes are given. Other nodes are labeled as receivers, and the sources demanded by those

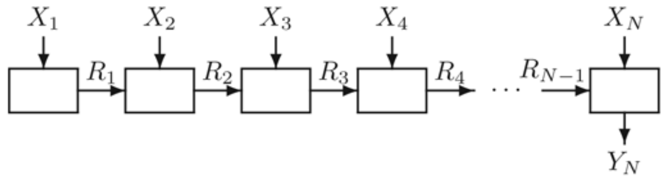


Figure 19. A single-demand line network.

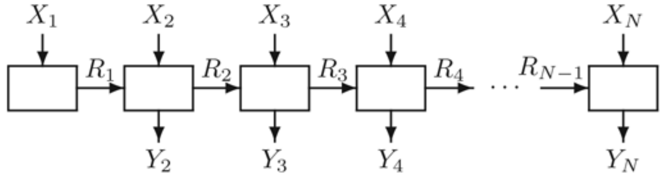


Figure 20. A multi-demand line network.

nodes are also given. While the network coding literature often focuses on the special case where all sources are independent and uniformly distributed and all receivers demand all sources (called *multicast* coding), generalizations to dependent sources [9] and arbitrary demands are also topics of significant interest.

The question asked in network coding is whether (or under what conditions) we can build a code that meets the given demands. While the derivation of a rate region yields a complex subset of a space whose dimension equals the number of links in the network, the *supportability* question raised by network coding never requires that we list all rates over all links that allow us to meet the given demands. Instead, at least for the multicast demand structure, the conditions for supportability are extremely simple, and practical code design to asymptotically achieve the performance limit are also well understood. (See, for example, [9], [20], [21].) The shift from rate regions to questions of supportability is only one of many factors that contribute to the success of network coding in making precise information theoretic statements about even very large networks. A similar shift in mindset may provide a useful tool for continuing progress in other network source coding problems.

B. Equivalence Classes

In addition to pursuing alternatives to rate regions, it is instructive to consider other forms of progress that may be useful when rate regions are desired but difficult to derive. Understanding the relationships between network source coding problems is one potential way to make progress. The idea here is to demonstrate that the rate regions for one family of network source coding problems can be derived from the rate regions for a more limited family of networks. For example, [15] shows that solving the rate regions for the family of "normal source networks" yields rate regions for a broader family of networks. Similarly, [22] demonstrates that finding solutions for the family of multiple unicast network coding problems yields solutions to network coding problems with arbitrary demands. Likewise, [23] proves that for certain families of joint source distributions, solutions for the family of network source coding problems on single-demand line networks (see Figure 19) yields solutions for general line networks (see Figure 20). Each of these examples demonstrates relationships between network source coding problems even in cases where the underlying rate regions remain unsolved. The given results are useful both because they direct our attention to the most central problems in the field and because they allow us to apply bounds derived for one problem to a whole family of related problems.

C. Approximation Algorithms

The preceding sections discuss alternatives to rate regions and means of gaining insight into relationships between rate regions While such alternative paths are important, the question of how to characterize rate regions if and when they are desired remains. Like many mathematical disciplines, the field of information theory has focused primarily on complete and precise solutions. A problem is considered solved only when precise and matching upper and lower bounds are demonstrated. Yet a variety of stumbling blocks arise in finding rate regions for sources encountered in practice. Since real sources often have complicated memory structures, single letter characterizations often do not exist. Source distributions are typically unknown, leaving source samples as a starting point for rate region characterization. Finally, as noted previously, the rate region is unknown for many networks and prohibitively difficult to calculate even when known. Combining these many factors leads to the current state of affairs: numerical characterizations of rate regions for real sources and networks are almost non-existent.

The question, then, is how to achieve numerical bounds for real sources and real networks. Approximation algorithms may be a useful tool for just such progress. I illustrate that point by way of an example.

The *nth order operational rate-distortion region* describes the optimal performance theoretically achievable by an n-dimensional vector quantizer. When $n = 1$, globally optimal solution for the nth order operational rate-distortion function is computationally feasible [24]–[27]. For $n > 1$, the problem is NP-hard [28], and the information theory literature has focused instead on locally optimal solutions achieved by iterative descent techniques [29]–[32].

Approximation algorithms present a potentially useful alternative for characterizing operational rate-distortion regions. Let $D_n(R)$ be the nth order operational distortion-rate region for a point-to-point network. A $(1+\epsilon)$ approximation algorithm is a code design algorithm that guarantees a rate-R code achieving distortion $\hat{D}_n(R) \in [D_n(R), (1+\epsilon)D_n(R)]$ for any $R \geq 0$. Thus approximation algorithms yield solutions that are not necessarily perfect but are provably good. Since the guarantee holds even when $D_n(R)$ is unknown, approximation algorithms are useful both for code design and for approximating the operational distortion-rate region.

A variety of low-complexity approximation algorithms for lossy coding in the point-to-point network appear in the literature. Among them are techniques ranging from randomized sub-sampling of the source distribution [33] to deterministic construction of a set of candidate codewords [34]. The approach introduced in

[35], [36] gives a deterministic algorithm for $(1+\epsilon)$ optimal code design. The given algorithm is the only approximation algorithm that we know of that applies not only to the point-to-point network but also to a more general family of network source coding problems—including a variety of problems for which asymptotic rate regions are currently unknown.

Approximation algorithms also present a low-complexity alternative to algorithms like the Blahut-Arimoto [37], [38] for rate region calculation. An example of such an approach for the coded side-information problem appears in [39].

III. Summary and Conclusions

The widening chasm between network source coding theory and real network environments provides a call to action for groundbreaking new work in the field of network source coding. Looking back on the history of the field to date, it seems likely that the greatest strides forward will come not from developing new techniques to solve old problems, but from questioning the questions themselves. Supportability, equivalence classes, and approximation algorithms are just three of the many possible new directions for bringing the field forward from Slepian-Wolf to the internet.

References

[1] M. Effros, "Network source coding: Pipe dream or progress," In *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, June 2007. IEEE.

[2] C.E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[3] L. Li, *Topologies of complex networks: Functions and structures*. Ph.D. Dissertation, California Institute of Technology, Pasadena, CA, 2007.

[4] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. IT-19, pp. 471–480, 1973.

[5] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," *IEEE Transactions on Information Theory*, vol. IT-28, no. 5, pp. 1788–1791, Sept. 1982.

[6] H. Feng, M. Effros, and S.A. Savari, "Functional source coding for networks with receiver side information," In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, pp. 1419–1427, Monticello, IL, Sept. 2004. IEEE.

[7] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[8] R. Ahlswede, N. Cai, S.-Y. R. Li, and R.W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. IT-46, no. 4, pp. 1204–1216, July 2000.

[9] T. Ho, M. Médard, R. Koetter, D.R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, October 2006.

[10] T.M. Cover, "Multiple access channels with arbitrarily correlated sources," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 648–657, 1980.

[11] S. Ray, M. Effros, M. Médard, T. Ho, D. Karger, R. Koetter, and J. Abounadi, "On separation, randomness, and linearity for network codes over finite fields, Technical Report 2687, MIT LIDS, 2006.

[12] S. Ray, M. Médard, M. Effros, and R. Koetter, "On separation for multiple access networks," In *Information Theory Workshop*, pp. 399–403, Chengdu, China, October 2006. IEEE.

[13] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Transactions on Information Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.

[14] A.D. Wyner, "On source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-21, no. 3, pp. 294–300, Nov. 1975.

[15] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Transactions on Information Theory*, vol. IT-26, no. 2, pp. 155–165, Mar. 1980.

[16] R.H. Zakon, Hobbes' internet timeline, http://www.zakon.org/robert/internet/timeline, Nov. 2006.

[17] H.S. Witsenhausen and A.D. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 493–501, 1975.

[18] D. Marco and M. Effros, "A partial solution for lossless source coding with coded side information," In *Proceedings of the Information Theory Workshop*, Punta del Este, Uruguay, March 2006. IEEE. Invited paper.

[19] W. Gu, R. Koetter, M. Effros, and T. Ho, "On source coding with coded side information for a binary source with binary side information," In *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, June 2007. IEEE.

[20] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[21] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, "Polynomial time algorithms for network code construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, June 2005.

[22] R. Dougherty and K. Zeger, "Nonreversability and equivalent constructions of multiple-unicast networks," *IEEE Transactions on*

Information Theory, vol. 52, no. 11, pp. 5067–5077, 2006.

[23] M. Bakshi, M. Effros, W. Gu, and R. Koetter, "On network coding of independent and dependent sources in line networks," In *Proceedings of the IEEE International Symposium on Information Theory*, Nice, France, June 2007. IEEE.

[24] J.D. Bruce, "*Optimum Quantization*," PhD thesis, M.I.T., Cambridge, MA, May 1964.

[25] D.K. Sharma, "Design of absolutely optimal quantizers for a wide class of distortion measures," *IEEE Transactions on Information Theory*, vol. IT-24, no. 6, pp. 693–702, Nov. 1978.

[26] X. Wu, *Algorithmic approach to mean-square quantization*. PhD thesis, University of Calgary, 1988.

[27] D. Muresan and M. Effros, "Quantization as histogram segmentation: Globally optimal scalar quantizer design in network systems," *IEEE Transactions on Information Theory*, vol. 56, no. 1, Jan. 2008. To appear.

[28] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay, "Clustering in large graphs and matrices," In *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1999.

[29] S.P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–136, March 1982. Previously an unpublished Bell Laboratories Technical Note (1957).

[30] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, Jan. 1980.

[31] P.A. Chou, T. Lookabaugh, and R.M. Gray, "Entropy-constrained vector quantization," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, no. 1, pp. 31–42, Jan. 1989.

[32] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Transactions on Information Theory*, vol. 50, no. 8, pp. 1584–1604, Aug. 2004.

[33] S. Har-Peled and S. Mazumdar, "On coresets for *k*-means and *k*-median clustering," In *36th STOC*, pp. 291–300, 2004.

[34] J. Matousek, "On approximate geometric *k*-clustering," *Discrete & Computational Geometry*, vol. 24, pp. 61–84, 2000.

[35] M. Effros and L. Schulman, "Deterministic clustering with data nets," *Electronic Colloquium on Computational Complexity, TR04-050*, June 2004. http://eccc.uni-trier.de/eccc-reports/2004/TR04-050/.

[36] M. Effros and L. Schulman, "Rapid near-optimal VQ design with a deterministic data net," In *Proceedings of the IEEE International Symposium on Information Theory*, p. 298, Chicago, June 2004. IEEE.

[37] S. Arimoto, "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel," *IEEE Transactions on Information Theory*, vol. IT-19, pp. 357–359, 1973.

[38] R.E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, July 1972.

[39] W. Gu and M. Effros, "On approximating the rate region for source coding with coded side information," In *Proceedings of the IEEE Information Theory Workshop*, pp. 432–435, Lake Tahoe, CA, Sept. 2007.

Competition and Collaboration in Wireless Networks

Plenary talk presented on Wednesday June 27, 2007 at the 2007 IEEE International Symposium on Information Theory, Nice France

H. Vincent Poor

Introduction: Over the past twenty to twenty-five years there has been considerable interest, and a number of major research developments, in the physical layer of wireless networks. Prominent examples include multiple-antenna (MIMO) systems, multiuser detection, and turbo processing, to name just three. These major advances in the physical layer have made a major difference in the way we think about wireless networks at the link level. Although of course there remain many interesting problems at the physical layer, since roughly the turn of this century interest in the area of wireless has shifted noticeably from the performance of individual links to the interactions among the nodes of the network. There are two basic modes of such interaction, *competition* and *collaboration*. Examples of competitive behavior in wireless networks include cognitive radio, in which users compete for spectrum; information theoretic security, in which an intended recipient of a message is in essence competing with an eavesdropper; and the use of game theory in the modeling, analysis and design of networks, in which terminals compete for the physical resources of a network to accomplish their own objectives. Alternatively, the use of collaboration among the nodes of a wireless network has also gained currency. Notable examples include network coding, cooperative transmission and relaying, multihop networking, collaborative beam-forming, and collaborative inference. All of these topics are areas that have arisen relatively recently in the community and they involve primarily node interactions rather than link layer issues.

Despite its title, this talk is not intended to be a comprehensive overview of this general problem area, but rather is concerned with one illustrative example of each type of node interaction. One of these is the area of *energy games* in multiple access communication networks, in which terminals compete for wireless resources so as to optimize their individual energy efficiencies. The other is the problem of collaborative inference, which arises naturally in wireless sensor networks. Of these two problems, the first is fairly well developed, while the second is at a somewhat earlier stage of development. But there are many interesting open questions in both of these areas.

Competition in the Multiple-Access Channel: Consider the uplink of a wireless infrastructure network, in which there are a number of terminals communicating up to an access point, using a multiple access protocol. We can think of such a network as being like an economic system, in which the terminals behave as economic agents competing for resources in order to maximize their own utilities. In this case, utility is based on the transfer of data up to the access point as efficiently as possible. Because this is a multiple access network, the actions of one terminal affect the utilities of the other terminals. So, we can model this situation as a competitive, or non-cooperative, game, and we can examine the particular situation in which the utility to the terminals is measured in terms of energy efficiency - that is, in bits-per-joule.

In particular, we can think of game having K players, where K is the number of terminals that are competing to communicate to the access point. Each terminal has a set of strategies, or actions, and a utility function. As noted above, the utility function of interest here is the energy efficiency, which can be measured as throughput divided by transmit power. Since throughput is measured in units of bits-per-second, and transmit power in units of watts, or joules-per-second, the units of this utility are bits-per-joule, as desired.

The transmit power is a very straightforward quantity, but the throughput needs to be modeled a bit more carefully in order to have an analytically tractable problem. Generally speaking, the throughput of a given terminal is its transmission rate multiplied by its frame success rate; i.e., this is the “goodput”. Adopting a model from [1], the frame success rate can be modeled as an instantaneous function, say f, of the received signal-to-interference-plus-noise ratio (SINR) with which that user's transmissions are received at the access point, post processing but before data detection. This is, of course, an idealization, since we know that the frame success rate is not always an exact function of the SINR. But this is a reasonable approximation, particularly for some of the receivers to be considered in this talk. The nice thing about this model is that it allows us to abstract the physical layer into the function f. That is, we can lump the entire physical layer, except for the SINR, into f. So, f includes all the things you might think as lying in physical layer - the modulation, the noise and channel models, the packet length, and so forth. Because the SINR for a given user contains the transmitted power of that user in its numerator, and the powers of all the other users in its denominator (recall that this is a multiple access channel), we can see why this is a competitive game: if one user turns up it transmit power, it will positively influence its own SINR and negatively influence the SINRs of all the other users. So, this is essentially a game on SINR.

Let us examine a specific example of this game, in which we fix the uplink receiver to be a linear multiuser detector - say, a decorrelator or minimum-mean-square-error (MMSE) detector. The set of strategies available to each player in this game is to vary its transmitter power so as to maximize its own utility; that is, this is a *power control* game. This game has been studied in [2] where it is shown that, if the function f is sigmoidal, then the game has a Nash equilibrium (i.e., an equilibrium at which no user can unilaterally improve its own utility). This equilibrium is achieved when each user chooses a transmit power to attain a certain target SINR, which is given by the solution of a simple nonlinear scalar equation. So, for this problem, Nash equilibrium requires *SINR balancing*.

This game theoretic problem is in itself quite interesting. For example, the Nash equilibrium is unique, it has useful iterative properties, there are distributed algorithms, etc. But it also gives rise to a very useful analytical tool. That is, we can consider the value of utility at the Nash equilibrium to be a measure of the energy efficiency of the network. Thus, we can examine that equilibrium utility as it is affected by various network design choices. For example, we can compare different multiuser receiver choices by comparing the Nash equilibria arising from the use of those receivers. Such a comparison was conducted in [2] for the case in which the signaling protocol is random code-division multiple-access (RCDMA), with the conclusion that considerable improvement in the energy efficiency of the terminals can be achieved by proper choice of the access point receiver; in particular, the MMSE receiver offers considerable improvement over both the matched filter and the decorrelating detector. Similarly, the addition of antennas at the access point can improve energy efficiency, at a rate that is approximately linear with the number of antennas. (Note that this latter conclusion is reminiscent of the similar well-known relationship between spectral efficiency and the number of antennas arising in MIMO systems.) An interesting point about this conclusion is that the complexity of the receiver under consideration is at the access point, whereas the energy efficiency accrues to the terminals. In a typical wireless infrastructure network, the terminals are battery-powered, and hence are the most energy-needy, whereas energy efficiency and complexity at the access point are typically of less concern.

A natural question that arises here is why we would consider a competitive game in a situation in which there is an access point that could control the power of the terminals centrally. It turns out, though, that centralized solutions to this problem yield results similar to those described above, but with much greater difficulty. For example, we can consider the Pareto, or socially optimal, solution, at which no user's utility can be improved without decreasing another's utility. Unlike the Nash equilibrium, which we have seen to be very easy to characterize and analyze, the Pareto solution is very difficult to determine, requiring numerical techniques. Moreover, while it is usually true that the Nash equilibrium is not generally Pareto optimal, it turns out to be quite close in this situation, as has been shown in [2]. So, even though we might think about using a centralized control here, the Nash equilibrium is much cleaner and it is almost the same as the more global, centralized equilibrium. Moreover, conclusions drawn from the analysis of the Nash equilibrium can be extrapolated to the behavior to be expected from centralized approaches.

Of course, in wireless networks we are interested in more than just throughput, and in particular, we are interested in other measures of quality of service (QoS). Such measures can also be addressed through the above formulation. For example, the issue of delay QoS has been considered in this context in [3] and [4]. In [3], delay is modeled in units of packet re-transmissions, resulting in a version of the Nash game described above with SINR constraints. It is seen that, as one might expect, delay-sensitive users penalize their own utilities because of their need to contain the number of packet re-transmissions required for successful reception. But, moreover, the presence of delay-sensitive users also penalizes delay-tolerant users sharing the same network, as the increased power needed by the delay-sensitive users increases interference to the other users. These effects can be analyzed precisely, and the impact of different receiver choices on them is examined in [3]. In [4], a finite-delay model is examined by introducing queueing of packets at the terminals, resulting in a QoS model involving both (source) rate and delay. Here, an SINR-constrained game similarly arises, but the set of strategies for the terminals is expanded to include the control of transmit power and transmit rate (i.e., bandwidth). A new concept – that of the "size" of a user - arises in this formulation, and it is seen that a Nash equilibrium exists only when the total sizes of all users is less than one. This gives rise to the need for access control, which can be implemented through this notion of size, which is a measure of how much of the physical layer resources a user requires in order to achieve its specified QoS parameters. Again, a closed-form expression can be obtained for the equilibrium energy efficiency, and this is used in [4] to examine the effects on energy efficiency of rate and delay constraints.

In summary of this work, we see that this game-theoretic analysis provides a very useful tool for examining the issue of energy efficiency in multiple-access wireless communication networks. Before moving to the next topic, it is of interest to mention a few additional contributions in this area. In particular, recent work has generalized this type of analysis to nonlinear multiuser receivers [5], such as maximum likelihood, maximum a posteriori probability, and interference canceling receivers; multi-carrier systems [6], in which the choice of carrier is also a user action; ultra-wideband signaling [7], in which rich scatter must be considered; and adaptive modulation, where we now let the terminals choose their modulation index for quadrature amplitude modulation [8] or their transmitter waveforms [9]. (An overview of many of these results is found in [10].) A general conclusion arising in most of these analyses is that equilibria in energy efficiency are often achieved from actions that are quite distinct from, and even the opposite of, those one would take to achieve spectral efficiency. Thus, the achievement of energy efficiency typically runs counter to the achievement of spectral efficiency, a phenomenon known from Shannon-theoretic analyses as well.

Collaborative Inference: To consider a natural framework for collaboration among network nodes, we now shift our focus to wireless sensor networks. Wireless sensor networks are distinguished from other wireless networks by several salient features. For example, whereas in other communication networks the applications could be web browsing, telephony, and so forth, in sensor networks the application is primarily inference; that is, detection, estimation or mapping. Another important distinctive feature of sensor networks is that the information at different terminals is often correlated; the sources often are not independent, since two sensors that are close together will likely have highly correlated measurements of the physical phenomenon they are sensing. And, finally, energy is often severely limited in wireless sensor networks. Of course, energy is limited in essentially all wireless networks, but in contrast to, say a cellular network or a Wi-Fi network, sensors are usually deployed in places where humans do not want to go or do not often go. So, batteries cannot be recharged and replaced easily in a sensor network. Thus, even more so than other wireless networks, energy is of major concern in sensor networks. Collaborative inference addresses these features of wireless sensor networks by having sensors work together to make inferences while conserving the resources of the network.

By way of background, we review briefly the classical supervised learning problem, in which we start with a set of real-vector inputs (or observations) X and corresponding real-valued outputs (or targets) Y. The goal of inference is to predict Y from observation of X, and to accomplish this we would like to design a function g to map values of X into predictions of Y satisfying some criterion, for example, minimizing the mean-square error, $E\{|Y - g(X)|^2\}$. Of course, if we know the joint distribution of X and Y, the solution to this problem lies in the province of classical Bayesian inference; in the particular case of minimizing the mean-square error, the proper choice of g is the conditional mean of Y given X. However, in many applications, and certainly in sensor networking where sensors are often distributed in unknown territory, the joint distribution between X and Y is unknown. So, we would like to use learning to construct a suitable function g from a set of examples (or *exemplars*), which are realizations of the inputs paired with the corresponding correct outputs. This is the classical supervised learning problem.

Now, we turn specifically to learning in wireless sensor networks. Supposing that sensors are spatially distributed, we can think of a wireless sensor network as a distributed sampling device with a wireless interface. And, as such, we can consider a learning problem in which each sensor measures some subset of the exemplars so that the exemplars are distributed around to the network. In classical learning theory it is assumed that all of the exemplars can be processed jointly, and a wealth of methodology is available for such problems, including notably smoothness-penalized least-squares function fitting within a reproducing kernel Hilbert space (RKHS). Here, we address the question of what can be done when the exemplars are not collected together, but rather are distributed throughout the network in the above fashion.

We assume that energy and bandwidth constraints preclude the sensors from transferring their exemplars to a central collection point for processing, and thus they must cope with the data locally. A justification for this is the usual one for distributed networks; i.e., that local communication is efficient relative to global communication, and certainly in wireless networks that is true. Communication among neighbors requires less energy and bandwidth than does communication with a distant access point. This gives us the seed of a model, in which we have the data distributed according to a physical sensor network that gives rise to a neighborhood structure. These considerations lead to a fairly general model that applies to this situation and many others, and that shapes the nature of collaboration in such networks.

In particular, we consider a model in which there are a number, M, of *learning agents* which in this setting are sensors, and a number, N, of training examples. The topology of the network can be captured in a bipartite graph $G = (V_1 + V_2, E)$, in which the vertices in one set V_1 of vertices represent the M learning agents, the vertices in the other set V_2 of vertices represent the N training examples, and an edge extends from a vertex in V_1 to a vertex in V_2 if the corresponding learning agent in V_1 has access to the corresponding exemplar in V_2. This model provides a general framework for examining distributed inference; it is quite general, encompassing fully centralized processing, decentralized processing with a public data-base, fully decentralized processing, etc. (See [11] for further details.) But generally, to analyze this model we do not need to be concerned with its particular topology. We can simply think about a general bipartite graph consisting of learning agents connected to a training database, and examine inference in that context.

A natural approach to consider in such a setting is *local* learning, in which each learning agent learns according to kernel regression applied to those exemplars to which it is directly connected. In this way, each learning agent creates a local model for what is going on in the entire field being sensed. A problem with local learning is that it is provably *locally incoherent*; that is, if two learning agents examine the same datum, their estimates may not agree on that datum. This is an undesirable property, and it can be shown that it generally leads to a penalty in terms of overall fitting error. So, a natural question that arises is whether there is some way for the agents to collaborate to eliminate this local incoherence, while retaining the efficiency of locality, i.e., without having to communicate more broadly than their neighborhoods. One way to do this is to begin with local learning, in which each of the learning agents performs a kernel regression locally. But rather than stopping there, after performing local regression, each agent then updates all of the data points to which it has access with its own estimates of the targets. This process can be further iterated over time in addition to iterating over sensors by adding an inertia-penalty to the regression.

Such a message-passing algorithm is proposed and analyzed in [11] for the MMSE regression problem, where it has been shown to have very favorable properties. In particular, the algorithm converges in norm to a relaxation of the centralized kernel estimator as we iterate in time, and the limit is locally coherent. Moreover, the iterates lie in the span of a finite number of functions so that the estimation problem being solved at each iteration is actually parameter estimation rather than function estimation, which is very efficient computationally. Also, the global fitting error improves with every update. These properties hold for any bipartite graph of the form described above, regardless of whether it is connected or not. But, under a connectivity assumption and the assumption of independent and identically distributed exemplars, the local estimates can all be made to converge in RKHS norm to the conditional mean by proper programming of the penalty terms. Further description of this algorithm, together with some illustrative numerical examples, is found in [11].

The message-passing approach to collaborative inference of [11] provides a general formalism for designing and analyzing algorithms for collaborative inference in a distributed sensing environment. Another approach to distributed learning is considered in [12], in which, rather than collaborating, sensors communicate directly back to an access point via links with very limited capacity. Questions then arise regarding whether consistent learning of regression or classification functions can take place in this setting. These questions are answered in [12] largely in the affirmative, in that consistent regression is possible with the transmission of only $\log_2 3$ bits per sensor per decision, while consistent classification is possible with transmission of only a single bit per sensor per decision. Another way in which wireless sensors can collaborate locally to save energy is via *collaborative beam-forming* [13], in which sensors that cluster together physically can collaborate to form a common message and then to form a beam on the access point. A further problem of interest is *judgment aggregation*, considered in [14], which is related to the problem of local coherence mentioned earlier. This is a problem that arises in many applications beyond sensor networking, including the combination of judgments of individuals in panels of experts. A review of these and related issues in this general area is found in [15].

Conclusions: As can be seen from the above discussion, the two problems discussed here are quite rich in themselves. However, they are only representative of a large class of emerging research problems in which node interaction is the primary consideration. Of course these issues are not really removed from the physical layer, as all such problems are driven largely by the need to make optimal use of the physical-layer resources.

References

[1] D. J. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Commun.*, Vol. 7, pp. 48 – 54, Apr. 2000.

[2] F. Meshkati, H. V. Poor, S. C. Schwartz and N. Mandayam, "An energy-efficient approach to power control and receiver design in wireless data networks," *IEEE Trans. Commun.*, Vol. 53, No. 11, pp. 1885 -1894, Nov. 2005.

[3] F. Meshkati, H. V. Poor and S. C. Schwartz, "Energy efficiency-delay tradeoffs in multiple-access networks," *IEEE Trans. Inform. Theory*, under review.

[4] F. Meshkati, H. V. Poor, S. C. Schwartz and R. Balan, "Energy-efficient resource allocation in wireless networks with quality-of-service constraints," *IEEE Trans. Commun.*, to appear.

[5] F. Meshkati, D. Guo, H. V. Poor and S. C. Schwartz, "A unified approach to energy-efficient power control in large CDMA systems," *IEEE Trans. Wireless Commun.*, to appear.

[6] F. Meshkati, M. Chiang, H. V. Poor and S. C. Schwartz, "A game-theoretic approach to energy-efficient power control in multi-carrier CDMA systems," *IEEE J. Selected Areas Commun.*, Vol. 24, No. 6, pp. 1115 - 1129, Jun. 2006.

[7] G. Bacci, M. Luise, H. V. Poor and A. Tulino, "Energy-efficient power control in impulse radio UWB wireless networks," *IEEE J. Selected Topics Signal Process.*, Vol. 1, No. 3, pp. 508 – 520, Oct. 2007.

[8] F. Meshkati, A. Goldsmith, H. V. Poor and S. C. Schwartz, "A game theoretic approach to energy-efficient modulation in CDMA networks with delay QoS constraints," *IEEE J. Selected Areas Commun.*, Vol. 25, No. 6, pp. 1069 - 1078, Aug. 2007.

[9] S. Buzzi and H. V. Poor, "Non-cooperative games for spreading code optimization, power control and receiver design in wireless data networks," *Proc. 13th European Wireless Conf.*, Paris, France, Apr. 1 - 4, 2007.

[10] F. Meshkati, H. V. Poor and S. C. Schwartz, "Energy-efficient resource allocation in wireless networks," *IEEE Signal Proc. Mag.*, Vol. 24, No. 3, pp. 58 - 68, May 2007.

[11] J. B. Predd, S. Kulkarni and H. V. Poor, "Distributed kernel regression: An algorithm for training collaboratively," *Proc. 2006 IEEE Inform. Theory Workshop*, Punta del Este, Uruguay, Mar. 13 - 17, 2006.

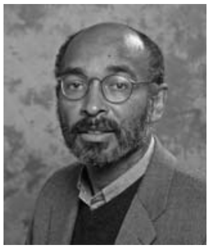
[12] J. B. Predd, S. Kulkarni and H. V. Poor, "Consistency in models for distributed learning under communication constraints," *IEEE Trans. Inform. Theory*, Vol. 52, No. 1, pp. 52 - 63, Jan. 2006.

[13] H. Ochiai, P. Mitran, H. V. Poor and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Process.*, Vol. 53, No. 11, pp. 4110 - 4124, Nov. 2005.

[14] J. B. Predd, S. Kulkarni, D. Osherson and H. V. Poor, "Aggregating forecasts of probability from incoherent and abstaining experts," *Proc. 2006 INFORMS Annual Meeting*, Pittsburgh, PA, Nov. 5 – 8, 2006.

[15] J. B. Predd, S. Kulkarni and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Process. Mag.*, Vol. 23, No. 4, pp. 56 – 69, Jul. 2006.

Signal Processing Algorithms to Decipher Brain Function



Plenary talk presented on Friday June 29, 2007 at the 2007 IEEE International Symposium on Information Theory, Nice France

Emery N. Brown

About the author: Emery N. Brown, M.D., Ph.D. is a Professor of Computational Neuroscience, and Professor of Health Sciences and Technology at the Massachusetts Institute of Technology, Massachusetts General Hospital, and Professor of Anaesthesia at Harvard Medical School.

The nervous system is composed of specialized cells called neurons that transmit information through electrical impulses known as action potentials. Because the shape of the action potential is stereotypical, information is carried in the occurrences times of the action potentials or spikes. The sequence of spikes or spike train forms point process time-series (0-1 measurements made in continuous time) (Daley and Vere-Jones, 2003). The nervous systems of animals at all levels of the evolutionary scale use the firing patterns of spike trains to encode representations of relevant biological signals and external stimuli. For example, when a musical score is played neurons in the ear discharge spike trains that in turn, induce firing along the entire pathway from the ear to the primary auditory cortex located in the temporal lobe of the brain. The patterns of these spike trains in the auditory pathway encode a representation of the music. The code is both dynamic and stochastic; it depends probabilistically on the acoustic properties of the music, previous experience of the system with the same score, as well as the most recent and current state of the auditory system and other parts of the brain. Even though the stimulus (musical score) is often continuous, its representation in the nervous systems is as a high-dimensional point process time-series. In addition, the receptive fields or response properties of neurons are plastic. That is, as a neuron in the auditory pathway is stimulated repeatedly by a sound it changes how it responds to the sound stimulus.

To study these two fundamental properties of neural systems, we have developed algorithms and an analysis paradigm for state estimation from point processes observations by modeling the biological signal as unobservable or hidden state processes and the neural spike trains as point processes. We have used these signal processing algorithms to characterize the properties of neural systems.

First, we established a nonlinear, recursive filter algorithm to estimate an unobserved state process from an ensemble of point process observations (Brown et al., 1998). We formulated the state estimation problem using the Bayes' theorem and Chapman-Kolmogorov (BCK) system of equations in which observations are the point processes and the state is a linear Gaussian system. This algorithm, which we termed the Bayes' filter, computes recursive Gaussian approximations to this BCK system. For a linear Gaussian system observed through a point process, the Bayes' filter is an analog of the Kalman filter for a linear Gaussian system observed through Gaussian observations. In addition to relating our algorithms to other standard recursive filtering procedures, this analysis made it possible to derive as special cases several currently used neural spike train decoding algorithms, i.e. algorithms to estimate a biological signal from neural spiking activity, such as the maximum likelihood, maximum correlation and population vector methods (Brown et al., 1998). We used this algorithm to show how the position of a rat could be estimated from the ensemble spiking activity of 30 neurons, in the hippocampus, a brain region responsible for short-term memory formation, with an accuracy of 8 cm during 10 minutes of foraging in an open environment (Brown et al. 1998).

Second, we developed a steepest point process adaptive filter algorithm for tracking a signal observed through point process observations (Brown et al. 2001). The algorithm is derived by using the instantaneous log likelihood function of a point process in lieu of the local quadratic loss function used to develop adaptive filter algorithms for continuous-valued observations. This algorithm is for point processes an analog of a steepest descent adaptive filter for continuous-valued observations. We used our adaptive filter algorithms to establish that the temporal evolution of receptive fields of neurons in the hippocampus and entorhinal cortex can be reliably tracked on a millisecond timescale (Brown et al., 2001; Frank et al., 2002). With these algorithms we demonstrated that when a rat learns a novel environment, the hippocampus can form new spatial receptive fields within 5 to 6 minutes (Frank et al. 2004). However, stable hippocampal receptive fields are not sufficient for the animal to treat the new location as familiar.

Third, we developed an approximate Expectation-Maximization (EM) algorithm to solve the problem of simultaneously estimating an unobservable state process along with the unknown parameters in state and point process observation models (Smith and Brown, 2003). The algorithm was derived by using an extension of our Bayes' filter algorithm with the fixed interval smoothing and covariance smoothing algorithms to evaluate efficiently the algorithm's E-step. We established a similar algorithm for state estimation from binary observations (0-1 processes measured in discrete time). In an analysis that used both our adaptive filter and EM state estimation algorithms, we showed that monkey hippocampal neurons signaled learning of novel tasks by changing their stimulus-selective response properties (Wirth et al., 2003). This change in the pattern of selective neural activity occurred before, at the same time as, or after learning, strongly suggesting that these neurons are involved in the initial formation of new associative memories.

Fourth, we generalized the derivation of our Bayes' filter so that the algorithm could be applied to any point process observation model defined in terms of a conditional intensity function (Barbieri et. al 2004a). Because the conditional intensity function is a history dependent rate function, this formulation opens the paradigm to use with a broad class of point process models for the study neural spike trains. Using the algorithm to analyze the firing patterns of approximately 30 simultaneously recorded hippocampal neurons, we demonstrated that the position of a rat could now be predicted with a median accuracy of 6 cm or less during 10 minutes of foraging. These results suggested that a highly accurate, dynamic representation of the animal's position is maintained in the ensemble spiking activity of hippocampal neurons. We also showed that the commonly used linear filter algorithms for neural spike train decoding are far less accurate than our state-space estimation algorithm. Furthermore, we established an important link between our analysis framework and the widely used information theory paradigm in computational neuroscience by demonstrating how to compute from the recursive estimates of the BCK posterior probability densities, dynamic estimates of the Shannon mutual information between the animal's position (the biological signal) and the ensemble spiking activity (the neural response).

Finally, we demonstrated that analogs for point process observations of the steepest descent, recursive least-squares and Kalman filter algorithms for continuous-valued observations can be derived from a unified estimation framework (Eden et al., 2004). By analogy with continuous-valued observation models, we also showed that by using an augmented state-space model both the unobserved state and the dynamic unknown parameters of the state and observation processes can be simultaneously estimated. We demonstrated the significance of this last result for neural computations by showing how the ensemble representation of a biological signal can be estimated (decoded) from joint spiking activity even as the receptive fields of the individual neurons in the ensemble evolve. This problem occurs naturally in the control of a neural prosthetic either because neurons change their response properties and or because which neurons are being recorded by the chronically implanted electrode array changes with time (Eden et al., 2004; Eden et al., 2005; Truccolo et al., 2005; Barbieri et al. 2005). We are using these results to design optimal algorithms to control neural prosthetic devices (Srinivasan et al. 2005; Srinivasan et al., 2006; Srinivasan et al., 2007). We have also developed sequential Monte Carlo versions of our algorithms (Ergun et al., 2007).

Our research on signal processing algorithms for deciphering brain function has led to new approaches to analyzing performance in learning experiments (Smith et al., 2004; Smith et al., 2005; Smith et al., 2007) and human heart beat time-series (Barbieri and Brown, 2005; Barbieri and Brown, 2006).

References

Barbieri R, Frank LM, Nguyen DP, Quirk MC, Solo V, Wilson MA, Brown EN, Dynamic analyses of information encoding by neural ensembles. *Neural Computation*, 2004, 16(2): 277-307.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells, *Journal of Neuroscience* 1998; 18: 7411-7425.

Eden UT, Frank LM, Barbieri R, Solo V, Brown EN, Dynamic analyses of neural encoding by point process adaptive filtering, *Neural Computation*, 2004, 16(5): 971-998.

Eden UT, Ph.D. Dissertation, Division of Engineering and Applied Sciences, Harvard Univeristy, 2005.

Ergun A, Barbieri R, Eden UT, Wilson MA, Brown EN. Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. *IEEE Transactions on Biomedical Engineering*, 2007; 54(3):419-428.

Smith AC, Brown EN. Estimating a state-space model from point process observations. *Neural Computation*. 2003; 15: 965-91.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki W, Brown EN. Dynamic analysis of learning in behavioral experiments, *Journal of Neuroscience*, 2004, 15: 965-91.

Smith AC, Wirth A, Suzuki W, Brown EN. Bayesian analysis of interleaved learning and response bias in behavioral experiments. *Journal of Neurophysiology*, 2007, Mar; 97(3):2516-24.

Srinivasan L, Eden UT, Willsky AS, Brown EN. A state-space analysis for reconstruction of goal-directed movements using neural signals. *Neural Computation*, 2006, 18(10): 2465-2494.

Srinivasan L, Brown EN. A state-space framework for movement control to dynamic goals through brain-driven interfaces. *IEEE Transactions on Biomedical Engineering*, 2007, 54(3): 526-535.

Srinivasan L, Eden UT, Mitter SK, Brown EN. General purpose filter design for neural prosthetic devices. *Journal of Neurophysiology*, 2007 (Published online May 23, 2007).

Truccolo W, Eden U, Fellow M, Donoghue JD, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble and covariate effects. *Journal of Neurophysiology*, (published online Sept. 8, 2004), 2005, 93: 1074-1089.

Wirth S, Yanike M. Frank LM, Smith AC, Brown EN, Suzuki WA. Single neurons in the monkey hippocampus and learning of new associations. Science 2003, 300: 1578-81.

Brown EN and Barbieri R. Dynamic Analyses of Neural Representations Using the State-Space Modeling Paradigm. In: Madras B, Von Zastrow M, Colvis C, Rutter J, Shurtleff D, Pollock J. *The Cell Biology of Addiction*, New York, Cold Spring Harbor Laboratory Press, 2005, p. 415-432.

Barbieri R, Brown EN. Analysis of heart beat dynamics by point process adaptive filtering. *IEEE Transactions on Biomedical Engineering*, 2006, 53(1): 4-12.

Brown EN, Nguyen DP, Frank LM, Wilson MA, Solo V. An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of the National Academy of Sciences* 2001; 98:12261-66.

Daley D and Vere-Jones D. An Introduction to the Theory of Point Process, 2nd edition, New York: Springer-Verlag, 2003.

Frank LM, Eden UT, Solo V, Wilson MA, Brown EN. Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. Journal of Neuroscience, 2002, 22:3817-30.

Frank LM, Stanley GB, Brown EN. Hippocampal plasticity across multiple days of exposure to novel environments. *Journal of Neuroscience*, 2004, 24(35):7681-89.

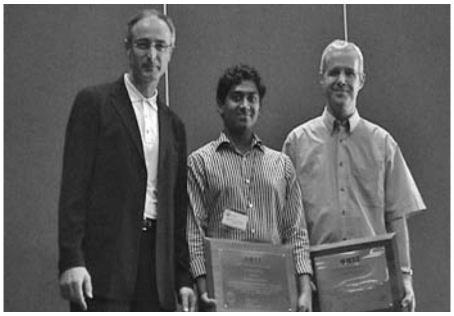
Smith AC, Wirth A, Suzuki WA, Brown EN. Bayesian analysis of interleaved learning and response bias in behavioral experiments. *Journal of Neurophysiology*, 2007, 97(3): 2516-2524.

Reflections on Universal Compression of Memoryless Sources

Alon Orlitsky,

Narayana P. Santhanam,

Junan Zhang



From left to right: Bixio Rimoldi, Narayana P. Santhanam, and Alon Orlitsky.

In preparation for your next safari, you observe a random sample of African animals. You find 3 giraffes, 1 zebra, and 2 elephants. How would you estimate the probability of the various species you may encounter on your trip?

The empirical-frequency, (maximum likelihood) estimator assigns probability 1/2 to giraffes, 1/6 to zebras, and 1/3 to elephants. But the poor estimator will be completely unprepared for an encounter with an offended lion.

Furthermore, how is it to distinguish between the probabilities of a lion and a tiger, both of which are unseen?

The situation we encounter here is one of *large alphabets*. The probability distribution to be estimated, is over an alphabet (the set of all animals in this case) much larger than the data sample of size 6 observed.

On the other hand, many results in both statistics and information theory hold when the data we have to work with is much larger than the alphabet size. For the problem mentioned above, and several others involving topics from text and images to biology today, this is not the case. As in the above problem, conventional approaches do not work and we have to rework such topics.

We consider the allied problems of distribution estimation and universal compression, both in large alphabet settings. The results in [1] were phrased in the language of universal compression literature and the connection to estimation was alluded to therein.

Following the reasoning in Section 2, we shall review the two problems together, and demonstrate in addition that these results also yield good solutions to text classification applications.

We then consider an approach taken by Laplace for probability estimation in Section 3. While Laplace's estimator doesn't really estimate probabilities well in the large alphabet setting, we can glean important insights from it. Besides no estimator can esatimate large alphabet sequence probabilities well, as we will see from a result by Kieffer in Section 4.

We take a different approach, involving *patterns* to resolve this impasse. In particular, we show that a large alphabet estimator proposed by Good and Turing bears striking resemblance to the intuition obtained in the Laplace case. However, it is still hard to make their approach work for some applications.

We continue with the formalism motivated by universal compression, but combine it with the patterns approach to avoid overfitting in the large alphabet setting. We build on our new estimators in Section 7 to obtain good results for text classification in Section 8.

1 Universal Compression

The universal compression framework tackles compression when the source distribution is unknown. To do so, it is typically assumed that the source distribution belongs to some natural class $\mathcal{P}$, such as the collection of *i.i.d.*, Markov, or stationary, distributions, but that the precise distribution within $\mathcal{P}$ is not known [2, 3]. The objective then is to compress the data almost as well as when the distribution is known in advance, namely to find a *universal* compression scheme that performs almost optimally by approaching the entropy no matter which distribution in $\mathcal{P}$ generates the data. Extensive overviews can be found in [4–7].

Let a source X be distributed over a support set $\mathcal{X}$ according to a probability distribution p. An *encoding* of X is a prefix free 1-1 mapping $\phi : \mathcal{X} \to \{0, 1\}^*$. Every probability assignment q over $\mathcal{X}$ corresponds to an encoding of X, where the number of bits allocated to $x \in \mathcal{X}$ is approximately $\log(1/q(x))$ and as such we will use the terms interchangeably where there is no ambiguity.

In case, the underlying distribution p is unknown, the extra number of bits required to encode x when a distribution q is used instead of p is

$$\log \frac{1}{q(x)} - \log \frac{1}{p(x)} = \log \frac{p(x)}{q(x)}.$$

The *worst-case redundancy* of q with respect to the distribution $p \in \mathcal{P}$ is

$$\hat{R}(p, q) \overset{\text{def}}{=} \max_{x\in\mathcal{X}} \log \frac{p(x)}{q(x)} ,$$

the largest number of extra bits allocated for any possible x. The *worst-case redundancy* of q with respect to the collection $\mathcal{P}$ is

$$\hat{R}(\mathcal{P}, q) \overset{\text{def}}{=} \max_{p\in\mathcal{P}} \hat{R}(p, q) ,$$

the number of extra bits used for the worst distribution in $\mathcal{P}$ and worst $x \in \mathcal{X}$. The *worst-case redundancy* of $\mathcal{P}$ is

$$\hat{R}(\mathcal{P}) \overset{\text{def}}{=} \min_{q} \hat{R}(\mathcal{P}, q) = \min_{q} \max_{p\in\mathcal{P}} \max_{x\in\mathcal{X}} \log \frac{p(x)}{q(x)} , \qquad (1)$$

the lowest number of extra bits required in the worst case by any possible encoder q.

In problems considered in this article, the support $\mathcal{X}$ is a collection of strings of a particular length, say n, which we refer to as the *blocklength*. The encoder q is hence a probability distribution on strings of length n. The redundancy will therefore depend on n, and we will be interested in how the redundancy increases with n.

If the redundancy grows $o(n)$, the excess number of bits we use per symbol is asymptotically zero. In such cases, some universal encoder can compress almost as well as every possible source in the distribution class.

A variation of the above problem is *sequential* universal compression. In this setting, the universal encoders for different blocklength cannot be arbitrary. For all $n \geq 1$, the universal encoders q_n and q_{n+1} on length-n and length-$(n+1)$ strings respectively must satisfy for all strings $x_1 \ldots x_n$,

$$q_n(x_1, \ldots , x_n) = \sum_{x_{n+1}} q_{n+1}(x_1, \ldots , x_n x_{n+1}) ,$$

namely the marginals must be "consistent". The question we ask again is how the redundancy of sequential encodings grows with the blocklength.

Example 1. Let I_k^n be the collection of *i.i.d.* distributions on length n sequences of symbols from an alphabet of size k.

Consider the universal encoder q_L that describes any sequence by (i) describing its type by an encoding corresponding to a uniform distribution over the $\binom{n+k-1}{k-1}$ possible types, and then (ii) identifying the sequence among all sequences of the type using the encoding corresponding to a uniform distribution over sequences of the type.

Conditioned on a type, the distribution over sequences of that type is uniform no matter what the underlying *i.i.d.* distribution is (as long as the type probability is non-zero). Hence, the extra codelength q_L uses is at most the number of bits used in step (i), which is at most $\log \binom{n+k-1}{k-1} \leq (k-1)\log(\frac{n}{k}+1)$ bits. □

2 Distribution Estimation

The distribution estimation problem we consider is identical to the sequential universal compression problem for *i.i.d.* sequences. The set of possible distributions that could be in effect is $\mathcal{P}$, and we come up with a sequential *probability estimator* over the same support. This approach is applied in a variety of fields other than universal compression: finance [8], online algorithms, and learning, *e.g.*, [9–11].

To evaluate the performance of an estimator, we apply it not just once but repeatedly to an *i.i.d.* sequence of elements drawn from the unknown distribution. Before each element is revealed, we use the estimator to evaluate its conditional probability given the previous elements. Multiplying the conditional probability estimates together, we obtain the probability that the estimator assigns to the whole sequence.

As before, let $\mathcal{P}$ be the collection of all possible distributions that could be in force. We derive sequential estimators that assign to every sequence a probability that is not much lower than the highest probability assigned to it by any distribution. The *sequence attenuation* of an estimator q for a sequence x_1^n to be

$$\hat{A}\left(q, x_1^n\right) \overset{\text{def}}{=} \frac{\max_{p\in\mathcal{P}} p(x_1^n)}{q(x_1^n)} \,,$$

the ratio between the highest probability assigned to x_1^n by any distribution and the probability assigned to it by q. Like the worst case redundancy, the *worst-case sequence attenuation* of q for patterns of length n is

$$\hat{A}_n(q) \overset{\text{def}}{=} \max_{x_1^n \in \mathcal{X}^n} \hat{A}\left(q, x_1^n\right) \,,$$

the largest sequence attenuation of q for any length-n sequence. Normalizing, $(\hat{A}_n(q))^{1/n}$ is the *worst-case per-symbol attenuation* of q for sequences of length n. Attenuation is simply “redundancy without the logs”, it is defined for convenience.

A natural objection to using a compression framework is that overestimation is also undesirable while estimating distributions, not just underestimation.

It is easy to see however that overestimation is not a serious issue since for all distributions p and q over $\mathcal{X}$.

$$p\{x : q(x) \ge Ap(x)\} \le \frac{1}{A} \,.$$

Let a universal encoder compress a collection of distributions $\mathcal{P}^n$ over sequences of length n with redundancy R_n, which implies, that the encoder has attenuation 2^{R_n}. It follows that for all sources $p \in \mathcal{P}^n$, with probability $\ge 1 - \frac{1}{2^{R_n}}$,

$$\frac{1}{n}\left|\log\frac{p(X^n)}{q(X^n)}\right| \le \frac{R_n}{n} \,. \tag{2}$$

Therefore, if the per symbol redundancy, $R_n/n \to 0$, then

$$\frac{1}{n}\left|\log\frac{p(X^n)}{q(X^n)}\right| \to 0 \,.$$

in probability. The above statement follows directly if $R_n \to \infty$, and it is easy to see that the above statement is true even if R_n is bounded.

Therefore, if a collection of sources incurs diminishing per-symbol redundancy, any source in the collection can be estimated well using a scheme with diminishing per-symbol redundancy.

3 Older Approaches: Laplace

Laplace first addressed the unseen-elements problem partially. Laplace [12] proposed adding one to the count of each species, including to the collection of unseen ones, thereby assigning probability $(3+1)/10 = 0.4$ to giraffes, $(1+1)/10 = 0.2$ to zebras, $(2+1)/10 = 0.3$ to elephants, and $(0+1)/10 = 0.1$ to lions.

From the viewpoint of an information theorist, Laplace does something very interesting in the case of finite alphabets. Recall the estimator q_L from Example 1. Laplace's add-one rule is a sequential realization of q_L. Namely for all sequences x^n, $q_L(x^n)$ can be obtained by using the add-one estimator's probability for x_1 given the empty string λ, multiplied by the estimators probability of x_2 given x_1, and so on.

As mentioned before, the principle here is that the number of types is polynomial in the sequence length (hence do not matter) *under the implicit assumption* that the alphabet size is much smaller than the sequence length. When this assumption is no longer true, the estimator of Laplace holds no charm for us—it attenuation grows exponentially with the alphabet size.

4 Problems with Large Alphabets

The problem with the Laplace estimator is fundamental. Kieffer [5] showed that for the collection $\mathcal{I}_*$ of all *i.i.d.* distributions over countably infinite alphabets, there exists no universal encoder. Namely, there cannot exist an estimator q such that for all $p \in \mathcal{I}_*^n$, the collection of all *i.i.d.* distributions over length-n sequences from a countably infinite alphabet, as $n \to \infty$, $\frac{1}{n}D(p(X^n)\|q(X^n)) \to 0$. It is not surprising this should be the case—after all, in our safari example of giraffes, elephants and hippos, how can one possibly distinguish between lions and tigers, neither of which we have seen?

We explore the result in detail in [13] where we analyze the redundancy in all regimes — in particular, when the alphabet size is $\Omega(n)$, universal schemes incur $\Omega(n)$ redundancy.

5 Patterns

In order to resolve this impasse, we ask—what if we only ask the probability that we will repeat an observation in the past, and the probability we will encounter something previously unseen?

Formalizing this approach, we define the *pattern* of a string.

Consider, for example, the string "abracadabra" over the Roman alphabet. We convey the relative precedence of the symbols in the string through it's *pattern*

$$12314151231\,,$$

and to recover the orginal sequence, we need the *dictionary*

$$\{a \to 1, b \to 2, r \to 3, c \to 4, d \to 5\}\,.$$

Note that while describing a pattern sequentially, we have to describe whether the forthcoming symbol is any of the symbols seen thus far (in which case, we reuse the index given to the symbol), or if it is new (we use the lowest unused number as the index). We will consider universally compressing patterns of *i.i.d.* strings, and equivalently, estimating distributions induced over patterns of *i.i.d.* strings.

A string of positive integers is the pattern of some sequence if and only if the first appearance of any $i \geq 1$ precedes that of $i+1$. For example, the empty string λ and the strings 1, 12, and 121 are patterns (of the empty string, and, say, "a", "ad", and "ada", respectively), while 2, 21, and 132 are not.

Let Ψ^n denote the set of length-n patterns. For example, $\Psi^0 = \{\Lambda\}$, $\Psi^1 = \{1\}$, $\Psi^2 = \{11, 12\}$, and $\Psi^3 = \{111, 112, 121, 122, 123\}$. It can be shown that every length-n pattern corresponds to a partition of a set of cardinality n, hence $|\Psi^n|$ is the n'th Bell number.

Let $\mathcal{A}^n$ denote the set of n-element sequences over an alphabet $\mathcal{A}$. For example, $\{a, b\}^2 = \{aa, ab, ba, bb\}$. If p be a probability distribution over an alphabet $\mathcal{A}$, then for every $n \in \mathbb{Z}^+$, p induces a probability distribution p over Ψ^n where

$$p(\overline{\psi}) \overset{\text{def}}{=} p\{\overline{x} \in \mathcal{A}^n : \Psi(\overline{x}) = \overline{\psi}\}$$

denotes the probability that a sequence of elements, each selected according to p, will form the pattern $\overline{\psi} \in \Psi^n$. For example, for any probability p over an alphabet $\mathcal{A}$, $p(1) = p(\mathcal{A}) = 1$, indicating that the first element of any pattern is 1 ("new"). If p is a distribution over $\{a, b\}$ where $p(a) = p$ and $p(b) = 1 - p \overset{\text{def}}{=} \overline{p}$, then $p(11) = p\{aa, bb\} = p^2 + \overline{p}^2$, the probability that two elements will be identical, and $p(12) = p\{ab, ba\} = 2p\overline{p}$, the probability that the two elements will be distinct.

Continuous (*i.e.*, non-atomic) distributions induce probabilities over patterns as well. For example, if p is any continuous distribution, then for all n, $p(12\ldots n) = p\{x_1 \ldots x_n : x_i \neq x_j\} = 1$, indicating that, with probability 1, a finite number of elements selected according to a continuous distribution are all distinct. It follows that for continuous distributions, every $\overline{\psi} \in \Psi^n - \{12\ldots n\}$, namely every pattern with repetitions, has $p(\overline{\psi}) = 0$.

Our goal is to derive an estimator that, though unaware of the underlying probability p, assigns to every pattern $\overline{\psi}$ a probability that is not much smaller than the induced probability $p(\overline{\psi})$. Since we do not know the underlying distribution, we consider the one that assigns to $\overline{\psi}$ the highest probability. The maximum probability of a pattern $\overline{\psi}$ is

$$\hat{p}(\overline{\psi}) \overset{\text{def}}{=} \max_p p(\overline{\psi})\,,$$

the highest probability assigned to the pattern by any distribution. For example, since any distribution p has $p(1) = 1$, we have $\hat{p}(1) = 1$. Since any distribution p concentrated on a single element has $p(1\ldots1) = 1$ for any number of 1's, we obtain $\hat{p}(1\ldots1) = 1$, and, since any continuous distribution p has $p(12\ldots n) = 1$, we derive $\hat{p}(12\ldots n) = 1$. In general however it is difficult to determine the maximum probability of a pattern. For example, some work [13] is needed to show that $\hat{p}(112) = \frac{1}{4}$.

To study patterns of *i.i.d.* sequences, we introduce the notion of a *profiles*, which play the role of types for sequence. Profiles form sufficient statistics for patterns.

The *multiplicity* of $\psi \in \mathbb{Z}^+$ in a pattern $\overline{\psi}$ is

$$\mu_\psi \overset{\text{def}}{=} \mu_\psi(\overline{\psi}) \overset{\text{def}}{=} |\{1 \leq i \leq |\overline{\psi}| : \psi_i = \psi\}|\,,$$

the number of times ψ appears in $\overline{\psi}$. The *prevalence* of a multiplicity $\mu \in \mathbb{N}$ in $\overline{\psi}$ is

$$\varphi_\mu \overset{\text{def}}{=} \varphi_\mu(\overline{\psi}) \overset{\text{def}}{=} |\{\psi : \mu_\psi = \mu\}|\,,$$

the number of symbols appearing μ times in $\overline{\psi}$. The *profile* of $\overline{\psi}$ is

$$\overline{\varphi} \overset{\text{def}}{=} \varphi(\overline{\psi}) \overset{\text{def}}{=} \left(\varphi_1, \ldots, \varphi_{|\overline{\psi}|}\right)$$

the vector of prevalences of μ in $\overline{\psi}$ for $1 \leq \mu \leq |\overline{\psi}|$.

Using the combinatorial properties of patterns and profiles, we show that the redundancy of compressing patterns of length n *i.i.d.* sequences is $\mathcal{O}(\sqrt{n})$, therefore the per-symbol redundancy diminishes to zero. Equivalently, it is possible to build an estimator with per-symbol attenuation 1, if all we care about is an estimate of probability that the next symbol is the same as one that has appeared before and the probability of seeing a previously unseen symbol.

6 Good Turing estimators

Good and Turing came up with a surprising estimator that, on the surface, bears little resemblance to either the empirical-frequency or the Laplace estimators above. After the war, Good published the estimator [14] mentioning that Turing had an "intuitive demonstration" for it, but not describing what this intuition was.

6.1 History

Good and Turing encountered this problem while trying to break the Enigma cipher during World War II [15]. The British intelligence was in possession of the Kenngruppenbuch, the German cipher book that contained all possible secret keys, and used previously decrypted messages to document the page numbers of keys used by various U-boat commanders. They wanted to use this information to estimate the distributions of pages that each U-boat commander picked secret keys from.

6.2 Similarity with Laplace

There is a striking similarity between Good Turing estimators and the Laplace estimator. The Good Turing estimator attempts to do with patterns exactly what the Laplace estimator was doing with sequences. The following description is an analog of Example 1 for patterns.

Consider the universal encoder q_{GT} that describes any pattern in Ψ^n by (i) describing its profile by an encoding corresponding to a uniform distribution over the $2^{\mathcal{O}(\sqrt{n})}$ possible profiles, and then (ii) identifying the pattern among all patterns of the profile using the encoding corresponding to a uniform distribution over patterns of the profile.

Conditioned on a profile, the distribution over patterns of that profile is uniform no matter what the underlying *i.i.d.* distribution is (as long as the profile probability is non-zero). Hence, the extra codelength q_{GT} uses is at most the number of bits used in step (i), which is at most $\mathcal{O}(\sqrt{n})$ bits.

Unlike in the sequence case, it is not possible to achieve the target distribution q_{GT} over all patterns in a sequential manner. The Good Turing estimate is a sequential approximation where the conditional probability of ψ_1^{n+1} given ψ_1^{n} is obtained by dividing the target probability for $\psi_1, \ldots, \psi_{n+1}$ by the target probability of $\psi_1, \ldots, \psi_n$ and in addition, normalizing. Since it is an approximation, in practice, the estimator is always used with some kind of *smoothing method* that processes the profile of the data first.

The Good-Turing estimator has since been incorporated into a variety of applications such as information retrieval [16], spelling correction [17], word-sense disambiguation [18], and speech recognition, *e.g.*, [19] where it is applied to estimate the probability distribution of rare words.

While the Good-Turing estimator performs well in general, it is suboptimal for elements that appear frequently, hence was modified in subsequent estimators, *e.g.*, the Jelinek-Mercer, Katz, Witten-Bell, and Kneser-Ney estimators [19]. However, in text classification applications for example, Laplace outperforms several smoothing methods of Good Turing.

On the theoretical side, interpretations of the Good-Turing estimator have been proposed [20–22], and its convergence rate was analyzed [23]. Yet, lacking a measure for assessing the performance of an estimator, no objective evaluation or optimality results for the Good-Turing estimator have been established.

7 New Estimators

We develop universal *sequential* encodings of patterns with diminishing per-symbol redundancy. Our first universal encoding scheme has redundancy at most

$$\frac{2}{\sqrt{2}-1}\left(\pi\sqrt{\frac{2}{3}}\log e\right)\sqrt{n} = \frac{4\pi\log e}{(2-\sqrt{2})\sqrt{3}}\sqrt{n}.$$

This implies an estimator with per-symbol attenuation $2^{\mathcal{O}(\frac{1}{\sqrt{n}})}$. However, this estimator has high complexity of implementation. Hence, we also derive a linear-complexity sequential algorithm whose redundancy is at most $\mathcal{O}(n^{2/3})$, where the implied constant is less than 10, implying a linear complexity estimator with per-symbol attenuation $2^{\mathcal{O}(\frac{1}{\sqrt[3]{n}})}$.

8 Applications: Text Classification

Probability estimation, while interesting in itself, also happens to be fundamental to several other applications. Improved probability estimates can potentially lead to better algorithms in machine learning, where large alphabets are handled in several applications. Therefore, a natural test for the large alphabet estimators developed here would be to incorporate them for machine learning tasks.

To illustrate the power of the estimators developed for the estimation problem above, we consider the problem of text classification. In this application, documents from an archive (*e.g.*, archived news reports) are assigned labels in some natural way (*e.g.*, their topic—finance, sports, opinion, etc). The task is to use hand labeled documents as training data in order to predict labels on new documents.

In practice, for text classification one of the techniques known to work reasonably is the so-called "Naive Bayes" approach where, conditioned on a class, words are assumed to be independent of each other. Of course, such an assumption would be very wrong from a linguistic point of view. The naive bayes classifier learns one distribution for each class. Each document is labeled with the class whose associated distribution gives the document the highest probability.

We combine the naive bayes approach with the estimators developed in Section 7.

Because of the nature of the problem, the most commonly accepted evaluations of classification techniques are empirical. There are several large datasets made available as benchmarks, and the results on some of the commmonly used ones are outlined below. Text classification techniques are not strongly dependent on language, and we present results on datasets in English and Portugese.

The collection *Newsgroups* is a list of 1000 articles collected from 20 newgroups. Among the newsgroups are collections of closely

Table 1: Percentage of documents accurately classified.p (SVM: Support Vector Machine classifier; Laplace: Naive Bayes classifier with Laplace smoothing; OSZ: New classifier)

	R52	Nwsgrps	CADE
SVM	*93.57*	73.09	52.84
Laplace	88.43	70.75	50.27
OSZ	91.98	*73.46*	*59.24*

related newsgroups such as comp.os.ms-windows.misc, comp.windows.x or rec.autos and rec.motorcycles. The task is to identify all the newsgroups. Roughly 10% of the documents were randomly chosen for training, while the rest are used for testing, and the results confirmed by repeat trials over random splits.

The other two datasets have much more skewed training data---some classes have thousands of documents, while some have as few as one. The Reuters 21758 R52 dataset is an archive of the Reuters newswire in 1987. It is hand classified into 52 categories, each denoting a topic for the article (cocoa, coffee, income, heat, grain, etc.).

The CADE dataset is a collection of Portugese web documents that are hand classified into 12 classes, each denoting a topic for the document (sports, culture, classifieds, etc.). It is known to be a very hard dataset to classify, and all classifiers have low accuracy for this dataset.

All the data sets can be obtained from [24].

We adapted the code from the libraries provided in [25]. The classification error is the proportion of documents that are misclassified. We compare our performance (OSZ row) with support vector machine (SVM row) results, generally considered to be the state of the art in text classification as well as the previous best results following the Naive Bayes approaches that used Laplace smoothing (Laplace). The improvements reported here are significant at 0.05 level.

Acknowledgements

We thank N. Alon, Y. Freund, I.J. Good, N. Jevtić, Sajama, W. Szpankowski, and K. Viswanathan for helpful discussions.

References and Notes

[1] A. Orlitsky, N.P. Santhanam, and J. Zhang, "Universal Compression of Memoryless Sources over Unknown Alphabets," *IEEE Trans. on Info Theory* 50 (7): 1469-1481, July 2004.

[2] B. Fittingoff, "Universal methods of coding for the case of unknown statistics," In *Proceedings of the 5th Symposium on Information Theory*, pp. 129–135. Moscow-Gorky, 1972.

[3] L.D. Davison, "Universal noiseless coding," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 783–795, Nov. 1973.

[4] J. Shtarkov. Coding of discrete sources with unknown statistics. In I. Csiszár and P. Elias, editors, *Topics in Information Theory (Coll. Math. Soc. J. Bolyai, no. 16)*, pages 559–574. Amsterdam, The Netherlands: North Holland, 1977.

[5] J.C. Kieffer, "A unified approach to weak universal source coding," *IEEE Transactions on Information Theory*, vol. 24, no. 6, pp. 674–682, Nov. 1978.

[6] J. Rissanen, "Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 629–636, Jul. 1984.

[7] N. Merhav and M. Feder, "Universal prediction," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2124–2147, Oct. 1998.

[8] T.M. Cover, "Universal portfolios," *Mathematical Finance*, vol. 1, no. 1, pp. 1–29, Jan. 1991.

[9] N. Littlestone and M.K. Warmuth, "The weighted majority algorithm," In *IEEE Symposium on Foundations of Computer Science*, 1992.

[10] V.G. Vovk, "A game of prediction with expert advice," *Journal of Computer and System Sciences*, vol. 56, no. 2, pp. 153–173, 1998.

[11] N. Cesa-Bianchi and G. Lugosi, "Minimax regret under log loss for general classes of experts," In *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*, pp. 12–18, 1999.

[12] P.S. Laplace, *Philosphical Essays on Probabilities*. Springer Verlag, New York, Translated by A.I. Dale from the 5th (1825) edition, 1995.

[13] A. Orlitsky, N.P. Santhanam, IEEE Trans. on Info Theory 50 (10): 2215-2230 October 2004.

[14] I.J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, no. 3/4, pp. 237–264, Dec. 1953.

[15] F.H. Hinsley and A. Stripp. *Codebreakers: The Inside Story of Bletchley Park*. Oxford University Press, 1993.

[16] F. Song and W.B. Croft, "A general language model for information retrieval (poster abstract)," In *Research and Development in Information Retrieval*, pp. 279–280, 1999.

[17] K.W. Church and W.A. Gale, "Probability scoring for spelling correction," *Statistics and Computing*, vol. 1, pp. 93–103, 1991.

[18] W.A. Gale, K.W. Church, and D. Yarowsky. A method for disambiguating word senses. *Computers and Humanities*, vol. 26, pp. 415–419, 1993.

[19] S.F. Chen and J. Goodman. An empirical study of smoothing techniques for language modeling. In *Proceedings of the Thirty-Fourth Annual Meeting of the Association for Computational*

Linguistics, pp. 310–318, San Francisco, 1996. Morgan Kaufmann Publishers.

[20] A. Nadas, "On Turing's formula for word probabilities," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-33, no. 6, pp. 1414–1416, Dec. 1985.

[21] A. Nadas, "Good, Jelinek, Mercer, and Robins on Turing's estimate of probabilities," *American Journal of Mathematical and Management Sciences*, vol. 11, pp. 229–308, 1991.

[22] I.J. Good, "Turing's anticipation of Empirical Bayes in connection with the cryptanalysis of the Naval Enigma," *Journal of Statistics Computation and Simulation*, vol. 66, pp. 101–111, 2000.

[23] D. McAllester and R. Schapire, "On the convergence rate of Good Turing estimators," In *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, 2000.

[24] Ana Cardoso-Cachopo. Datasets for single-label text categorization. From: http://www.gia.ist.utl.pt/acardoso/datasets/.

[25] Andrew Kachites McCallum. Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. http://www.cs.cmu.edu/mccallum/bow, 1996.

Reflections on the Capacity Region of the Multi-Antenna Gaussian Broadcast Channel

Hanan Weingarten, Yossef Steinberg, and Shlomo Shamai (Shitz)

From left to right: Yossef Steinberg, Hanan Weingarten, and Shlomo Shamai (Shitz).

Abstract—**We give an overview of the results and techniques used to determine the capacity region of the Multi-Antenna Broadcast Channel [18]. We provide a brief historical perspective on this problem, and finally, indicate recent developments.**

I. Introduction

The problem of the Gaussian multi-antenna broadcast channel (BC) has been in the limelight of information theoretic research in recent years and turned out to be most relevant in modern communication systems. This short overview is intended to provide a taste of this exciting field and mainly to highlight, the results and techniques, reported in [18], where the capacity region of this channel has been established (This paper is the winner of the 2007 IT Society Best Paper Award). Further, a short historical perspective as well as some recent developments are also mentioned, by pointing out relevant references. We examine a BC that has one transmitter which sends several independent messages to several users, numbered $k = 1, \ldots, K$, and equipped with receivers which cannot cooperate with one another. Each message is intended for a different user. Therefore, such a channel is used to model the downlink channel in cellular systems, other wireless links and ADSL wire-line links. Independently, there has been a massive interest in multi-antenna channels, mainly in the context of a point to point channel. In such a channel both the transmitter and receiver are equipped with several antennas, thereby obtaining almost a linear increase in the channel capacity, proportional to the minimum number of transmit and receive antennas, without requiring additional bandwidth or transmit power [12]. It is no wonder then that there has been interest in combining the broadcast and multi-antenna models, into a natural setting of a multi-antenna BC.

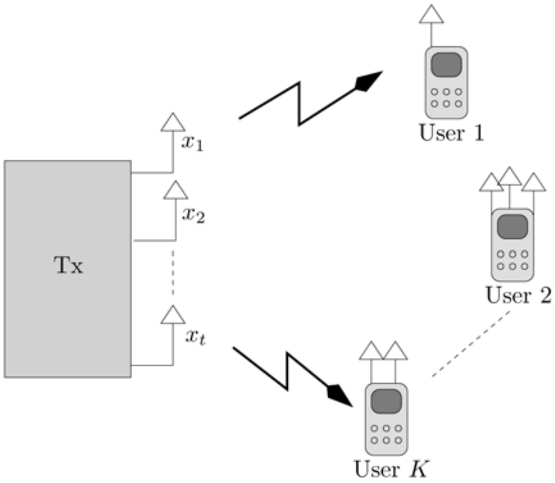


Figure 1. A multi-antenna broadcast channel.

In a Gaussian multi-antenna BC, the transmitter is equipped with t antennas and each of the receivers is equipped with r_i antennas (see Figure 1). Due to a multitude of reflections in wireless links and wire-line coupling in ADSL links, the receivers obtain some linear mixture of the transmitted signals. By using a vector notation to denote the transmitted and received signals at any given time, we can use the following vector equation to define a time sample of the BC

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \qquad k = 1, \ldots, K, \tag{1}$$

where $\mathbf{y}_k$ is a $r_k \times 1$ vector received by user k, $\mathbf{H}_k$ is an $r_k \times t$ fading matrix, defining the linear mixture, and $\mathbf{x}$ is a $t \times 1$ vector which elements denote the levels being transmitted. $\mathbf{n}_k \sim \mathcal{N}(0, \mathbf{I})$ is a Gaussian noise vector added at the receiver of user k. In addition we assume that there is a power constraint, P, on the input such that $E[\mathbf{x}^T\mathbf{x}] < P$.

We can consider several encoding methods for the multi-antenna BC. Linear encoding schemes are the simplest to implement yet show the great appeal of using multi-antenna transmitters in a BC setting. As an example, we consider the simple case of two users, each equipped with one antenna and a transmitter equipped with two antennas. The transmitter constructs the following signal:

$$\mathbf{x} = s_1 \mathbf{u}_1 + s_2 \mathbf{u}_2 \tag{2}$$

where s_1 and s_2 are scalars carrying independent information streams for the first and second users, respectively, and $\mathbf{u}_1$ and $\mathbf{u}_2$ are 2×1 unitary vectors which form beams which direct the information streams at the direction of the respective users.

At each user we receive the following signal

$$y_1 = \mathbf{h}_1^T \mathbf{x} + n_1 = s_1 \mathbf{h}_1^T \mathbf{u}_1 + z_1$$
$$y_2 = \mathbf{h}_2^T \mathbf{x} + n_2 = s_2 \mathbf{h}_2^T \mathbf{u}_2 + z_2$$

where $z_1 = s_2 \mathbf{h}_1^T \mathbf{u}_2 + n_1$ and $z_2 = s_1 \mathbf{h}_2^T \mathbf{u}_1 + n_2$ are the overall interference at each receiver. When zero-forcing beam-forming is considered, the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are chosen in such a manner that $\mathbf{h}_1^T \mathbf{u}_2 = \mathbf{h}_2^T \mathbf{u}_1 = 0$. Thus, the signal intended to one user does not interfere at the receiver of the second user. Alternatively, $\mathbf{u}_1$ and $\mathbf{u}_2$ may be chosen to maximize the signal to interference ratio at the receivers (see, for example, [2], [20]).

Figure 2 is an eye opener and shows the overall transmission rate (sum-rate) that can be obtained when using the beam-forming construction or when using time domain multiple access (TDMA) (i.e., transmitting only to one user at a time). It can be seen that as the SNR increases, the slope (alternatively, multiplexing gain) using beam-forming is twice that obtained when the transmitter directs its signal only to the strongest user. Thus, it is possible to approximately double the communication rates at high SNRs when the transmitter has two antennas. Note that unlike the point to point channel, we did not require that each of the users will be equipped with two antennas.

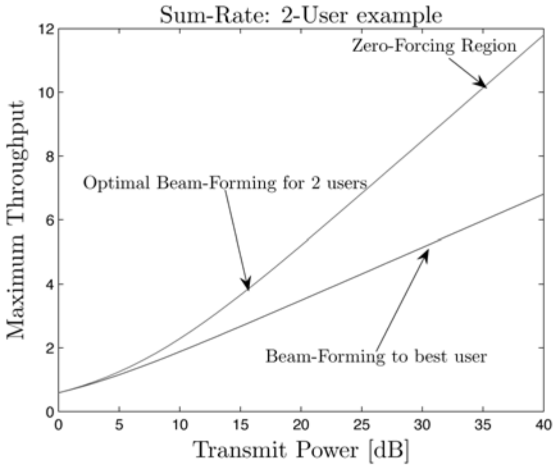


Figure 2. Linear beam-forming and TDMA in a 2 user channel: $\mathbf{h}_1^T = (1 \quad 0.5)$, $\mathbf{h}_2^T = (0.2 \quad 1)$. At high SNRs, zero-forcing beam-forming is best. At mid-range SNRs, the beam-forming vectors are optimized to maximize the sum-rate.

II. Towards the Capacity Region

We are now left with the question of what is the best performance which can be obtained in multi-antenna BC, or alternatively, what is the capacity region of this channel. Unlike the point to point channel, we are interested in a capacity region and not a scalar. In a BC we have several rates, one for each receiving user. We collect these rates onto a vector and the capacity region is defined as the set of all rate vectors which are achievable with an arbitrarily small probability of error for sufficiently large codeword lengths. The first to address this question for the multi antenna BC were Caire and Shamai [4]. In their paper, they suggested using a nonlinear coding scheme known as dirty paper coding (DPC). This coding scheme is based on a result by Costa [5] who investigated a scalar, point to point Gaussian channel, which apart from additive noise also suffers from a Gaussian interference which is known non-causally at the transmitter but not at the receiver. A time sample of this channel is defined by $y = x + s + n$ where x is a power constrained channel input ($Ex^2 < P$), $s \sim \mathcal{N}(0, S)$ is a gaussian interference which is non-causally known at the transmitter but not at the receiver and $n \sim \mathcal{N}(0, N)$ is an additive noise which is not known at either end.

Costa showed that the capacity of this channel is given by $\frac{1}{2}\log(1 + \frac{P}{N})$. That is, we can obtain the same rate as if the interference, s, did not exist. This is not trivial as the input is power constrained and the interference can not be directly cancelled out. Costa referred to coding in such a scenario as writing on dirty paper, as the interference which is non-causally known at the transmitter may be compared to an initially dirty paper over which we may wish to deliver a message.

Caire and Shamai suggested using this scheme for transmitting over a BC. To demonstrate this, let us return to the two user example and the linear encoding scheme as in (2). Now assume that we use a standard Gaussian codebook for encoding the signal s_1. Note that when encoding the signal s_2, the transmitter is fully aware of the interference signal $s_1\mathbf{h}_2^T\mathbf{u}_1$ and may treat it as non-causally known interference at the receiver y_2. Thus, using DPC as costa did, the effect of the interference may be completely eliminated when deciphering s_2 at user 2. On the other hand, user 1 still suffers from interference due to the signal s_2. Thus, the following rates can be obtained:

$$R_1 \le \frac{1}{2}\log\left(1+\frac{S_1|\mathbf{h}_1^T\mathbf{u}_1|^2}{1+S_2|\mathbf{h}_1^T\mathbf{u}_2|^2}\right)$$
$$R_2 \le \frac{1}{2}\log\left(1+S_2|\mathbf{h}_2^T\mathbf{u}_2|^2\right) \qquad (3)$$

where (S_1, S_2) such that $S_1 + S_2 \le P$ are the powers allotted for the transmission of the two signals. Note that the encoding order may also be reversed. DPC may be used to encode the signal s_1 on top of s_2. The resulting rates will be the same as in the equation above, except the indexes of the users will be reversed.

Figure 3 shows the achievable rate regions using TDMA, linear precoding and DPC with optimized beamforming vectors and power allocations. Evidently, DPC performs bettter than the other options and a natural question is whether the DPC region is the capacity region. It should be noted that unlike the point to point channel, there is no known single letter information theoretic expression for the capacity region of the general BC. See [6] for a short overview. Yet, there are some classes of the BC for which we do have an expression for the capacity region. One such class is the class of stochastically degraded BCs [7]. In this class, the output of one user is in essence the output of the other user after it has been passed through an additional point to point channel. One BC which belongs to this class is the Gaussian BC with a single transmit antenna. In this case, the output of one user is equal to a linear combination of the output of the other user and an independent Gaussian noise. The multi-antenna BC is not degraded in general and that poses a major hardship in characterizing its capacity region, as even a general information theoretic characterization is unavailable, let alone the optimized variables therein.

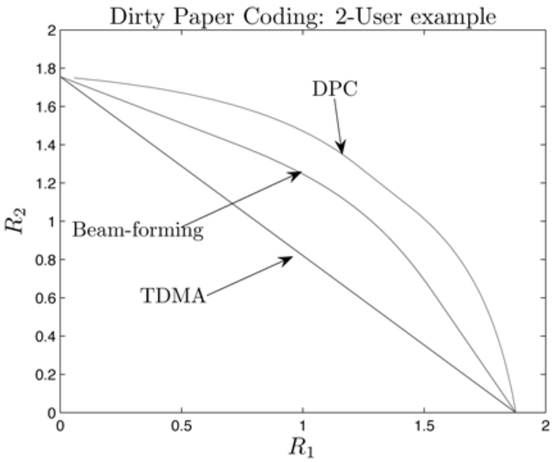


Figure 3. TDMA, Linear beam-forming, and DPC regions in a 2 user channel: $\mathbf{h}_1^T = (1 \quad 0.5)$, $\mathbf{h}_2^T = (0.2 \quad 1)$ and $P = 10$.

Nevertheless, Caire and Shamai [4] were able to show, using the Sato outer bound that in a two user BC with receivers that have a single antenna, the sum-rate points on the boundary of the DPC region coincide with the capacity region. These points are also known as sum-capacity points. That is, the points that lie on the line $R_1 + R_2 = \max(R_1 + R_2)$. This result was later extended to any multi-antenna BC in [14], [16], [21], introducing interesting concepts of duality between Gaussian multi-antenna broadcast and multiple-access channels.

III. The Enhanced Channel and the Capacity Region

In our work [18], we have shown that all points on the boundary of the DPC region correspond to the capacity region. To simplify the problem, we consider a vector version of the channel where the number of receive antennas at each user is equal to the number of transmit antennas and the fading matrices are identity matrices. However, we now allow the noise covariance matrices to take any form. Therefore, a time sample of the channel takes the following form

$$\mathbf{y}_i = \mathbf{x} + \mathbf{n}_i, \qquad i = 1, 2 \qquad (4)$$

where $\mathbf{n}_i \sim \mathcal{N}(0, \mathbf{N}_i)$ ($\mathbf{N}_i$ is a positive semi-definite -PSD- matrix). It is not difficult to see that if indeed all matrices $\mathbf{H}_i$ in (1) are square and invertible, the BCs defined by (4) and (1) are equivalent if $\mathbf{N}_i = \mathbf{H}_i^{-1}\mathbf{H}_i^{-T}$. If the fading matrices are not invertible, it is possible to show that the BC in (1) may be approximated by a sequence of vector BCs (4) in which some of the eigenvalues of $\mathbf{N}_i$ go to infinity [18]. An additional restriction is put on the power constraint. Instead of a total power constraint where we limit the total power transmitted over all antennas, we consider a covariance matrix constraint such that $E\mathbf{x}\mathbf{x}^T \le \mathbf{S}$ ($\mathbf{S}$ is a PSD matrix). The capacity region under such a matrix constraint may be generalized to a broad range of power constraints such as the total power constraints and individual power constraints.

In order to transmit over this BC we consider a vector version of DPC [22]. We transmit a superposition of two signals $\mathbf{x} = \mathbf{s}_1 + \mathbf{s}_2$ where $\mathbf{s}_1$ is a codeword taken from a Gaussian codebook and $\mathbf{s}_2$ is a DPC codeword, where $\mathbf{s}_1$ acts as a non-causally known interference. Thus, we obtain the vector version of (3) as follows:

$$R_1 \le \frac{1}{2}\log\frac{|\mathbf{S}+\mathbf{N}_1|}{|\mathbf{S}_2+\mathbf{N}_1|}$$
$$R_2 \le \frac{1}{2}\log\frac{|\mathbf{S}_2+\mathbf{N}_2|}{|\mathbf{N}_2|} \qquad (5)$$

where $E\mathbf{s}_1\mathbf{s}_1^T = \mathbf{S} - \mathbf{S}_2$ and $E\mathbf{s}_2\mathbf{s}_2^T = \mathbf{S}_2$ are covariance matrices allotted for the messages. We can also switch between the users, reversing the precoding order, thus obtaining the same rates with

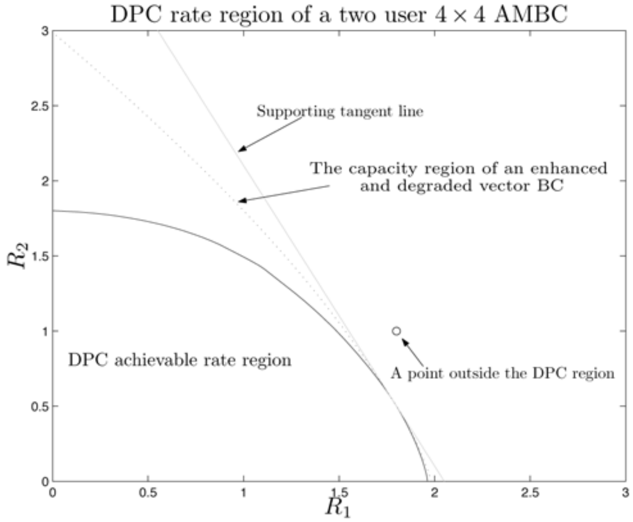


Figure 4. Schematic view of the capacity region proof.

the user indexes reversed.

Note that if $\mathbf{N}_2 \leq \mathbf{N}_1$ (i.e., $\mathbf{N}_1 - \mathbf{N}_2$ is a PSD), the vector BC is a degraded BC as the output of the first user may be given by $\mathbf{y}_1 = \mathbf{y}_2 + \tilde{\mathbf{n}}$ where $\tilde{\mathbf{n}} \sim \mathcal{N}(0, \mathbf{N}_1 - \mathbf{N}_2)$. Yet, even though there is an information theoretic expression for the capacity region of a degraded BC, proving that (5) gives the capacity region in the vector case is an elusive task (see [13], [15]). The difficulty arises as the central tool, the entropy power inequality as used by Bergmans in the Gaussian scalar broadcast channel [1], is not tight for Gaussian vectors, unless they posses proportional covariance matrices.

We circumvent this hurdle by the introduction of an *enhanced channel*. A vector BC with noise covariances $\mathbf{N}'_2 \leq \mathbf{N}_2$ and $\mathbf{N}'_1 \leq \mathbf{N}_1$ is said to be an enhanced version of the vector BC in (4). We showed [18] that for a point on the boundary of the DPC region of a degraded vector BC, denoted by (R_1^*, R_2^*) and obtained by choosing $\mathbf{S}_2 = \mathbf{S}_2^*$, there exists an enhanced and degraded version of the vector BC such that the same rates are obtained in the enhanced BC using the same power allocation. Furthermore, $\mathbf{S}_2 + \mathbf{N}'_2$ is proportional to $\mathbf{S}_2 + \mathbf{N}'_1$. This proportionality result allowed us to show that indeed (R_1^*, R_2^*) lies on the capacity region of the enhanced vector BC using Bergmans' classical approach [1]. Note that as the capacity region of the enhanced channel must contain that of the original channel (the enhanced channel has less noise), (R_1^*, R_2^*) must also lie on the capacity region of the original vector BC. Furthermore, as this can be done for every point on the boundary of the region in (5), this region must be the capacity region.

We generalized the proof to non-degraded vector BCs by finding a set of enhanced channels for this more general set of BCs. Note that the DPC region must always be convex as time sharing may be used. Therefore, any point that lies outside this region may be separated from the region by a tangent line. We showed [18] that for every tangent line we can find a specific enhanced and degraded vector BC which capacity region is also supported by the same tangent line (see Figure 4). Therefore, we conclude that every point outside the DPC region also lies outside the capacity region of an enhanced and degraded BC. However, as the capacity region of the enhanced channel contains that of the original channel, we conclude that every point outside the DPC region must lie outside the capacity region of the original channel and thus prove that the DPC region (5) is indeed the capacity region. In [18] this result is extended to any multi-antenna BC (1) and any linear power constraint on the input.

IV. Future Work

Up to this point we assumed that each message has only one intended user. However, we may also consider the case where some of the messages are common to a number of users. The capacity region in this case is still an open problem and some progress has been reported in [17], [19]. Another open problem is the capacity region of fading and compound Gaussian BCs. We assumed that the fading matrix is perfectly known at the transmitter (and receivers). However, in reality, the transmitter may have only a partial knowledge of the fading conditions, giving rise to a fading or a compound BC [8], [9]and [23].

Applications and implications of the multi-antenna broadcast channel in wireless and wire-line communications are reported extensively in recent years. The impact of theoretically motivated state of the art communications technologies for this channel can be substantial as is demonstrated in certain cellular models [10]. For an extended perspective and details see [3], [11].

References

[1] Patrick P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise," *IEEE Trans. on Information Theory*, vol. 20, no. 2, pp. 279–280, Mar. 1974.

[2] H. Boche, M. Schubert, and E.A. Jorswieck, "Throughput maximization for the multiuser MIMO broadcast channel," in *proceedings of Acoustics, Speech and Signal Processing (ICASSP03)*, pp. 4.808–4.811, Apr. 5–10, 2003.

[3] G. Caire, S. Shamai, Y. Steinberg, and H. Weingarten, *Space-Time Wireless Systems: From Array Processing to MIMO Communications*, chapter On Information Theoretic Aspects of MIMO-Broadcast Channels, Cambridge University Press, Cambridge, UK, 2006.

[4] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[5] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[6] T.M. Cover, "Comments on broadcast channels," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2524–2530, Sept. 1998.

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

[8] S. Shamai (Shitz), H. Weingarten, and G. Kramer, "On the compound MIMO broadcast channel," in *Proceedings of the Information*

Theory and Applications Workshop (ITA 2007), UCSD, San Diego, CA, Jan. 29–Feb. 2, 2007.

[9] A. Lapidoth, S. Shamai (Shitz), and M.A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," http://www.arxiv.org/pdf/cs.IT/0605079, p. arXiv:cs.IT/0605079, May 2006.

[10] O. Somekh, B. Zaidel and S. Shamai (Shitz), "Sum rates characterization of joint multiple cell-site processing", *IEEE Trans. on Information Theory*, vol. 53, no. 12, Dec. 2007.

[11] S. Shamai, "Reflections on the gaussian broadcast channel: Progress and challenges," in *Plenary address, http://www.isit2007.org/index.php 2007 IEEE International Symposium on Information Theory (ISIT2007)*, Nice, France, June 24–29, 2007.

[12] I. E. Teletar, "Capacity of multi-anteanna gaussian channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[13] D. Tse and P. Viswanath, "On the capacity of the multiple antenna broadcast channel," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G.J. Foschini and S. Verdú, Eds., pp. 87–105. DIMACS, American Mathematical Society, Providence, RI, 2003.

[14] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of gaussian MIMO broadcast channels," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[15] S. Vishwanath, G. Kramer, S. Shamai (Shitz), S. Jafar, and A. Goldsmith, "Capacity bounds for gaussian vector broadcast channels," in *Multiantenna Channels: Capacity, Coding and Signal Processing*, G. J. Foschini and S. Verdú, Eds., pp. 107–122. DIMACS, American Mathematical Society, Providence, RI, 2003.

[16] P. Viswanath and D. Tse, "Sum capacity of the vector gaussian channel and uplink-downlink duality," *IEEE Trans. on Information Theory*, vol. 49, no. 8, pp. 1912– 1921, Aug. 2003.

[17] H. Weingarten, T. Liu, S. Shamai (Shitz), Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple input multiple output broadcast compound channel," *Submitted to IEEE Trans. on Information Theory.*, 2007.

[18] H.Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. on Information Theory*, vol. 52, no. 9, pp. 3936–3964, Sept. 2006.

[19] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "On the capacity region of the multi-antenna broadcast channel with common messages," in *2006 IEEE Int. Symp. on Inf. Theory (ISIT2006)*, Seattle, Washington, USA, July 9–14 2006, pp. 2195–2199.

[20] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Trans. on Signal Processing*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[21] W. Yu and J. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Trans. on Information Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.

[22] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang, "Writing on colored paper," in *proc. of Int. Symp. on Inf. Theory (ISIT2001)*, Washington, DC, p. 309, June 24–29, 2001.

[23] D. Tuninetti and S. Shamai (Shitz), "Fading gaussian broadcast channels with state information at the receiver," Advances in Network Information Theory, P. Gupta, G. Kramer, and A. J. van Wijngaarden, Eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol.~66, American Mathematical Society, 2004.

Reflections on Fountain Codes

Michael Luby and Amin Shokrollahi

I. INTRODUCTION

The binary erasure channel (BEC) of communication was introduced by Elias [1] in 1955, but it was regarded as a rather theoretical channel model until the large-scale deployment of the Internet about 40 years later.

On Internet Protocol (IP) based networks data is transmitted in the form of packets. Each packet is equipped with a header that describes the source and the destination of the packet, and often also a sequence number describing the absolute or relative position of the packet within a given stream. These packets are routed on the network from the sender to the receiver. Due to various reasons, for example buffer overflows at the intermediate routers, interference and poor signal conditions, some packets may get lost and never reach their destination. Other packets may be declared as lost if the internal checksum of the packet does not match. Therefore, an IP-based network is a very good real-world model of the BEC.

The general Internet is the largest and most well-known example of an IP-based network. Increasingly, other network infrastructure and applications has moved to be IP-based, e.g., mobile phone cellular networks, Television cable, satellite and DSL networks, defense networks, automobile broadcast networks, etc.

Michael Luby is co-founder and Chief Technology Officer of Digital Fountain, Inc. Amin Shokrollahi is a professor of Mathematics and Computer Science at EPFL, and the Chief Scientist of Digital Fountain.

Reliable transmission over IP-based networks has been the subject of much research. For the most part, reliability is guaranteed by use of appropriate protocols. For example, the ubiquitous TCP/IP ensures reliability based on a sender/receiver packet-by-packet acknowledgment protocol. It is well-known that such protocols exhibit poor behavior in many cases, such as transmission of data from one sender to multiple receivers, or transmission of data over impaired channels, such as wireless or satellite links.

For these reasons other transmission solutions have been proposed. One class of such solutions is based on Forward Erasure Correction (FEC) coding over multiple packets. The original data is encoded using some erasure correcting code. If during the transmission some part of the data is lost, then it is possible to recover the lost data using erasure correcting algorithms. For applications it is crucial that the codes used are capable of protecting a flexible amount of data (from very small to very large), of correcting a flexible number of erasures (from very few to a lot), that they operate close to optimally irrespective of the loss characteristics of the underlying network, and that they have very fast encoding and decoding algorithms.

To satisfy all these requirements, a new class of codes is needed. Fountain Codes constitute such a class. Suitable subclasses of these codes, like LT- and Raptor Codes satisfy all the requirements given above. An example of the usage of such codes is the following. A server sending data over a one-way broadcast network to many recipients can apply a Fountain Code for a given piece of data to generate as many encoding packets as needed on the fly. When the receiver joins the session, packets are received by the recipient, who collects the output symbols, and leaves the transmission as soon as it has received enough of them. More loss of symbols just translates to a longer waiting time to receive the desired number of packets.

In order to make Fountain Codes work in practice, one needs to ensure that they possess a fast encoder and decoder, and that the decoder is capable of recovering the original symbols from any set of output symbols whose size is close to optimal with high probability. We call such Fountain Codes *universal*. The first class of universal Fountain Codes was invented by Luby [2], [3]. The codes in this class are called LT-Codes. The other class of Fountain Codes, called Raptor codes [4], [5] generalize LT-Codes by adding a precoder. This generalization makes it possible to design linear time encoding and decoding algorithms while keeping the universality. Highly optimized versions of Raptor Codes are being used in commercial systems of Digital Fountain, a company specializing in fast and reliable delivery of data over heterogeneous networks. A version of Raptor Codes that works for a flexible range of data lengths and encoding lengths, and requires very little processing power, has been selected into a variety of standards, such as 3GPP MBMS for multimedia broadcast streaming and file delivery, the IETF for multicast/broadcast file delivery and streaming, DVB-H IPDC for broadcast file delivery, DVB IPTV for streaming delivery of television services, and others.

II. FOUNTAIN, LT, AND RAPTOR CODES

Fountain Codes are a novel class of codes designed for transmission of data over time varying and unknown erasure channels. They were first mentioned without an explicit construction in [6], and the first efficient construction was invented by Luby [2]. One type of Fountain Codes designed for k source symbols may be specified by a probability distribution $\mathcal{D}$ on the set of binary strings of length k. Operationally, such a Fountain Code can produce from the vector x a potentially limitless stream of symbols $y_1, y_2, y_3, \ldots$, called *output symbols*. Each output symbol is independently generated by sampling the distribution $\mathcal{D}$ to obtain a mask $(a_1, \ldots, a_k)$. The value of the output symbol is then $\bigoplus_{i=1}^{k} a_i x_i$, where $\bigoplus$ denotes the XOR operation, and $a_i x_i$ is x_i if $a_i = 1$, and is 0 else. The fundamental requirement is that the source symbols can be recovered from any set of n output symbols with high probability, where n is close to the length k of x. The number $n/k— 1$ is called the overhead of the decoder. The last condition shows that such Fountain Codes are robust against erasures, since only the number of received output symbols is important for decoding. In operation the output symbols need to contain indications that allow the receiver to recover the mask of each of these symbols. This is accomplished by equipping output symbols with Encoding Symbol ID's (ESI's). In the standardized Raptor code, an ESI is a 16-bit integer which facilitates the creation of the mask associated to an output symbol [7].

Different Fountain Codes differ in terms of their overhead for a given error probability. But they also differ in terms of the computational efficiency of the encoding and decoding processes.

LT-Codes, invented by Luby [3], form the first realization of Fountain Codes. LT-Codes exhibit excellent overhead and error probability properties. For LT-Codes the probability distribution $\mathcal{D}$ has a particular form which we describe by outlining its sampling procedure. At the heart of LT-Codes is a probability distribution Ω on the integers $1, \ldots, k$, called the *weight* or *degree* distribution. To create an output symbol, the following procedure is applied: the distribution is sampled to obtain an integer $w \in \{1, \ldots, k\}$; next, a binary vector $(a_1, \ldots, a_k)$ of Hamming weight w is chosen uniformly at random, and the value of the output symbol is set to $\bigoplus_{i=1}^{k} a_i x_i$. Decoding of LT-Codes is done using a greedy algorithm which is the specialization of the beliefpropagation algorithm to the case of the erasure channel [8].

Despite the excellent performance of LTCodes, it is not possible to give a construction with average constant (per symbol) encoding and a decoding cost linear in the length of the input vector, without sacrificing the error probability. In fact, a simple analysis shows that to obtain constant encoding cost with reasonable overheads, the error probability has to be constant as well [5].

An extension of LT-Codes, Raptor Codes are a class of Fountain Codes with constant per symbol encoding and linear decoding cost. They achieve their computational superiority at the expense of an asymptotically higher overhead than LT-Codes, although in most practical settings Raptor Codes outperform LT-Codes in every aspect. Raptor Codes achieve their performance using a simple idea: the source is *precoded* using a suitable linear code. An appropriately chosen LT-code is then applied to the encoded vector to create the output symbols.

Note that LT-Codes form a special subclass of Raptor Codes: for

3

$$
\begin{aligned}
z_1 &= x_1\\
z_5 &= x_2 \oplus z_1\\
z_8 &= x_6 \oplus z_5\\
z_4 &= x_5 \oplus z_8\\
z_3 &= x_3 \oplus z_1 \oplus z_4\\
z_6 &= x_4 \oplus z_1 \oplus z_3 \oplus z_4\\
z_2 &= s_2 \oplus z_1 \oplus z_4 \oplus z_5 \oplus z_6 \oplus z_8\\
z_7 &= s_1 \oplus z_3 \oplus z_4 \oplus z_6
\end{aligned}
$$

$$
\begin{aligned}
y_1 &= z_2 \oplus z_4 \oplus z_5 \oplus z_7\\
y_2 &= z_2 \oplus z_3 \oplus z_6\\
y_3 &= z_3 \oplus z_5 \oplus z_8
\end{aligned}
$$

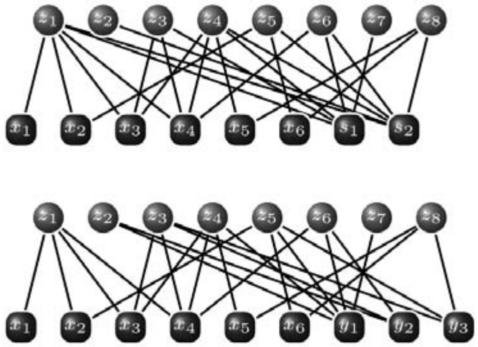


Fig. 1. Toy example of a systematic Raptor code. The source symbols are $x_1, \ldots, x_6$. The nodes with labels s_1, s_2 are obtained from the relations dictated by the precode, and their values are 0. In a first step, the intermediate symbols $z_1, \ldots, z_8$ are obtained from the source symbols by applying a decoder. The sequence of operations leading to the zi is given on the left. Then the output symbols are generated from these intermediate symbols. Examples for three output symbols y_1, y_2, y_3 are provided. Note that by construction the x_i are also XORs of those z_1 to which they are connected.

these codes the precode is trivial. At the other extreme there are the *precode-only* (PCO) codes [5] for which the degree distribution Ω is trivial (it assigns a probability of 1 to weight 1, and zero probability to all other weights). All Raptor Codes in use are somewhere between these two extremes: they have a nontrivial (high-rate) precode, and they have an intricate (though low-weight) degree distribution.

Raptor Codes can be decoded in a variety of ways. The conceptually simplest decoder sets up a system of linear equations and solves it using Gaussian elimination. The system is constructed by obtaining for each received output symbol the set of input symbols of which the output symbol is the XOR. This information is extracted from the ESI of the output symbol. This system describes the values of the collected output symbols in terms of the (unknown) values of the input symbols, and can be solved for the latter if the rank of the matrix is full. This decoder is optimal as far as the success of the recovery procedure is concerned: decoding (by means of *any* algorithm) fails if and only if the Gaussian elimination decoder fails. However, the running time of this decoder is prohibitively large. A different decoder with much lower complexity is the belief-propagation decoder. A modification of this algorithm has been completely analyzed in [5] and designs have been presented which show that the failure probability of the algorithm is very small even for small overheads, if *k* is in the range of tens of thousands.

The superior computational performance of the greedy decoding algorithm comes at the expense of large overheads for small values of *k*. It seems hard to be able to control the variance for small values of *k*. To remedy this situation, a different decoding algorithm has been devised [9]. Called *inactivation decoder*, this decoder combines the optimality of Gaussian elimination with the efficiency of the greedy algorithm.

In a variety of applications it is imperative to have the source symbols as part of the transmission. The straightforward idea of sending the source symbols alongside the normal output symbols of a Raptor code fails miserably. This is because there is large discrepancy between the statistics of the source symbols and that of the repair symbols. A method that *does* work has been outlined in [5] and in [10]. The main idea behind the method is the following: we start with a nonsystematic Raptor code, and generate *k* output symbols. We then see whether it is possible to decode the input symbols using these output symbols. If so, then we identify these output symbols with the source symbols, and decode to obtain a set of *m intermediate symbols*. The repair symbols are then created from the intermediate symbols using the normal encoding process for Raptor Codes. An example of a systematic Raptor Code together with its encoding procedure is provided in Figure 1.

III. SOME APPLICATIONS

Fountain Codes in general, and Raptor Codes in particular, can be used in a variety of data transmission scenarios. In a point-to-point communication where there is one sender and one receiver the use of Raptor Codes allows for faster transmission when compared to, say, a TCP based transmission, particularly when the distance between sender and receiver is large. In a broadcast/multicast application, such as the ones envisioned by the 3GPP MBMS and DVBIP standards the use of Raptor Codes is essential for guaranteeing quality of service, and minimizing the overall connection bandwidth of the sender. In a multipoint-to-point application, in which a receiver attempts to receive the same file from multiple uncoordinated senders, Raptor Codes are used to make sure that the information received from the different senders is indeed "additive," in the sense that it does not make any difference which packet is received from which sender; even if all but one of the senders leave the transmission, the receiver still obtains useful information, albeit at a lower rate. Very similar comments are applicable to multipoint-tomultipoint applications in which many uncoordinated receivers want to receive different pieces of data residing with different senders.

Some applications, in particular those dealing with file download, use the Fountain property of Raptor Codes, and their ability to recover the data in close to optimal time, even on devices with limited computational and memory resources. These applications include data download and data broadcast on mobile using wireless or satellite transmission.

Yet other applications use the property of Raptor Codes that make recovery of a file possible from reception of data from uncoordinated senders. Each sender has access to the same piece of data, and creates an independent fountain for this piece. The receivers then combine data received from each of the senders. Because the output symbols have been generated independently at each of the senders, from the point of view of the receiver the data received may as well have been generated by one fountain, at the cumulative rate of all the fountains. Since this way of receiving data is robust against partial or total failure of the senders, this is the method of choice in applications where there is expectation of such failures.

In other applications, such as IPTV and more generally video transmission over networks, systematic Raptor Codes are used as "tunable fixed rate codes." This means that only a fixed amount of redundancy (or repair data) is generated in a fixed application, this amount may vary depending on the particular network conditions encountered. Each instantiation would correspond to a code of a different rate, but all these codes are generated the same way using the same Raptor code. This flexibility is a key feature for providing video of very good quality on networks with widely varying characteristics.

REFERENCES

[1] P. Elias, "Coding for two noisy channels," in *Information Theory, Third London Symposium*, 1955, pp. 61–76.

[2] M. Luby, "Information additive code generator and decoder for communication systems," October 23 2001, U.S. Patent No. 6,307,487.

[3] ——, "LT codes," in *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002.

[4] A. Shokrollahi, S. Lassen, and M. Luby, "Multi-stage code generator and decoder for communication systems," June 27, 2006, U.S. Patent No. 7,068,729.

[5] A. Shokrollahi, "Raptor Codes," *IEEE Transactions on Information Theory*, no. 6, pp. 2551–2567, June 2006.

[6] J. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in *proceedings of ACM SIGCOMM '98*, 1998.

[7] 3GPP TS 26.346 V6.1.0, *Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service; Protocols and Codecs*, June 2005.

[8] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, "Efficient erasure correcting codes," *IEEE Transactions on Information Theory*, vol. 47, pp. 569–584, 2001.

[9] A. Shokrollahi, S. Lassen, and R. Karp, "Systems and processes for decoding chain reaction codes through inactivation," 2005, U.S. Patent number 6,856,263.

[10] Shokrollahi, A., and Luby, M., "Systematic Encoding and Decoding of Chain Reaction Codes," U.S. Patent 6 909 383, June 21, 2005.

Shannon, Beethoven, and the Compact Disc

Kees A. Schouhamer Immink

Kees A. Schouhamer Immink with Claude Shannon.

About the author:

Dr. Kees A. Schouhamer Immink worked from 1968 till 1998 at Philips Research Labs, Eindhoven. In 1998, he founded Turing Machines Inc, where he currently serves as its CEO and president. Since 1994, he has been an adjunct professor at the Institute for Experimental Mathematics, Essen-Duisburg University, Germany, and a visiting professor at the Data Storage Institute in Singapore. The photo shows the author (left) with Claude Shannon during the awards dinner at the Convention of the Audio Engineering Society (AES) in New York, October 1985, where Shannon received the AES Gold Medal 'For contributions that made digital audio possible'.

Abstract – *An audio compact disc (CD) holds up to 74 minutes, 33 seconds of sound, just enough for a complete mono recording of Ludwig von Beethoven's Ninth Symphony ('Alle Menschen werden Brüder') at probably the slowest pace it has ever been played, during the Bayreuther Festspiele in 1951 and conducted by Wilhelm Furtwängler. Each second of music requires about 1.5 million bits, which are represented as tiny pits and lands ranging from 0.9 to 3.3 micrometers in length. More than 19 billion channel bits are recorded as a spiral track of alternating pits and lands over a distance of 5.38 kilometers (3.34 miles), which are scanned at walking speed, 4.27 km per hour.*

This year it is 25 years ago that Philips and Sony introduced the CD. In this jubilee article I will discuss the various crucial technical decisions made that would determine the technical success or failure of the new medium.

Shaking the tree

In 1973, I started my work on servo systems and electronics for the videodisc in the Optics group of Philips Research in

Eindhoven. The videodisc is a 30 cm diameter optical disc that can store up to 60 minutes of analog FM-modulated video and sound. It is like a DVD, but much larger, heavier, and less reliable. The launch of the videodisc in 1975 was a technical success, but a monumental marketing failure since the consumers showed absolutely no interest at all. After two years, Philips decided to throw in the towel, and they withdrew the product from the market.

While my colleagues and I were working on the videodisc, two Philips engineers were asked to develop an audio-only disc based on optical videodisc technology. The two engineers were recruited from the audio department, since my research director believed a sound-only disc was a trivial matter given a video and sound videodisc, and he refused to waste costly researcher's time. In retrospect, given the long forgotten videodisc and the CD's great success, this seems a remarkable decision.

The audio engineers started by experimenting with an analog approach using wide-band frequency modulation as in FM radio. Their experiments revealed that the analog solution was scarcely more immune to dirt and scratches than a conventional analog LP. Three years later they decided to look for a digital solution. In 1976, Philips demonstrated the first prototypes of a digital disc using laser videodisc technology. A year later, Sony completed a prototype with a 30 cm diameter disc, the same as the videodisc, and 60 minutes playing time [2].

The Sony/Philips alliance

In October 1979, a crucial high-level decision was made to join forces in the development of a world audio disc standard. Philips and Sony, although competitors in many areas, shared a long history of cooperation, for instance in the joint establishment of the compact cassette standard in the 1960's. In marketing the final products, however, both firms would compete against each other again. Philips brought its expertise and the huge videodisc patent portfolio to the alliance, and Sony contributed its expertise in digital audio technology. In addition, both firms had a significant presence in the music industry via CBS/Sony, a joint venture between CBS Inc. and Sony Japan Records Inc. dating from the late 1960s, and Polygram, a 50% subsidiary of Philips [4]. Within a few weeks, a joint task force of experts was formed. As the only electronics engineer within the 'Optics' research group, I participated and dealt with servos, coding, and electronics at large. In 1979 and 1980, a number of meetings, alternating between Tokyo and Eindhoven, were held. The first meeting, in August 1979 in Eindhoven, and the second meeting, in October 1979 in Tokyo, provided an opportunity for the engineers to get to know each other and to learn each other's main strengths. Both companies had shown prototypes and it was decided to take the best of both worlds. During the third technical meeting on December 20, 1979, both partners wrote down their list of preferred main specifications for the audio disc. Although there are many other specifications, such as the dimensions of the pits, disc thickness, diameter of the inner hole, etcetera, these are too technical to be discussed here.

Item	**Philips**	**Sony**
Sampling rate (kHz)	44.0 - 44.5	44.1
Quantization	14 bit	16 bit
Playing time (min)	60	60
Diameter (mm)	115	100
EC Code	t.b.d.	t.b.d.
Channel Code	M3	t.b.d.

t.b.d. = to be discussed

As can be seen from the list, a lot of work had to be done as the partners agreed only on one item, namely the one-hour playing time. The other target parameters, sampling rate, quantization, and notably disc diameter look very similar, but were worlds apart.

Shannon-Nyquist sampling theorem

The Shannon-Nyquist sampling theorem dictates that in order to achieve lossless sampling, the signal should be sampled with a frequency at least twice the signal's bandwidth. So for a bandwidth of 20 kHz a sampling frequency of at least 40 kHz is required. A large number of people, especially young people, are perfectly capable of hearing sounds at frequencies well above 20 kHz. That is, in theory, all we can say. In 1978, each and every piece of digital audio equipment used its own 'well-chosen' sampling frequency ranging from 32 to 50 kHz. Modern digital audio equipment accepts many different sampling rates, but the CD task force opted for only one frequency, namely 44.1 kHz. This sampling frequency was chosen mainly for logistics reasons as will be discussed later, once we have explained the state-of-the-art of digital audio recording in 1979.

Towards the end of the 1970s, 'PCM adapters' were developed in Japan, which used ordinary analog video tape recorders as a means of storing digital audio data, since these were the only widely available recording devices with sufficient bandwidth. The best commonly-available video recording format at the time was the 3/4" U-Matic.

The presence of the PCM video-based adaptors explains the choice of sampling frequency for the CD, as the number of video lines, frame rate, and bits per line end up dictating the sampling frequency one can achieve for storing stereo audio. The sampling frequencies of 44.1 and 44.056 kHz were the direct result of a need for compatibility with the NTSC and PAL video formats. Essentially, since there were no other reliable recording products available at that time that offered other options in sampling rates, the Sony/Philips task force could only choose between 44.1 or 44.056 KHz and 16 bits resolution (or less).

During the fourth meeting held in Tokyo from March 18-19, 1980, Philips accepted (and thus followed Sony's original proposal) the 16-bit resolution and the 44.1 kHz sampling rate. 44.1 kHz as opposed to 44.056 kHz was chosen for the simple reason that it was easier to remember. Philips dropped their wish to use 14 bits resolution: they had no technical rationale as the wish for the 14 bits was in fact only based on the availability of their 14-bit digital-analog converter. In summary, Compact Disc sound quality followed the sound quality of Sony's PCM-1600 adaptor since logistically speaking there was no other choice.

Thus, quite remarkably, in recording practice, an audio CD starts life as a PCM master tape, recorded on a U-Matic videotape cassette, where the audio data is converted to digital information superimposed within a standard television signal. The industry standard hardware to do this was the Sony PCM-1600, the first commercial video-based 16-bit recorder, followed by the PCM-1610 and PCM-1630 adaptors. Until the 1990s, only video cassettes could be used as a means for exchanging digital sound from the studios to the CD mastering houses. Later, Exabyte computer tapes, CD-Rs and memory sticks have been used as a transport vehicle.

Coding systems

Coding techniques form the basis of modern digital transmission and storage systems. There had been previous practical applications of coding, especially in space communications, but the Compact Disc was the first mass-market electronics product equipped with fully-fledged error correction and channel coding systems. To gain an idea of the types of errors, random versus burst errors, burst length distribution and so on, we made discs that contained known coded sequences. Burst error length distributions were measured for virgin, scratched, or dusty discs. The error measurement was relatively simple, but scratching or fingerprinting a disc in such a way that it can still be played is far from easy. How do you get a disc with the right kind of sticky dust? During playing, most of the dust fell off the disc into the player, and the optics engineers responsible for the player were obviously far from happy with our dust experiments. The experimental discs we used were handmade, and not pressed as commercial mass-produced polycarbonate discs. In retrospect, I think that the channel characterization was a far from adequate instrument for the design of the error correction control (ECC).

There were two competing ECC proposals to be studied. Experiments in Tokyo and Eindhoven -Japanese dust was not the same as Dutch dust- were conducted to verify the performance of the two proposed ECCs. Sony proposed a byte-oriented, rate 3/4, Cross lnterleaved Reed-Solomon code (CIRC) [6]. Vries of Philips designed an interleaved convolutional, rate 2/3, code having a basic unit of information of 3-bit characters [9]. CIRC uses two short RS codes, namely (32, 28, 5) and (28, 24, 5) RS codes using a Ramsey-type of interleaver. If a major burst error occurs and the ECC is overloaded, it is possible to obtain an approximation of an audio sample by interpolating the neighboring audio samples, so concealing uncorrectable samples in the audio signal. CIRC has various nice features to make error concealment possible, so extending the player's operation range [10].

CIRC showed a much higher performance and code rate (and thus playing time), although extremely complicated to cast into silicon at the time. Sony used a 16 kByte RAM for data interleaving, which, then, cost around $50, and added significantly to the sales price of the player. During the fifth meeting in Eindhoven, May 1980, the partners agreed on the CIRC error correction code since our experiments had shown its great resilience against mixtures of random and burst errors. The fully correctable burst length is about 4.000 bits (around 1.5 mm missing data on the disc). The length of errors that can be concealed is about 12.000 bits (around 7.5 mm). The largest error burst we ever measured during the many long days of disc channel characterization was 0.1 mm.

We also had to decide on the channel code. This is a vital component as it has a great impact on both the playing time and the quality of 'disc handling' or 'playability'. Servo systems follow the track of alternating pits and lands in three dimensions, namely radial, focal, and rotational speed. Everyday handling damage, such as dust, fingerprints, and tiny scratches, not only affects retrieved data, but also disrupts the servo functions. In worst cases, the servos may skip tracks or get stuck, and error correction systems become utterly worthless. A well-designed channel code will make it possible to remove the major barriers related to these playability issues.

Both partners proposed some form of (d, k) runlength-limited (RLL) codes, where d is the minimum number and k is the maximum number of zeros between consecutive ones. RLL codes had been widely used in magnetic disk and tape drives, but their application to optical recording was a new and challenging task. The various proposals differed in code rate, runlength parameters d and k, and the so-called spectral content. The spectral content has a direct bearing on the playability, and we had to learn how to trade playability versus the code rate (and thus playing time). In their prototype, Philips used the propriety M3 channel code, a rate $\frac{1}{2}$, d=1, k=5 code, with a well-suppressed spectral content [1]. M3 is a variation on the M2 code, which was developed in the 1970s by Ampex Inc. for their digital video tape recorder [5]. Sony started with a rate 1/3, d=5, RLL code, but since our experiments showed it did not work well, they changed horses halfway, and proposed a propriety rate $\frac{1}{2}$, d=2, k=7 code, a type of code that was used in an IBM magnetic disk drive. Both Sony codes did not have spectral suppression, and the engineers had opposing views on how the servo issue could be solved. Synchronization of signals with unknown speed read with constant linear velocity (disc rotational speed varies with the radius) was another issue. Little was known, and every idea had to be tried on the testbed, and this took time. So that, at the May 1980 meeting, the choice of the channel code remained open, and 'more study was needed'. Before continuing with the coding cliffhanger, we take a musical break.

Playing time and Beethoven's Ninth by Wilhelm Furtwängler

Playing time and disc diameter are probably the parameters most visible for consumers. Clearly, these two are related: a 5% increase in disc diameter yields 10% more disc area, and thus an increase in playing time of 10%. The Philips' top made the proposal regarding the 115mm disc diameter. They argued 'The Compact Audio Cassette was a great success', and, 'we don't think CD should be much larger'. The cross diameter of the Compact Audio Cassette, very popular at that time and also developed by Philips, is 115 mm. The Philips prototype audio disc and player were based on this idea, and the Philips team of engineers restated this view in the list of preferred main parameters. Sony, no doubt with portable players in mind, initially preferred a 100 mm disc.

During the May 1980 meeting something very remarkable happened. The minutes of the May 1980 meeting in Eindhoven literally reads:

disc diameter:	120 mm,
playing time:	75 minutes,
track pitch:	1.45 μm,

can be achieved with the Philips M3 channel code. However, the negative points are: large numerical aperture needed which entails smaller (production) margins, and the Philips' M3 code might infringe on Ampex M2.

Both disc diameter and playing time differ significantly from the preferred values listed during the Tokyo meeting in December 1979. So what happened during the six months? The minutes of the meetings do not give any clue as to why the changes to playing time and disc diameter were made. According to the Philips' website with the 'official' history: "The playing time was determined posthumously by Beethoven". The wife of Sony's vice-president, Norio Ohga, decided that she wanted the composer's Ninth Symphony to fit on a CD. It was, Sony's website explains, Mrs. Ohga's favorite piece of music. The Philips' website proceeds:

"The performance by the Berlin Philharmonic, conducted by Herbert von Karajan, lasted for 66 minutes. Just to be quite sure, a check was made with Philips' subsidiary, Polygram, to ascertain what other recordings there were. The longest known performance lasted 74 minutes. This was a mono recording made during the Bayreuther Festspiele in 1951 and conducted by Wilhelm Furtwängler. This therefore became the maximum playing time of a CD. A diameter of 120 mm was required for this playing time".

Everyday practice is less romantic than the pen of a public relations guru. At that time, Philips' subsidiary Polygram –one of the world's largest distributors of music– had set up a CD disc plant in Hanover, Germany. This could produce large quantities CDs with, of course, a diameter of 115mm. Sony did not have such a facility yet. If Sony had agreed on the 115mm disc, Philips would have had a significant competitive edge in the music market. Sony was aware of that, did not like it, and something had to be done. The result is known.

Channel code continued, EFM

Popular literature, as exemplified in Philips' website mentioned above, states that the disc diameter is a direct result of the requested playing time. And that the extra playing time for Furtwängler's Ninth subsequently required the change from 115mm to a 120 mm disc (no one mentions Sony's 100 mm disc diameter). It suggests that there are no other factors affecting playing time. Note that in May 1980, when disc diameter and playing time were agreed, the channel code, a major factor affecting playing time, was not yet settled. In the minutes of the May 1980 meeting, it was remarked that the above (diameter, playing time, and track pitch) could be achieved with Philips' M3 channel code. In the mean time, but not mentioned in the minutes of the May meeting, the author was experimenting with a new channel code, later coined EFM [3]. EFM, a rate 8/17, *d*=2, code made it possible to achieve a 30 percent higher information density than the Philips' M3. EFM also showed a good resilience against disc handling damage such as fingerprints, dust, and scratches. Note that 30 percent efficiency improvement is highly attractive, since, for example, the increase from 115 to 120 mm only offers a mere10 percent increase in playing time.

A month later, in June 1980, we could not choose the channel code, and again more study and experiments were needed. Although experiments had shown the greater information density that could be obtained with EFM, it was at first merely rejected by Sony. At the end of the discussion, which at times was heated, the Sony people were specifically opposing the complexity of the EFM decoder, which then required 256 gates. My remark that the CIRC decoder needed at least half a million gates and that the extra 256 gates for EFM were irrelevant was jeered at. Then suddenly, during the meeting, we received a phone call from the presidents of Sony and Philips, who were meeting in Tokyo. We were running out of time, they said, and one week for an extra, final, meeting in Tokyo was all the lads could get. On June 19, 1980 in Tokyo, Sony agreed to EFM. The 30 percent extra information density offered by EFM could have been used to reduce the diameter to 115mm or even Sony's original target diameter100mm, with, of course, the demanded 74 minutes and 33 seconds for playing Mrs. Ohga's favorite Ninth. However such a change was not considered to be politically feasible, as the powers to be had decided 120mm. The option to increase the playing time to 97 minutes was not even considered. We decided to improve the production margins of player and disc by lowering the information density by 30 percent: the disc diameter remained 120mm, the track pitch was increased from 1.45 to 1.6μm, and the user bit length was increased from 0.5 to 0.6μm. By increasing the bit size in two dimensions, in a similar vein to large letters being easier to read, the disc was easier to read, and could be introduced without too many technical complications.

The maximum playing time of the CD was 74 minutes and 33 seconds, but in practice, however, the maximum playing time was determined by the playing time of the U-Matic video recorder, which was 72 minutes. Therefore, rather sadly, Mrs. Ohga's favorite Ninth by Furtwängler could not be recorded in full on a single CD till 1988 (EMI 7698012), when alternative digital transport media became available. On a slightly different note, Jimi Hendrix's Electric Ladyland featuring a playing time of 75 minutes was originally released as a 2 CD set in the early 1980s, but has been on a single CD since 1997.

The inventor of the CD

The Sony/Philips task force stood on the shoulders of the Philips' engineers who created the laser videodisc technology in the 1970s. Given the videodisc technology, the task force made choices regarding various mechanical parameters such as disc diameter, pit dimensions, and audio parameters such as sampling rate and resolution. In addition, two basic patents were filed related to error correction, CIRC, and channel code, EFM. CIRC, the Reed-

Solomon ECC format, was completely engineered and developed by Sony engineers, and EFM was completely created and developed by the author.

Let us take a look at numbers. The size of the taskforce varied per meeting, and the average number of attendees listed on the official minutes of the joint meetings is twelve. If the persons carrying hierarchical responsibility of the CD project are excluded (many chiefs, hardly any Indians) then we find a very small group of engineers who carried the technical responsibility of the Compact Disc 'Red Book' standard.

Philips' corporate public relations department, see The Inventor of the CD on Philips' website [7], states that the CD was "too complex to be invented by a single individual", and the "Compact Disc was invented collectively by a large group of people working as a team". It persuades us to believe that progress is the product of institutions, not individuals. Evidently, there were battalions of very capable engineers, who further developed and marketed the CD, and success in the market depended on many other innovations. For example, the solid-state physicists, who developed an inexpensive and reliable laser diode, a primary enabling technology, made CD possible in practice. Credit should also be given to the persons who designed the transparent Compact Disc storage case, the ' jewel box', made a clever contribution to the visual appeal of the CD.

Philips and Sony agreed in a memorandum dated June 1980, that their contributions to channel and error correction codes are equal. Sony's website, however, with their 'official' history is entitled 'Our contributions are equal' [8]. The website proceeds, “We avoid such comments as, 'We developed this part and that part' and to emphasize that the disc's development was a joint effort by saying, 'Our contributions are equal'. The leaders of the task force convinced the engineers to put their companies before individual achievements.” The myth building even went so far that the patent applications for both CIRC and EFM were filed with joint Sony/Philips inventors. Philips receives the lion's share of the patent royalty income, which is far from equally shared between the two partners.

Everything else is gaslight

A favorite expression of audiophiles –particularly during the early period, when they were comparing both vinyl LP and CD versions of the same recordings– was: "It is as though a veil has been lifted from the music". Or, in the words of the famous Austrian conductor Herbert von Karajan, when he first heard CD audio: "Everything else is gaslight". Von Karajan was fond of the gaslight metaphor: he first conducted Der Rosenkavalier in 1956 with the soprano Elisabeth Schwarzkopf. Later, when he revived the opera in 1983 with Anna Tomowa, he referred to his 1956 cast as "gaslight", which rather upset Schwarzkopf.

Philips and Sony settled the introduction of the new product to be on November 1, 1982. The moment the ink of the “Red Book”, detailing the CD specifications, was dry, the race started, and hundreds of developers in Japan and the Netherlands were on their way. Early January 1982 it became clear that Philips was running behind, the electronics was seriously delayed, and they asked Sony to postpone the introduction. Sony rejected the delay, but agreed upon a two-step launch. Sony would first market their CD players and discs in Japan, where Philips had no market share, and half a year later, March 1983, the worldwide introduction would take place by Philips and Sony. Philips Polygram could supply discs for the Japanese market. This gave Philips some breathing space for the players, but not enough, as in order to make the new deadline, the first generation of Philips CD players was equipped with Sony electronics.

The first CD players cost over $2000, but just two years later it was possible to buy them for under $350. Five years after the introduction, sales of CD were higher than vinyl LPs. Yet this was no great achievement, as in 1980 sales of vinyl records had been declining for many years although the music industry was all but dead. A few years later, the Compact Disc had completely replaced the vinyl LP and cassette tape. Compact Disc technology was ideal for use as a low-cost, mass-data storage medium, and the CD-ROM and record-once and re-writable media, CD-R and CD-RW, respectively, were developed. Hundreds of millions of players and more than two hundred billion CD audio discs were sold.

Further reading

[1] M.G. Carasso, W.J. Kleuters, and J.J Mons, Method of coding data bits on a recording medium (M3 Code), US Patent 4,410,877, 1983.

[2] T. Doi, T. Itoh, and H. Ogawa, A Long-Play Digital Audio Disk System, AES Preprint 1442, Brussels, Belgium, March 1979.

[3] K.A.S. Immink and H. Ogawa, Method for Encoding Binary Data (EFM), US Patent 4,501,000, 1985.

[4] T. Kretschmer and K. Muehlfeld, Co-opetition in Standard-Setting: The Case of the Compact Disc, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=618484

[5] J.W. Miller, DC Free encoding for data transmission (M2 Code), US Patent 4,234,897, 1980.

[6] K. Odaka, Y. Sako, I. Iwamoto, T. Doi, and L. Vries, Error correctable data transmission method (CIRC), US Patent 4,413,340, 1983.

[7] The inventor of the CD, Philips' historical website: http://www.research.philips.com/newscenter/dossier/optrec/index.html

[8] Our contributions are equal, Sony's historical website: www.sony.net/Fun/SH/1-20/h2.html

[9] L.B. Vries, The Error Control System of Philips Compact Disc, AES Preprint 1548, New York, Nov. 1979.

[10] K.A.S. Immink, "Reed-Solomon Codes and the Compact Disc" in S.B. Wicker and V.K. Bhargava, Eds., Reed-Solomon Codes and Their Applications, IEEE Press, 1994.

Seoul Chapter at a Glance

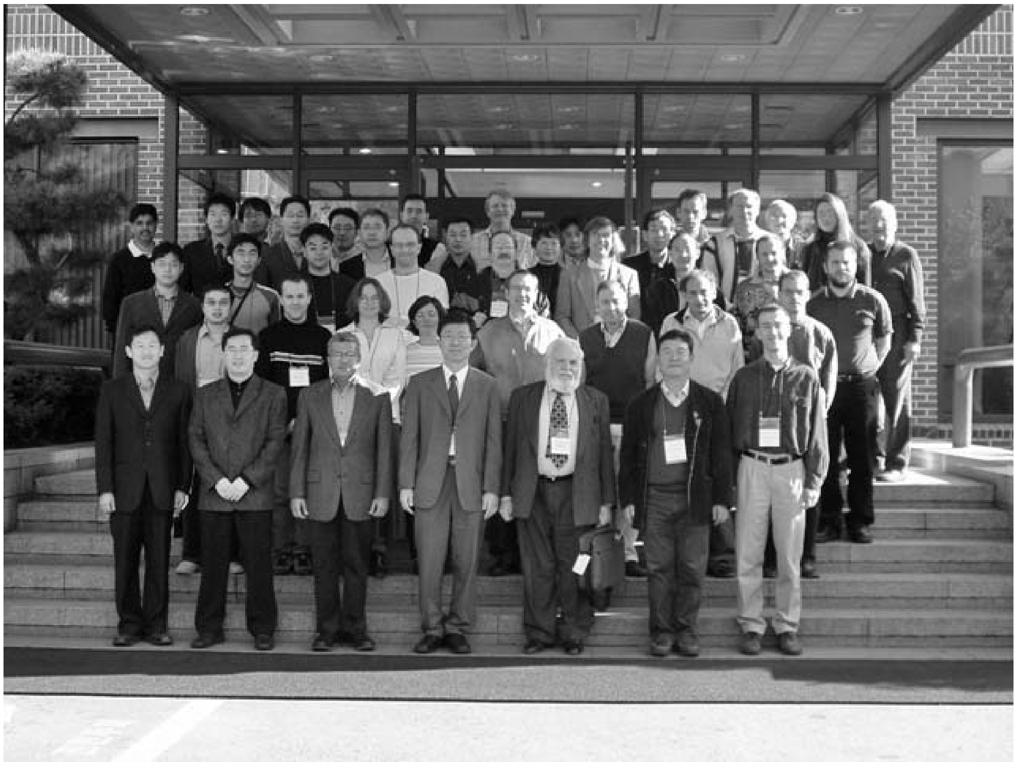
Jong-Seon No

The IEEE Information Theory Chapter of the Seoul Section was founded in 1996 by Jong-Seon No, as a founding chair, with many other leading Korean researchers from academia and industry. Compared to wireless communications and other related areas, Information Theory area was largely unattended in Korea until 1996. The first domestic workshop on Information Theory held in early 1997 attracted about 50 attendees including graduate students and 5 original papers were presented.

Now, the IT Seoul chapter has almost 100 active members including about 60 graduate students. The chapter annually hosts three half-day workshops, each composed of four paper presentations and one tutorial, attended by about 50 members including graduate students. In 2003, the chapter hosted a Distinguished Lecture by Vijay K. Bhargava at Seoul National University, whose title was "Power-Residues, Binary Matrices with Specified Properties and Error Correcting Codes." The chapter hosted the "Sequences and Their Applications (SETA '04)" workshop, Seoul in 2004, where 50 papers were presented and the proceedings were published by Springer as LNCS vol. 3486. Tor Helleseth and Dilip Sarwate served as TPC co-chairs. The chapter also hosted the "International Symposium on Information Theory and Its Applications (ISITA 2006)", Seoul in 2006, where 177 high quality papers from 22 countries were selected and presented. ISIT 2009 will be hosted by the Seoul chapter, where Jong-Seon No and Vincent Poor will serve as General co-chairs, and Robert Calderbank, Habong Chung, and Alon Orlitsky will serve as TPC co-chairs.

The chapter has been constantly making an effort to recruit young graduate students in the area of Information Theory by co-hosting or cooperating with a number of major domestic conferences on wireless communication area.

The chapter is honoured to receive the 2007 IEEE Information Theory Chapter of the Year Award and will keep working on upgrading and expanding its efforts and activities.



Sequences and Their Applications 2004 (SETA '04) Workshop in Seoul National University.

Activities of the IT Student Committee - Summer and Fall 2007

Andrea Goldsmith, Ivana Maric, Lalitha Sankar, Brooke Shrader and Aylin Yener

The Student Committee has been very active over the past several months and we summarize the happenings in this article.

In chronological order, we start with the ISIT 2007 event held in Nice, where following the tradition of the previous year, the Student Committee hosted two events: a research roundtable discussion and a panel discussion. As expected, both events raised a lot of interest – during the roundtable discussion we had nine research topics, and a full room with about 75 students in attendance (and even a few professors). The research topics and their leaders were as follows:

1. “Multiantenna and Multiuser Channels: Recent Results on the Capacity Regions and Degrees of Freedom”, leader: Hanan Weingarten, The Technion

2. “Rate Distortion Theory for Multi-terminal Networks” leader: Haim Permuter, Stanford University

3. “Cognitive Radios and Universal IT”, leader: Krish Eswaran, UC Berkeley

4. “Neuroscience and IT” leader: Bobak Nazer, UC Berkeley

5. “Multiaccess Channels with Feedback”, leader: Michele Wigger, ETHZ

6. “IT Security”, leader: Prasanth Ananthapadmanabhan, UMD/UMCP

7. “Transmission of Correlated Data in Wireless Networks”, leader: Stephan Tinguely, ETHZ

8. “Joint Source-Channel Coding”, leader: Deniz Gunduz, Brooklyn Poly

9. “Information Theoretical Aspects of Decentralized Detection with Applications”, leader: Amichi Sanderovich, The Technion.

More details on this event and papers discussed can be found on the student website: http://students.itsoc.org/.

The panel topic chosen this year was "Research and Funding: Academic and Industry Perspectives”. The panelist were: Jeffrey Andrews, University of Texas, Austin; João Barros , Universidade do Porto, Portugal; Robert Calderbank, Princeton University; Ralf Koetter, Technical University of Munich, Germany and Vincent Poor, Princeton University. The panel was moderated by Sergio Servetto. We would like to thank all the participants who helped make these events a success. As a custom by now, free IT student committee T-shirts were distributed and we thank all the Stanford students who helped in transporting T-shirts from California to Nice.

Round table at ISIT 2007 moderated by Sergio Servetto.

Shortly after, the Committee had to deal with the unbelievably tragic loss of its chair, Professor Sergio Servetto of Cornell University. The tragic event left us shocked and deeply saddened. Professor Servetto had been active in helping the student committee since its inception. Our website was set up by him. We had the privilege of working with him and were inspired by his immense enthusiasm, energetic participation, and an infectious interest in student activities. Together with Professor Aylin Yener and other colleagues, he was planning on organizing a Winter/Summer School for students.

At the Allerton Conference in September, the student committee hosted a tribute in the memory of Sergio Servetto. The challenging task of organizing and leading the tribute was done by Professor Andrea Goldsmith of Stanford along with help from Professor João Barros. The tribute was attended by Sergio’s wife Viviana and was also video-taped for his sons’ benefit. It was a poignant event with friends, colleagues, and students sharing fond memories of Sergio as well as their grief at his loss.

In addition to the tribute, two other events important to students happened at Allerton in the BOG meeting. First a Winter/Summer School of Information Theory proposal by Aylin Yener and Gerhard Kramer, presented by Aylin, was approved, including $10,000 of funding by the society for the first event. DARPA tentatively approved a similar level of funding, so the school is off to a great start in terms of funding and planning. The first school will be held early June next year at an east coast US location, most likely at the University Park Campus of Penn State, with more details to follow over the next few months. In addition, the student committee proposed to the BOG that a student paper award be given at each ISIT. This award was first proposed by the student committee, and awarded for the first time at ISIT’07 in Nice. The success and student enthusiasm for the award prompted the student committee to propose it be given annually, and this was approved by the BOG.

Inspired by Professor Sergio Servetto, The Student Committee is organizing a Panel Discussion at the CISS conference, to be held at Princeton University in March 2008. The topic of the panel is "What Makes a Great Researcher."

For all the events that the committee is organizing, we need more student volunteers. If interested, please send an email to our volunteer coordinator Lalitha lalitha@winlab.rutgers.edu. Some of the specific tasks that we need help with are: keeping the IT student committee website updated; organizing the events at the next CISS and ISIT. All the suggestions about interesting topics for the next panels and roundtable discussions are welcome.

We end this article with happy news announcing that Professor Aylin Yener of Pennsylvania State University is our new student committee chair. We are excited to work with her and already had our first meeting at the Allerton Conference in September. We welcome her to her new position and look forward to working with her!

Web Matters

Ulrich Speidel

As a member of the IEEE IT Society, you will most likely be aware that the society has, for some years now, had a web site (currently under www.itsoc.org). A recent online survey of members showed that most of those surveyed accessed the website around once a month. At present, the website provides information on, among other things, the society's publications, membership and governance, access to Pareja (the society's paper review system), author information, society news, noteworthy papers and presentations, student activities, and of course Claude E. Shannon's life and work.

Over the years, the website has changed hands a few times. It is now in the able care of Nick Laneman, whose job description as the society's "Online Editor" now amounts to a vast understatement of the services he has rendered to the society since he took up office. Among other things, such as providing a web area for BoG documents, Nick pondered where the web site was, what it offered, where it could and should be, and what it might offer.

Fast forward to ISIT 2007 in Nice, and the pondering had resulted in a preliminary proposal to the BoG to undertake a complete re-vamp of the society's online activities. Nick gathered an able team of helpers and has since managed to conduct a user survey and facilitate an extensive online discussion among the ad-hoc committee.

As a first outcome, this has produced a far more comprehensive and detailed proposal for a first stage, complete with cost estimates from developers. So it is time to update the society, and your newsletter editor has seen to it that yours truly got roped into reporting.

So, what's up? First and foremost, the driving idea behind the revamp is to make the site more useful to the society, its members and its officers, reviewers, and other volunteers, such as conference organizers. Central to it will be the move to a customized content management system (CMS), which should allow updates to the content (such as changes to the details of society officers) to be performed by a larger group of people, without the need for in-depth knowledge of web programming languages such as HTML.

There will also be an extension of the static content, such as a more systematic collection of technical material ranging from digital archives of Shannon and other plenary lectures and tutorials to classical out-of-print books and articles. Most of this is easy to do (provided copyright can be obtained), and in fact some of it is already part of the society's current web arrangements (see media.itsoc.org, for example).

On the more interactive front, improved web areas for BoG and the Awards Committee have been high on the wish list, as have been pages for job opportunities, for volunteers, and for the conference proposal, approval, and announcement process. Again, these seem to be within relatively easy reach.

Somewhat further down the track there are other features, which are a little more difficult to achieve than by just customizing a content management system. Significant input and thinking are still needed here.

These include features such as:

- an integrated paper review process with perhaps a uniform author and reviewer interface for the Transactions, conferences, and other refereed publication avenues.
- support for conference and workshop home pages with conference management facilities
- a preprint facility similar to ArXiv.

Some of these have pretty apparent benefits. In other cases, a closer look identifies need. Take conference support, for example: Almost the entire workflow of a conference or workshop can be mapped into WWW code - from paper submission to participant registration. It is not easy, though, and not usually worth doing for a single event. However, the society runs an ISIT and typically at least one IT workshop per year, oftentimes two, all of which need websites.

Offering conference information, online paper submission, if possible online refereeing, and online conference registration is now pretty much a standard feature of many such web sites. At present, many IT society events pay thousands of dollars each for a portal to offer conference information. For online submission and reviewing, EDAS charges $6 for the first 100 submissions, $5 thereafter. Finally, online registration costs a bundle or is factored somewhere into the fees charged by conference managers, which can be exorbitant.

Investing in a re-usable system which IT Society conferences can access free of charge could thus pay dividends in terms of lower conference fees for all of us, and save organizers time reinventing the wheel.

Anand Sarwate, Matthieu Bloch, and Nick Laneman have put an overview document together that outlines many of the issues involved in the transition to a content management system and the possible services it may provide for us (see http://dev.itsoc.org/bog/online-committee/ for this and other related information). The society is in the fortunate position to be able to implement at least some of these ambitions. It will require careful planning and consultation, and this means input from a large range of members (including you).

Quotes from developers have been sourced (also on http://dev.itsoc.org/bog/online-committee/). The BoG has unanimously approved up to $60k for work on the transition to a customized CMS.

As work gets underway, now is the time to have your say and champion your cause. In fact, many of the decisions that await the society in web matters are actually non-technical, so you do not have to be a web guru to contribute. David Neuhoff has been doing the “non-hacker type” reality checks so far, and having a few more voices like his in the design process would be helpful. So, hacker or not, join the web-discuss mailing list and make your views count.

Workshop Report: The 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks, July 1–6, 2007 Bergen, Norway

Øyvind Ytrehus

The week following ISIT 2007 in Nice, 86 participants were gathered at Solstrand Fjordhotel, just outside of Bergen, Norway, for the first 2007 IEEE Information Theory Workshop. The workshop focused on topics related to information theory for wireless networks, such as: Space-time codes and cooperative relay networks, Error-control for wireless communication: Forward error correction or ARQ?, Network coding for wireless communication, and Low correlation sequences and streamciphers. The single-track program consisted of 18 invited and 32 accepted talks. The corresponding papers can be found on IEEXplore. Vijay Kumar and Tor Helleseth were the program co-chairs, assisted on the program committee by the invited speakers. Yours truly kept an eye that things were running more or less smoothly, while Ángela Barbero was responsible for the artwork. Abundant pictorial documentation of the ITW can be found at http://itw2007.org (for a limited time) or at http://www.selmer.uib.no/ITW2007.html (for a long time).

Art work: Claude Shannon.

Workshop Report: The 2007 IEEE Information Theory Workshop, September 2-6 2006 in Lake Tahoe, CA USA

by William E. Ryan

The 2007 Information Theory Workshop at Lake Tahoe was held at the Granlibakken Conference Center and Lodge, Sept. 2-6, 2007. It was sponsored by the IEEE, the Information Theory Society, and Qualcomm, and was organized by Bill Ryan, Krishna Narayanan, Zixiang Xiong, Amir Banihashemi, and Tolga Duman together with a large technical program committee. The workshop theme was coding and included sessions on source coding, channel coding, joint source-channel coding, network coding, distributed source coding, iterative decoders, and coding for wireless channels.

The program consisted of four plenary talks, 33 invited talks, 50 contributed talks, and 32 contributed poster presentations. The contributed papers were selected from 116 submitted papers. Both invited and contributed papers will appear on IEEExplore. The plenary talks were: "On Modeling Distributions over Large Alphabets" by Alon Orlitsky; "Source and Channel Coding: Theory, Practice, and System Aspects" by Giuseppe Caire; "Pseudo-Codewords and Iterative Decoding: A Guided Tour" by Pascal Vontobel; and "Applications of Compressive Sampling to Error Correction" by Emmanuel Candes. In the spirit of a workshop, as opposed to a conference, the last two plenary talks were of a tutorial nature and were given in two 50-minute presentations.

Granlibakken, set on a hillside among beautiful trees, including a few sequoias, turned out to be an ideal place for the workshop. The food was great, the weather was ideal, and it was only about a 30-minute walk to the enormous and beautiful Lake Tahoe. Many beautiful birds were seen, including blue (Steller's) jays, cardinals, and ravens, and at least one bear was seen (running away). One highlight of the workshop was engineer-turned-comedian Don McMillan who performed what he calls "application-specific comedy" on banquet night. As an example, his idea of a Gaussian channel is "Cable Channel 134: All Gauss, all the time".

From the many compliments received by the organizers, ITW2007 at Lake Tahoe was a resounding success. For more information on the workshop, including the slides of the plenary talks, the program, and the organizing committee, visit http://www.ece.tamu.edu/itw2007/.

Alon Orlitsky giving his Plenary Talk on Sept. 3.

Giuseppe Caire giving his Plenary Talk on Sept. 4.

Pascal Vontobel giving his Plenary Talk on Sept. 5.

Emmanuel Candes giving his Plenary Talk on Sept. 6.

A collection of some of the participants of ITW2007 on the Granlibakken grounds.

Workshop report: The XI international symposium on problems of redundancy in information and control systems, July 2-6 2007, Saint-Petersburg, Russia

Sergei Fedorenko

Plenary lecture given by Hisashi Kobayashi.

Workshop participants.

The XI international symposium on problems of redundancy in information and control systems was held on 2-6 July 2007 on board of the comfortable cruise ship "Saint-Petersburg" (Saint-Petersburg, Russia).

The history of the symposium on redundancy problems in information systems is long-standing and dates from 1964 when the first symposium was held, initiated by Professor N.A. Zheleznov. The topics discussed on this symposium covered the widest area of problems connected with analysis and synthesis of the systems with structural, logical or functional redundancy. Thousands of professionals in the field of information theory, error-correcting coding, communication systems and networks, automated control systems, computer engineering and mathematical modeling participated in 10 symposiums held within 1964-1989. The symposium was considered as a complex and interbranch one, aimed at mutual ideas sharing in the different directions of informatics.

Among the conventional symposium sections outstood the section of the information theory and coding for the high quality of reports presented there. The majority of well-known Russian specialists and scientists took part in the work of this section from year to year. Alongside, starting from the first symposium, the section of computer engineering was traditionally held.

The X symposium on redundancy problems in information systems was held in 1989. Since then it has been interrupted by the well-known events in Russian history. Saint-Petersburg State University of Aerospace Instrumentation (SUAI) with the U.S.S.R. Academy of Sciences was the organizer of all the 10 symposiums on the redundancy problem. Now SUAI proposes to regenerate the traditional symposium taking into account changed scientific, technical and political situation.

The technical co-sponsors of the symposium are the IEEE Russia (Northwest) Section and the IEEE Information Theory Society.

About 120 participants from 8 countries took part in the work of the newly arranged Symposium of 2007, 55 reports were presented that cover the following scientific areas: Information theory, Coding theory, Communication systems, Cryptography, Theory of combinations, Software systems, Computational systems and networks. The Coding theory session traditionally attracted the majority of reports (15).

Two plenary sessions with invited lecturers were held. They are “35 Years of Progress in Digital Magnetic Recording” by Hisashi Kobayashi, Princeton University, USA, and “35 years Development of Multithreshold Algorithms” by Valery Zolotarev, Space Research Institute of Russian Academy of Science, Moscow, Russia.

In addition, a special session was held dedicated to the 70th anniversary of Victor Zyablov (Institute for Information Transmission Problems of Russian Academy of Science, Moscow, Russia).

Hisashi Kobayashi (Princeton University, USA) was awarded the first prize as the best reporter, and Thomas Berson (Anagram Labs, Palo Alto CA, USA) was acknowledged as the best session chairman.

The Proceedings of the Symposium are available at http://k36.org/redundancy2007/proceedings.php

Although the scientific discussion has been the main part of the Symposium program, it has not been the only. The participants had also a good chance to visit Valaam, the largest and finest island of the lake Ladoga. The island hosts a functioning friary with its chronicle dating back to the X century, Spaso-Preobrazhensky Cathedral, the Saint Peter and Paul Church and several skits that preserve spiritual values and are magnificent monuments of Russian architecture. Another place of interest that participants could enjoy was the world famous Kizhi Museum, one of the largest out-door museums in Russia. The museum collections contain 83 pieces of the wooden architecture. The core of the collection is the outstanding sample of the wooden architecture – the architectural ensemble of the Kizhi Pogost of Our Savior built on Kizhi Island in the 18th – 19th centuries.

The forthcoming XII Symposium will be held in Saint-Petersburg in 2009.

Call for Nominations

2008 IEEE Information Theory Society Paper Award

The Information Theory Society Paper Award is given annually for an outstanding publication in the fields of interest to the Society appearing anywhere during the preceding two calendar years.

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 a certificate and an honorarium of US $1,000 for a paper with a single author, or US $2,000 equally split among multiple authors. The 2008 award will be given for a paper published in 2006 and 2007.

NOMINATION PROCEDURE: By March 1, 2008, please email the name of the paper you wish to nominate, along with a supporting statement explaining its contributions, to the IT Transactions Editor-in-Chief, Ezio Biglieri, at ezio.biglieri@upf.edu.

2008 IEEE Joint Comsoc/IT Paper Award

The Joint Information Theory/Communications Society Paper Award recognizes one or two outstanding papers that address both communications and information theory. Any paper appearing in a ComSoc or IT Society publication during the year 2007 is eligible for the 2008 award. A Joint Award Committee will make the selection.

NOMINATION PROCEDURE: By February 1, 2008, please email the name of the paper you wish to nominate, along with a supporting statement explaining its contributions, to Andrea Goldsmith at andrea@ee.stanford.edu.

2008 IEEE Information Theory Society Aaron D. Wyner Distinguished Service Award

The IT Society Aaron D. Wyner Award honors individuals who have shown outstanding leadership in, and provided long standing exceptional service to, the Information Theory community. This award was formerly known as the IT Society Distinguished Service Award.

Nominations for the Award can be submitted by anyone. The individual or individuals making the nomination have the primary responsibility for justifying why the nominee should receive this award.

NOMINATION PROCEDURE: Letters of nomination should

- Identify the nominee's areas of leadership and exceptional service, detailing the activities for which the nominee is believed to deserve this award;
- Include the nominee's current vita;
- Include two letters of endorsement.

Current officers and members of the IT Society Board of Governors are ineligible.

Please send all nominations by April 15, 2008 to Bixio Rimoldi, at bixio.rimoldi@epfl.ch.

IEEE Awards

The IEEE Awards program has paid tribute to technical professionals whose exceptional achievements and outstanding contributions have made a lasting impact on technology, society and the engineering profession.

Institute Awards presented by the IEEE Board of Directors fall into several categories:

Medal of Honor	(Deadline: July 1)
Medals	(Deadline: July 1)
Technical Field Awards	(Deadline: January 31)
Corporate Recognitions	(Deadline: July 1)
Service Awards	(Deadline: July 1)
Prize Papers	(Deadline: July 1)
Fellowship	(Deadline: November 15)

The Awards program honors achievements in education, industry, research and service. Each award has a unique mission and criteria, and offers the opportunity to honor distinguished colleagues, inspiring teachers and corporate leaders. The annual IEEE Awards Booklet, distributed at the Honors Ceremony, highlights the accomplishments of each year's IEEE Award and Medal recipients.

For more detailed information on the Awards program, and for nomination procedure, please refer to http://www.ieee.org/portal/pages/about/awards/index.html.

IEEE Information Theory Society Board of Governors Meeting Nice, France, June 24, 2007

João Barros

Attendees: John Anderson, João Barros, Ezio Biglieri, Giuseppe Caire, Daniel Costello, Richard Cox, Anthony Ephremides, Dave Forney, Marc Fossorier, Andrea Goldsmith, Alex Grant, Tor Helleseth, Ralf Koetter, Ryuji Kohno, Frank Kschischang, J. Nicholas Laneman, Hans-Andrea Loeliger, Steven W. McLaughlin, Muriel Médard, Urbashi Mitra, Prakash Narayan, David L. Neuhoff, Vincent Poor, Bixio Rimoldi, Nela Rybowicz, Anant Sahai, Sergio D. Servetto, Shlomo Shamai, Ulrich Speidel, Joseph A. O'Sullivan, Daniela Tuninetti, David Tse, Adriaan J. van Wijngaarden, Ken Zeger.

The meeting was called to order at 13:00 by Society President Bixio Rimoldi, who welcomed the members of the Board.

1. The agenda was approved with a change of order. The change was due to the fact that Nicholas Laneman, who was expected to present a proposal for a new initiative, had to leave earlier than planned. Board members were encouraged to look at the agenda and other materials online at the Society development server.

2. The Board unanimously approved the minutes of the previous meeting (Baltimore, MD, USA, March 14, 2007).

3. Anant Sahai presented the Treasurer's report. The Society's finances were reviewed and found to be in good health. The Treasurer explained how, according to the new financial rules of IEEE, half of the operative surplus of the previous year can be spent in an off-budget way until the end of the year. This results in an amount of 208k that can be viewed as a "one-time" spending opportunity. The Treasurer also presented foreseeable risks in the future and expressed some concern regarding a possible reduction of the Society's IEEExplore revenues. Some caution was advised regarding the 10% surplus that each conference is expected to deliver to the Society. The Treasurer believes that although IEEE recommends a 20% surplus, the current value of 10% is reasonable in the short term to ensure the health of the Society's finances.

4. The Board agreed on (a) starting with the presentation of proposals for new initiatives and (b) making the corresponding motions after the last proposal was presented.

5. The Online Editor, J. Nicholas Laneman, presented the recommendations of the adhoc committee on online content and services with respect to how to invest part of the surplus in order to expand the Society's web infrastructure. The Online Editor underlined the necessity of having typical users testing the system under development to ensure that the interfaces are sufficiently pleasant to use. The recommendations elaborate on several possibilities of improving the use of repositories, providing support for publications (conferences, journals and preprints) and making lectures widely available on the web. A general strategy for improving the Society's use of the web was put to discussion. The Board discussed whether it would make sense for the Society to take a leadership role within IEEE with respect to using the web for member interactions. Rich Cox explained that IEEE is planning to put preprints available online, once they are accepted for publication. He also commented that the current recommendation of the ad-hoc committee goes well beyond IEEExplore in using the web to promote interaction between IEEE members. IEEE estimates that only 20% of its members interact at conferences and meetings. Andrea Goldsmith expressed some concern regarding forcing a cultural change within the Society and advised some caution regarding the investment and the size of the envisioned project. Giuseppe Caire explained the difficulties that conference organizers have in finding the right infrastructure for submission and registration support. He recommended that the Society's web-site should provide more help in this direction. Ulrich Speidel pointed out that the structure of the Society's conferences does not change substantially from year to year, so that the web infrastructure could be re-used easily. Anant Sahai elaborated on the importance of the web for the younger generation and how it can be a tool to attract undergraduate students to our field. Ralf Koetter suggested integrating existing software as opposed to designing a whole system from scratch. Sergio Servetto stressed the need for professional assistance in drawing adequate specifications. A small amount would be necessary to take such initial steps.

6. Sergio Servetto presented a proposal of the Student Committee. The main idea is to start an annual event similar to the Winter School on Information Theory in Europe with the goal of creating opportunities for students to meet each other informally. The Board discussed means to fund this kind of initiatives, including the use of surplus funds from conferences. Frank Kschischang suggested that the School be co-located with ISIT to save travel funds.

7. Ulrich Speidel presented a proposal of a fund of 10.000 USD to support graduate students from New Zealand in attending ITSOC-related conferences. He explained that since most conferences take place in Europe, North America and Asia, sending New Zealand students involves a large financial burden. The Board discussed the possibility of using part of the conference surpluses to support local activities. Andrea Goldsmith expressed the concern that the procedure to grant such funds should be fair with respect to other regions. Muriel Médard and Daniel J. Costello proposed that these grants be discussed by the Board on a case by case basis.

8. The President discussed the possibility of digitizing past proceedings of conferences, specifically CISS and Allerton. It is not yet clear whether the organizers of Allerton would be interested in this effort. The Board discussed the issues of copyright.

9. Steve McLaughlin presented a motion to allocate 10.000 USD for initial specifications regarding substantial (order of 100.000 USD) and modular improvements to the Society's web infrastructure. The proposal shall be presented at the next meeting of the Board at Allerton.

   The Board unanimously approved the motion.

10. A motion was presented by Prakash Narayan to allocate 10.000 USD to help fund a Winter or Summer School on Information Theory and Coding pending a specific proposal of the Student Committee (Sergio Servetto). Sergio Servetto clarified that the aforementioned amount would be used as seed money to start preparing the event. The Board discussed the possibility of discussing a more specific proposal at the Allerton meeting. The President suggested creating also a distinguished speaker program, as recommended by the IEEE. The Board unanimously approved a motion of Urbashi Mitra to table the previous motion until the Allerton meeting. A motion was presented by Ralf Koetter for the Society to charge the Student Committee with the organization of a regular International Winter or Summer School on Information Theory and Coding.

    The Board unanimously approved the motion.

    Giuseppe Caire suggested that the European Winter School be integrated in the activities of the Student Committee.

11. Ulrich Speidel amended his proposal to include not only IT-SOC conferences but any IT-SOC sponsored event.

12. A motion was presented to approve the amended proposal of Ulrich Speidel. Andrea Goldsmith presented a friendly amendment to provide funding exclusively for IT Society members.

    The Board unanimously approved the motion.

    The Conference Committee will inform the Organizers of ITW Uruguay about the possibility of submitting a similar proposal.

13. The President presented his report and miscellaneous announcements. The President reported on the state of membership in the Society in comparison to other IEEE Societies. The number of student members continues to decrease. The President reported that membership was still one of the main topics of the June TAB meeting. The Computer Society Chapter of the Santa Clara Valley Section runs a lecture series called Shannon Lectures. Contacts have been made to avoid name duplication with the Society's Shannon Lecture. The President thanked all the members who collaborated on the Publications Report and Society Report, which were used for the Quinquennial Review by IEEE. The President mentioned several suggestions made by IEEE, which are included in his report available online.

14. Vincent Poor presented the report of the Editor in Chief (EIC), including mail dates, page budgets and annual volume. The Five-Year Transactions Review emphasized that the Transactions are an exemplary periodical and mentioned that the time from submission to publication has been reduced. Muriel Médard made the suggestion of promoting a special issue of the Transactions jointly with a journal in biology to explore intersections between the two fields.

15. Ezio Biglieri presented five initiatives to be pursued during his tenure: (1) allowing comments made by readers of arxiv preprints to be taken into consideration in the paper review process of the Transactions, (2) reducing processing time through frequent automatic reminders, (3) re-evaluating the submission of correspondence items and content overlap with conference papers, (4) organizing a series of invited papers, and (5) starting a magazine with news and light tutorial papers. The Board discussed the page limit for transactions letters. The President and Anthony Ephremides suggested abolishing the distinction between Transactions papers and letters. Urbashi Mitra highlighted the implications of the differences between the quality of reviews in conferences and the Transactions. Giuseppe Caire suggested a single class of papers with no page limit and a flexible policy of re-submission of conference papers. Vincent Poor stressed that Transactions papers come from different sources, including computer science conferences where the reviewing process is, in some cases, very strict.

    The Board approved a motion by David Neuhoff for the appointment of an ad-hoc committee to re-evaluate the editorial guidelines and the creation of a magazine.

16. Daniela Tuninetti presented a proposal for an online version of the IT Newsletter.

    The Board approved a motion to increase the budget allocated to the IT Newsletter by 525 USD per issue to enable the trial of an HTML version.

17. Alex Grant presented the Conference Coordinator's report on the current status of symposia and workshops. The individual reports for each event are available online.

    It seems necessary to define whether the awards luncheon at ISIT is an ITSOC event (Society funding) or an ISIT event (conference funding). This issue will be discussed further at the next meeting in Allerton. The Board discussed the possibility of publishing ISIT tutorials through NOW publishers. Andrea Goldsmith presented a motion to withdraw from IEEExplore those ISIT papers, whose authors did not go to ISIT to present their papers, subject to the discretion of the TPC Chairs. The Board unanimously approved the motion. The Treasurer expressed some concerns with respect to the budget of ISIT 2011 in St. Petersburg. The Board discussed the necessity of demanding cautious budgets from conference organizers, in particular with respect to number of attendees and anticipated surplus. Dan Costello presented a motion to hold ISIT 2011 in St. Petersburg with a modified budget to account for a projected surplus of 10%. The budget must be approved by the Treasurer.

    The Board unanimously approved the motion.

    The Board discussed tentative dates for ISIT 2010 in the case that it is held in Austin, Texas.

18. Andrea Goldsmith presented a motion to approve formation of an ad-hoc committee on new initiatives. The goal is to provide a dedicated mechanism for soliciting, developing and evaluating new initiatives.

    The Board unanimously approved the motion.

    The committee will be formed immediately, with its first report given at Allerton.

19. Marc Fossorier presented the Awards Committee report. The Awards Committee recommends that the Best Paper Award be given to the following paper: H. Weingarten, Yossef Steinberg, S. S. Shamai: The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel. IEEE Transactions on Information Theory, vol. 52, No. 9, pp. 3936-3964, September 2006.

    The Board unanimously accepts the recommendation of the Awards Committee. Marc Fossorier raised the concern that the number of nominations for the Joint Com-Soc/IT Paper Award is low. The Board discussed possible measures to increase the number of nominations, including seeking stronger involvement by the Associated Editors of the Transactions.

20. Andrea Goldsmith explained the procedure to select the candidates for the Chapter of the Year Award. The winner this year is the Seoul Chapter.

21. Steve McLaughlin reported on the process of forming the list of BoG candidates.

    The Board unanimously accepts the recommendation of the Nominations Committee.

    The decision to extend the list is left at the discretion of the Nominations Committee.

22. David Neuhoff opened the discussion for nominations of new officers.

    The Board unanimously approved the nominations of Frank Kschischang and Robert Calderbank for 2nd Vice-President.

23. There was no new business.

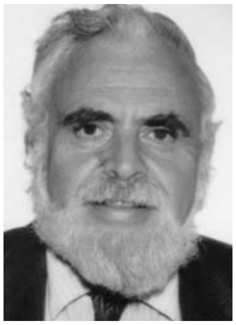
24. The next Board meeting will be held at Allerton in September.

25. The meeting was adjourned at 18:58.

GOLOMB'S PUZZLE COLUMN™

EASY PROBABILITIES SOLUTIONS

Solomon W. Golomb



1. There are $\binom{10}{5} = 252$ ways to select 5 of the 10 decimal digits. When these are arranged in ascending order as a $< b < c < d < e$, the only way $a + b + c > d + e$ can occur is when the five selected numbers are {5, 6, 7, 8, 9}, so that 5+6+7 > 8+9. Thus, the probability of this occurring "at random" is only $\frac{1}{252} = 0.00396825 \ldots$.

2. To maximize the probability that a green marble will be selected, place a single green marble in one jar, with the remaining n - 1 green marbles and all n red marbles in the second jar. If the contestant chooses the first jar (with probability $\frac{1}{2}$), the selected marble will be green. If the second jar is chosen, the probability of a green marble being drawn is $\frac{(n-1)}{(2n-1)}$. Thus the probability of a green marble being selected, for this arrangement, is $\frac{1}{2}(1 + \frac{n-1}{2n-1})$, which is $\frac{2}{3}$ if n has the minimum value $n = 2$, but tends to the limiting value of $\frac{3}{4}$ as n increases.

3. If North-South have all the hearts then East-West have none. Thus the probability that you and your partner (together) have all the hearts is the same as the probability that the two of you have none.

4. If A is a stronger player than B, your probability of winning two consecutive matches is better in the sequence ABA than in the sequence BAB. You can calculate this exactly if P_1 is the probability that A will beat you and P_2 is the probability that B will beat you, with $P_1 > P_2$. Intuitively, the result follows from: to win two matches (of the three) in a row, you *must* win the middle match, and B is weaker than A. Also, the sequence BAB gives you only *one* chance to defeat strong player A, while ABA gives you two chances.

5. (a) If *at least one* of the two children is a girl, there are three equally likely cases (in "birth order"): BG, GB, GG, and the probability that *both* are girls is $\frac{1}{3}$.

   (b) If *the older child* is a girl, the younger child is equally likely to be a boy or a girl; so in this case the probability that *both* are girls is $\frac{1}{2}$.

6. The marble originally in the jar is either black (B_1) or white (W_1). A new white marble (W_2) is inserted. The jar now contains either B_1W_2 or W_1W_2 (equally likely). When a marble is removed and observed to be white, there are now *three* equally likely cases: i) W_2 out, B_1 still in, ii) W_2 out, W_1 still in ; or iii) W_1 out, W_2 still in. (Observing the withdrawn marble to be white eliminated the case iv) B_1 out, W_2 still in.) Thus the probability that the unseen marble is white is $\frac{2}{3}$.

Reference. Problem 1 is my own, and previously unpublished. Some version of each of the remaining problems (all oldies) can be found in Martin Gardner's *Colossal Book of Short Puzzles and Problems*, W. Norton & Co. 2006, which identifies Problem 6 as coming from Lewis Carroll's book *Pillow Problems*.

Guest Column: News from the Communications Program at the National Science Foundation

Sirin Tekinay

Dear reader,

This is the ninth quarterly guest column in this series, signaling the start of my third year at the NSF. I am thrilled to see this space continue to serve its purpose of energizing our interaction on ideas, visions, and issues that impact us all as professionals in the communications community as I write about relevant NSF programs and news.

New and Upcoming Solicitations

In the last issue, I had reported on leading the NSF-wide Cyber Enabled Discovery and Innovation Implementation Team (CDIIT), a big team of program officers from every research directorate and programmatic office of the NSF. I am delighted to report that since then, the team has produced the CDI solicitation, which was posted on September 28 [1]. The five year, $750M program includes research in information theory, communication theory, network theory, signal processing, and complexity, to name a few. For FY 2008 NSF has allocated a minimum of $26M, pending the availability of funds, for CDI. The CDIIT has morphed into the CDI Working Group (CDIWG) and is now ready for the unique challenges of the multi-disciplinary review process.

CDI in a nutshell is computational thinking for science and engineering. Computational thinking refers to: creating and creatively using computational concepts, methods, models, algorithms, and tools. CDI accelerates the huge paradigm shift in the way in which science and engineering will be conducted in the future. It offers a means for tomorrow's scientists to be trained in and skilled with using computational methods and tools. CDI projects are to advance more than one field of science or engineering. For years, our community has been reaching beyond information and communication science and engineering, forming intellectual partnerships, and establishing the role of information theory, communication theory, signal processing, and network theory in other fields of science and engineering. We now have an exceptional opportunity to contribute in ways that are truly transformative for all of science and engineering research and education.

Letters of Intent (LoI) by project team leaders are required and allowed until November 30, 2007. The accuracy of information in the LoI is extremely important in that it will be the basis for NSF's preparation for the review of the corresponding proposal. Proposals will be reviewed in a two-tier process: preliminary proposals are due by January 8, 2008. Based on the expert panel evaluation, successful project teams will be invited to submit full proposals by April 29, 2008. The first cycle of CDI will conclude with awards made by the end of July 2008.

I would like to encourage researchers and educators in our community to work with other scientists and engineers, and by putting your collective expertise together, advance all frontiers of science to new ones. Send us your creative, bold, and ambitious ideas!

News on Communications Research

Theoretical Foundations 2007 competition was concluded promptly by the budget spend-out deadline of mid-August. As a result, communications and network theory related awardees received a combined total of $10M, all of which was spent in the form of standard grants. A standard grant is one where the total funding for a multi-year project is allocated all at once, as opposed to continuing grant increments. I am proud to have been able to sustain a twenty per cent success rate in our program without mortgaging the future of the program by committing continuing grant increments from future years' unknown funds. It was a hard decision to decline the last few proposals; however, despite the bitter medicine we took to ensure the future financial health of the program, I was still able to translate the slight increase in the program budget into a modest increase in the success rate. At the time of writing, the National Science Foundation is under continuing resolution; i.e., the 2008 budget has not been allocated by the congress, which proves the standard grants of 2007 a wise choice.

We, the Theoretical Foundation Cluster, will run the TF08 competition with a calendar similar to last year's. We have updated and uploaded the Theoretical Foundations solicitation in NSF's e-publishing system. We expect it to sail smoothly through the clearance process and make its public appearance by the end of October. This means the deadline will be towards the end of January 2008.

I plan to make funding decisions on the twenty CAREER proposals in the Communications Research Program by mid-November. This is one of the most exciting parts of the job, to have a launching, or fueling, impact on junior colleagues' professional lives.

NSF People

In every column, I introduce some of the people I work with; who embody the culture and spirit of NSF. This time, I am proud to introduce my two co-chairs for the CDI Working Group.

Dr. Thomas Russell of the Mathematical and Physical Sciences Directorate has been a program director at the NSF since 2003 [2]. He leads the Applied and Computational Mathematics Program. Tom is a permanent NSFer. Before NSF, he was a professor at the Department of Mathematics of the University of Colorado at Denver, where he served as the director of the Center for Computational Mathematics, and the chair of his department. Tom's impeccable attention to detail, unfaltering stand for highest quality in everything he produces, added to his fantastic wit make him an irreplaceable teammate.

Dr. Eduardo Misawa is a program director [3] of the Dynamic Systems Program in the Division of Civil, Mechanical and Manufacturing Innovation in the Engineering Directorate. Eduardo joined the NSF from Oklahoma State University, where he was a professor of Mechanical Engineering. Eduardo's tireless dedication to excellence while ensuring the most harmonious work environment with his signature sense of humor and smile make him exceptional among the exceptional.

It has been a privilege and a pleasure to work with such wonderful colleagues. After three months of spending long days together, Tom, Eduardo, and I became close friends: now we are completing each other's sentences, laughing at each other's jokes without telling any.

The Social Scene

The CDI Working Group celebrated the posting of the CDI solicitation, a major milestone, by gathering for happy hour at a local favorite place. It was so much fun that we decided to get together more often, about once a month. The team is such a positive force that it is a pleasure to spend time together, whether we are working or socializing.

On a Personal Note

As exciting, fulfilling, and rewarding as it is, starting and leading the next best NSF initiative took its toll on my personal research. My doctoral student back at NJIT has been infinitely patient with her advisor's absence. Now that my calendar is peppered with CDI steps, I am able to resume my monthly trips to NJ again: the CDIWG has produced, simultaneously with the program solicitation, the program management plan detailing the schedule of the review, recommendation, and award processes for the next year.

The speedy Acela train is pulling into the beautiful Washington Union Station.

Till next time, dream big, and keep in touch!

Sirin Tekinay
Program Director, Communications Foundations
National Science Foundation
4201 Wilson Blvd
Arlington VA 22230

REFERENCES:

[1] http://www.nsf.gov/pubs/2007/nsf07603/nsf07603.htm

[2] http://www.nsf.gov/staff/staff_bio.jsp?lan=trussell&org=NSF

[3] http://www.nsf.gov/staff/staff_bio.jsp?lan=emisawa&org=NSF

Call For Problems

Masimo Franceschetti, Tara Javidi, Paul Siegel, General Chairs, Third ITA Workshop;
Yoav Freund, Larry Milstein, Bhaksar Rao, Technical Chairs, Third ITA Workshop

Do you have a favorite open problem? A problem that you believe is the key to further advance in your field or research? Or, perhaps, a conjecture that, if proved, would make your results so much stronger? Well, now you have a chance to tell the world about it.

The Information Theory and Applications Center at the University of California San Diego is planning to have a session on "Open Problems in Information Theory and Its Applications" as part of the Third ITA Workshop, to be held January 28 through February 1, 2008 on the UCSD campus. The organizers of the Workshop are hereby soliciting submissions of open problems to be presented at this session.

Each submission should adhere to the following content and format guidelines. The content of the submission should include the following components.

Description: Please describe the problem clearly and completely, using succinct mathematical notation. Do *not* assume any prior background --- the description should be accessible to a broad audience of researchers in information theory and its applications.

Motivation: Please tell us why the problem is important. Describe the history of the problem. In which general context does it arise? What would be the ramifications of a solution to this problem?

Prior work: Please provide a brief summary of all relevant prior work, citing the appropriate references. If you have insights/intuition/ideas as to how the problem could be solved, please be generous and describe these as well.

Format: All submissions should consist of a brief cover letter and a PDF file, sent by e-mail to <open.problems@ita.ucsd.edu>. The PDF file should be formatted either as an article of at most 2-3 pages, or as a potential workshop presentation, not to exceed 10-15 slides.

All submissions received by **December 16** will be considered. Notifications of acceptance will be sent by e-mail not later than January 7. Authors of accepted problems will be expected to present them at the Workshop, and to contribute an article (one to three pages) to be posted on the Workshop website.

Note that this is the only open call for submissions for this ITA Workshop --- all other presentations at the Workshop are by invitation only. Also note that ITA may choose to assign monetary awards for the solution of some of the open problems that are accepted for presentation.

2008 IEEE International Symposium on Information Theory

Toronto, Canada, July 6–11, 2008

The 2008 IEEE International Symposium on Information Theory (ISIT 2008) will be held from Sunday, July 6th, to Friday, July 11th, 2008, at the Sheraton Centre Toronto Hotel, in Toronto, Ontario, Canada. Toronto is Canada's largest city, and is directly accessible by air from major cities around the world. The symposium hotel is in the city centre, conveniently located near shopping, museums, and public transportation.

Previously unpublished contributions from a broad range of topics in information theory are solicited, including (but not limited to) the following areas:

Coding theory and practice
Communication theory
Compression
Cryptography and data security
Detection and estimation
Information theory and statistics
Information theory in networks
Multi-terminal information theory
Pattern recognition and learning
Quantum information theory
Sequences and complexity
Shannon theory
Signal processing
Source coding

In addition to submitting new results in the above areas, researchers in related fields and researchers working on novel applications of information theory are encouraged to submit contributions. The paper submission deadline is **January 7, 2008**, with notification of acceptance by March 31, 2008.

Detailed information on paper submission, technical program, accommodation, tutorials, travel, and excursions will be posted on the symposium web site: `http://www.isit2008.org`.

For general inquiries, please contact one of the symposium co-chairmen:

Frank R. Kschischang
Dept. of Electrical and Computer Engineering
University of Toronto
10 King's College Road
Toronto, Ontario M5S 3G4
Canada
tel. +1 416 978 0461
`frank@comm.utoronto.ca`

En-hui Yang
Dept. of Electrical and Computer Engineering
University of Waterloo
200 University Avenue West
Waterloo, Ontario N2L 3G1
Canada
tel. +1 519 888 4567, ext. 32873
`ehyang@uwaterloo.ca`

Technical Program Committee
H. Bölcskei (co-chair)
R. Koetter (co-chair)
G. Kramer (co-chair)
V. Anantharam
A. Ashikhmin
J.-C. Belfiore
E. Biglieri
I. F. Blake
H. Boche
N. Cai
P. A. Chou
T. M. Cover
I. Csiszár
S. Diggavi
I. Dumer
H. El Gamal
Y. Eldar
U. Erez
M. Feder
G. D. Forney, Jr.
B. J. Frey
M. C. Gastpar
A. Grant
P. Gupta
B. Hajek
T. S. Han
T. Helleseth
M. L. Honig
J. B. Huber
H. Imai
N. Jindal
T. Johansson
I. Kontoyiannis
S. Kulkarni
P. R. Kumar
J. N. Laneman
A. Lapidoth
S. Litsyn
N. Merhav
A. Montanari
P. Moulin
R. R. Müller
P. Narayan
K. R. Narayanan
A. Orlitsky
K. G. Paterson
S. S. Pradhan
A. Ramamoorthy
K. Ramchandran
R. Renner
R. M. Roth
S. A. Savari
S. Shamai (Shitz)
A. C. Singer
E. Soljanin
R. Srikant
W. Szpankowski
G. Ungerboeck
R. L. Urbanke
A. Vardy
V. V. Veeravalli
S. Vishwanath
P. Viswanath
P. O. Vontobel
J. K. Wolf
K. Zeger

General Co-Chairs
F. R. Kschischang
E.-H. Yang
Student Travel Grants
R. Kerr

Finance
W. Yu
Recent Results
A. Banihashemi
N. Kashyap

International Liaison
H.-A. Loeliger
L. Ping
G. W. Wornell
Publicity
S. Yousefi

Tutorials
B. J. Frey
Local Arrangements
R. S. Adve
T. J. Lim

Publications
J.-Y. Chouinard
Companions' Program
C. Kschischang

WiOpt'08

6th Intl. Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks March 31 - April 4, Berlin, Germany http://www.wiopt.org

Scope of the Symposium

This symposium intends to bring together researchers and practitioners working on optimization of wireless network design and operations. It welcomes works on different perspectives, including performance analysis and simulation, protocol design, numerical communication and optimization theory, for all forms of wireless networks: cellular, wide, metropolitan, local and personal-area networks, dense and sparse ad-hoc networks, domain specific vehicular, public-transport, application-specific sensor networks, as well as any combination of these.

Topics of interest include, but are not limited to:

* Modeling, simulations and measurements

* Protocol design

* Spectrum allocation

* Security and intrusion detection

* Pricing and incentives

* Scalability, manageability and optimization

* System capacity and performance

* Mobility and multihoming

* Opportunistic and cooperative scheduling

* Cognitive radio

* Interference control

* Energy efficiency

Submissions

The submission format for the papers is an extended abstract, up to eight pages long. Please use the IEEE Transactions format, 11 pt character size, one column text, one-and-a-half line spacing, letter paper. This page budget should contain all figures, tables, references, etc. The extended abstract should also include a brief abstract of up to 150 words. The submission will be handled via EDAS (http://edas.info). Only PDF files are acceptable; please make sure that the paper prints without problems (take care to embed all required fonts, etc.).

Adjunct Workshops

Several one-day workshops are planned to accompany the main WiOpt Symposium:

WiNMee/WiTMeMo 2008 : International Workshop On Wireless Network Measurement

RAWNET 2008 : Resource Allocation in Wireless Networks

SPASWIN 2008: Spatial Stochastic Models for Wireless Networks

WNC3 2008 : Wireless Networks: Communication, Cooperation and Competition

PHYSCOMNET: Physics inspired Paradigms for Wireless Communications and Networks

WMCNR 2008 : Wireless Multihop Communications in Networked Robotics

Important Dates

Conference: April 1-3, 2008

Workshops: March 31-April 4, 2008

Submission deadline: October 1, 2007

Notification of acceptance: December 15, 2007

Camera-ready copy: January 15, 2007

CALL FOR PAPERS

Eleventh International Workshop on Algebraic and Combinatorial Coding Theory ACCT2008

ORGANIZED BY:

⋆ Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
⋆ Institute for Information Transmission Problems, Russian Academy of Sciences

TIME: Monday, June 16 (arrival day) - Sunday, June 22 (departure day), 2008.

PLACE: Hotel Finlandia (http://www.hotelfinlandia.com), Pamporovo, Bulgaria. Pamporovo is the sunniest winter resort in Bulgaria and is located at 230 km from Sofia and 85 km south from Plovdiv. There are good bus connections to Sofia and Plovdiv. Hotel Finlandia offers nice accommodation and conference facilities. More information from the organizers at `acct@moi.math.bas.bg`.

TOPICS:

- Linear codes
- Burst-correcting codes
- Self-dual codes
- Algebraic-geometric codes
- Decoding
- Combinatorial codes
- Covering problems
- Spherical codes and designs
- Cryptography
- Computer systems in coding theory
- Related topics

ORGANIZING COMMITTEE:

L. Bassalygo (Moscow, Co-Chairman)
S. Dodunekov (Sofia, Co-Chairman)
S. Kapralov (Gabrovo)
V. Lebedev (Moscow)
B. Kudryashov (St. Petersburg)
I. Landjev (Sofia)
V. Zyablov (Moscow)
S. Topalova (V. Tarnovo)

PROGRAMME COMMITTEE:

N. Manev (Sofia, Co-Chairman)
G. Kabatiansky (Moscow, Co-Chairman)
S. Bouyuklieva (V. Tarnovo)
P. Boyvalenkov (Sofia)
E. Kolev (Sofia)
V. Levenshtein (Moscow)
F. Solov'eva (Novosibirsk)
V. Zinoviev (Moscow)

REGISTRATION FEE: EURO 450/500 for double/single room prior to May 16, 2008, (EURO 500/550 after May 16, 2008) – includes hotel, full board, social events, an excursion, workshop proceedings, EURO 350 for PhD students, EURO 300 for spouses.

SUBMISSIONS: Authors are invited to submit camera ready papers in Latex (at most six pages) by e-mail to **peter@moi.math.bas.bg** (Peter Boyvalenkov).

DEADLINE FOR SUBMISSIONS: April 15, 2008.

WEB PAGE: http://www.moi.math.bas.bg/acct2008/acct2008.html

CISS 2008

Conference on Information Sciences and Systems

Call for Papers

Authors are invited to submit previously unpublished papers describing theoretical advances, applications, and ideas in the fields of information theory (including application to biological sciences); communication, networking; signal, image, and video processing; systems and control; learning and statistical inference.

Two types of contributed papers are solicited:

- Regular papers, requiring approximately 30 minutes for presentation; these will be reproduced in full (up to six pages) in the conference proceedings.

- Short papers, suitable for presentation in approximately 15 minutes; one-page summaries of these papers will be published in the proceedings.

Electronic summaries in Adobe PDF format, together with a "regular" or "short" designation and 2-3 keywords must be submitted by January 2, 2008 through the conference website. Please refer to the checklist for preparing and submitting conference publication PDF files in accordance with IEEE specifications. Summaries should be of sufficient detail and length to permit careful reviewing. Authors will be notified of acceptance no later than February 1, 2006. **Final manuscripts of accepted papers are to be submitted in PDF format no later than February 22, 2006. These are firm deadlines that will permit the distribution of a CD containing the conference proceedings at the Conference.**

Program Directors

Prof. Robert Calderbank
Prof. Bradley Dickinson
Dept. of Electrical Engineering
Princeton University

Conference Coordinators

Dorothy Coakley
Phone: (609) 258-3152

Beth Jarvie
Phone: (609) 258-2213

Technical Cosponsorship

IEEE Information Theory Society

Mailing Address

CISS 2008
Department of Electrical Engineering
Princeton University
E-Quad, B325
Princeton, NJ 08544
Fax: (609) 258-2158
Email: ciss@princeton.edu

Wireless Internet Access During Conference

http://www.princeton.edu/frist/OITVisitorWireless.pdf

Keep Your Professional Edge with Innovative Courses from IEEE

Staying technically current in today's ever-changing workplace is a career must if you want to maintain your professional edge or your P.E. license as required by more than 30 states in the US. IEEE offers an innovative new product called Expert Now as well as a growing service, Education Partners Program to help meet your continuing professional development needs.

Expert Now is a collection of over 65 one-hour long, interactive online courses on topics such as aerospace, circuits & devices, communications, computing, laser & optics, microwave theory & techniques, power, reliability, signal processing and software. Presented by experts in the field, each course brings to your desktop the best tutorial content IEEE has to offer through their technical meetings that take place worldwide. Continuing Education Units (CEUs) can be earned upon successful completion of the assessment. To review the course catalog visit http://ieeexplore.ieee.org/modules.modulebrowse.jsp.

For those looking for a more robust educational experience, more along the lines of a longer online course or a more traditional classroom setting, the IEEE Education Partners Program can prove helpful in your search for continuing professional development opportunities. Exclusive for IEEE members, it provides access to more than 6,000 on-line courses, certification programs and graduate degree programs at up to a 10% discount from academic and private providers that IEEE has peer reviewed to accept into the program. To review the current list of partners participating in the program visit http://www.ieee.org/web/education/partners/eduPartners.html

Another way to browse for a course or educational events taking place in your area is through the courses registered with IEEE to offer CEUs. To review what's available in your area visit http://www.ieee.org/web/education/ceus/index.html. IEEE is an Authorized provider of CEUs through the International Association for Continuing Education and Training as well as an authorized provider of CEUs for the Florida State Board. IEEE CEUs are also accepted by the New York State Board and can easily be converted into PDHs. One CEU is equal to 10 contact hours of instruction in a continuing education activity. IEEE CEUs readily translate into Professional Development Hours (PDHs) (1 CEU = 10 PDHs).

For more general information on IEEE's Continuing Education products and services, visit http://www.ieee.org/web/education/home/index.html. Specific inquiries can be directed to Celeste Torres via email, c.torres@ieee.org or by phone +1 732 981 3425.

Conference Calendar

DATE	CONFERENCE	LOCATION	CONTACT/INFORMATION	DUE DATE
Jan. 3-4, 2008	**4th Workshop on Network Coding, Theory and Applications (NETCOD2008)**	Hong Kong	http://netcod2008.ie.cuhk.edu.hk	Passed
Jan. 14-16, 2008	**7th International ITG Conference on Source and Channel Coding (SCC 08)**	Ulm, Germany	http://www.mk.tu-berlin.de/scc08	Passed
January 28 - February 1, 2008	**2008 Information Theory and Applications Workshop (ITA 2008)**	San Diego, CA, USA	http://ita.ucsd.edu/workshop.php	TBA
April 13 - 18, 2008	**2008 IEEE Conference on Computer Communications (INFOCOM 2008)**	Phoenix, AZ, USA	http://www.ieee-infocom.org/	Passed
March 12-14, 2008	**The 2008 International Zurich Seminar on Communications (IZS 2008)**	Zurich, Switzerland	http://www.izs2008.ethz.ch	Passed
March 12-14, 2008	**3rd International Symposium on Communications, Control and Signal Processing (ISCCSP08)**	St. Julian, Malta	http://guinevere.eng.um.edu.mt/ isccsp2008/	Passed
March 19-21, 2008	**Conference on Information Sciences and Systems (CISS08)**	Princeton, NJ USA	http://conf.ee.princeton.edu/ciss/	January 2, 2008
March 31 – April 4 2008	**6th International Symposium on Modeling and Optimization in Mobile, Ad-Hoc, and Wireless Networks (WiOpt'08)**	Berlin, Germany	http://wiopt.orh	October 1, 2008
May 5-9, 2008	**2008 IEEE Information Theory Workshop (ITW 2008)**	Porto, Portugal	http://www.dcc.fc.up.pt/~itw2008/	March 7, 2008
May 19 – 23, 2008	**2008 IEEE International Conference on Communications (ICC 2008)**	Beijing, China	http://www.ieee-icc.org/2008/	Passed
July 6 – 11, 2008	**2008 IEEE International Symposium on Information Theory (ISIT 2008)**	Toronto, Canada	http://www.isit2008.org	January 7, 2008
September 1 – 5, 2008	**2008 International Symposium on Turbo Codes and Related Topics**	Lausanne, Switzerland	http://www-turbo.enst-bretagne.fr/	TBA