IEEE Information Theory Society Newsletter

Vol. 61, No. 1, March 2011 Editor: Tracey Ho ISSN 1045-2362

Editorial committee: Helmut Bölcskei, Giuseppe Caire, Meir Feder, Joerg Kliewer, Anand Sarwate, and Andy Singer

President's Column

Giuseppe Caire

I am deeply honored to serve as your Society president for 2011. I'd like to begin with acknowledging my recent predecessors, Frank Kschischang (Junior Past President) and Andrea Goldsmith (Senior Past President), as well as Dave Forney, who has officially ended his service as an officer of the IT Society on December 31st, 2010, after a long and perhaps unmatched history of leadership and guidance, that stands as a reference for all of us. Under their lead, the IT Society has sailed through some significant storms, such as an unprecedented economic recession, causing turmoil for IEEE finances, and a number of sad events, including the premature passing of Shannon Lecturer Rudolf Ahlswede and other notables. Thanks to their commitment and vision, our Society is very much alive and kicking, we have healthy reserves and strong income, the membership downtrend has been reversed, we have an outstanding Society website that has become a very valuable tool for membership activities and visibility to the outside world, and we have put in place a number of highly innovative initiatives. These include the Student Committee, the Outreach Committee, WITHITS, the Distinguished Lecturers Program, the Annual School of Information Theory, the ever-growing posting of preprints on ArXiv, and the ISIT Student Paper Award. Such initiatives have successfully passed the beta-testing phase and have become pillars of our Society activity.

Also, I would like to acknowledge my outstanding team of volunteer officers, Muriel Médard (First VP), Gerhard Kramer (Second VP), Nihar Jindal (Treasurer) and Natasha Devroye (Secretary), who will share with me the burden of running the Society in 2011; all standing Committee Chairs and members (it would be too long to mention them explicitly, but please have a look at http://www.itsoc.org/people); the Information Theory Society Distinguished Lecturers (http://www.itsoc.org/people/committees/dlp); and all IT Society Chapter volunteers (http://www.itsoc.org/people/chapters).

Given the excellent state of the Information Theory Society, my main goal for 2011 will be "do no harm", a golden rule sadly too often forgotten by many politicians and people with responsibilities around the world. On the other hand, we have a lot of work to do and we must be proactive in facing the challenges that the future is putting in front of us.

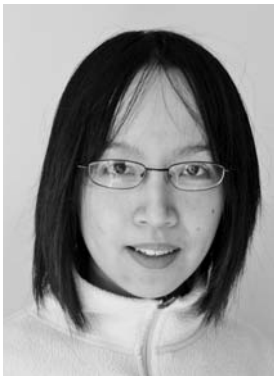
While our Information Theory Transactions continues to rank at the top of the ISI Journal Citation Report among all journals in Electrical and Electronic Engineering and Computer Science in total citations, and *first* among all journals in Electrical Engineering, Computer Science and Applied Mathematics according to the Eigenfactor [tm] score, the "sub-to-pub" time (i.e., the length of the time between first submission and publication of a paper) remains a concern. Some important steps have been taken in order to correct this problem without sacrificing the quality of our Transactions, which is ultimately our most cherished value. I have immense trust in our recently appointed EiC, Helmut Bölcskei, and in the newly created Executive Editorial Board, consisting of Dave Forney, Shlomo Shamai, Alexander Vardy and Sergio Verdú. Under their leadership, I am confident that the sub-to-pub time will be reduced to more acceptable numbers, in line with other comparable IEEE journals. While the transition to Scholar One, effective as of September 2010, is a first step in this direction, we can achieve this ambitious goal only if our collective sense of shared responsibility is further strengthened and we all feel committed to provide timely and high-quality reviews.

Looking at a broader picture, it is useful to consider what Information Theory has achieved and where it is heading. In the past few years, we have witnessed a period of vibrant new discoveries. Decades-long open problems have been solved, either exactly or via good engineering approximations. The mathematical techniques that form the foundation of our discipline, such as random coding, superposition coding, successive interference cancellation, lattice coding and quantization, binning (or hashing), linear and non-linear precoding, opportunistic scheduling and many more, are now

continued on page 3

From the Editor

Tracey Ho



Dear IT Society members,

In this first issue of 2011, we welcome our new IT Society President Giuseppe Caire, who shares his thoughts on our society's strengths and successes, and remaining challenges to be tackled. We also, with sadness, pay tribute to four distinguished colleagues in information theory who passed away recently: Rudolf Ahlswede, Frederick Jelinek, Joseph Ovseyevich and Raymond J. Solomonoff. On a happier note, we congratulate the winners of prestigious IEEE medals and awards, and the newly elevated IEEE Fellows from our society. And we recap the plenary talks by Ram Zamir at ISIT 2010 in Austin and Ian Blake at ITW 2010 in Dublin.

We also have an exciting new column on "Teaching IT". This was inspired by Sergio Verdu's "Teaching IT" Shannon lecture, so it is fitting that he has written the inaugural article, on teaching lossless data compression. As envisioned by Ezio Biglieri, who first raised the idea, and the newsletter editorial committee, the column focuses on the challenge of teaching seemingly complicated technical concepts in the simplest possible way, reducing each to its essence while not sacrificing scope and rigor.

As a reminder, announcements, news and events intended for both the printed newsletter and the website, such as award announcements, calls for nominations and upcoming conferences, can be submitted jointly at the IT Society website http://www.itsoc.org/, using the quick links "Share News" and "Announce an Event". Articles and columns that do not fall into the above categories should be e-mailed to me at tho@caltech.edu, with a subject line that includes the words "IT newsletter". The deadlines for the next few issues are:

Issue	Deadline
June 2011	April 10, 2011
September 2011	July 10, 2011
December 2011	October 10, 2011

Please submit plain text, LaTeX or Word source files; do not worry about fonts or layout as this will be taken care of by IEEE layout specialists. Electronic photos and graphics should be in high resolution and sent as separate files.

I look forward to your contributions and suggestions for future issues of the newsletter.

Tracey Ho

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 *continued from page 1*

at the heart of core technology developments, and are migrating into new systems and communication standards. Central to this recent development is the concept of *networks*. It is by now clear that the single-source single-link problem has arrived at a point where the marginal improvement, in most settings of engineering significance, is relatively small. In contrast, as soon as the problems are enriched with network aspects, such as distributed and correlated sources, interference and intermediate nodes that are neither sources nor destinations, the distance between theory and practice is still large, and the margins for dramatic improvements are potentially huge. Furthermore, even the theory offers plenty of long-standing or new open problems that will keep generations of information theorists busy for a long time. Once I heard that Gödel's Theorem is the "life insurance" of mathematicians. Without any claim of mathematical rigor, I'd like to borrow this statement and say that communication networks are the life insurance of information theorists: there is still a lot of work to do!

Despite recent successes and wide demonstration of the impact of Information Theory on technology and, ultimately, on society, wealth and jobs, sometimes I have the impression that the same excitement is not shared by funding agencies, college students enrolling in engineering programs, and university administrators. It is rare, today, to hear our broad area mentioned as a "strategic priority" in any of these environments. Admittedly, there are areas with a higher presence in the popular media and and with the public, which have drained much attention and much of the available research funding. Perhaps this is a backlash from the overemphasis on the broad Information Technology sector during the "dot-com bubble". However, in my humble view, this fact merits some serious reflections: how can we reach out to different areas, and bring the approach and the methods of Information Theory to the solution of a broader set of problems, while preserving the core aspects that made our field well-defined, rigorous and scientifically satisfactory? Does it make sense to broaden the class of problems beyond the classical scenarios? Some partial answers are already being provided in these days, with the enlarging of our field to computation, security, coordination of actions and controls through communication networks, with applications that go beyond the simple reliable communication of information from point A to point B, but touch aspects such as Smart Infrastructure, Information Dissemination in Social Networks and much more. I believe that this evolution is vital for the future of our area, provided that the new problems are stated and attacked with the same brilliant simplicity, broad applicability, and mathematical rigor that characterized the early days of Information Theory.

Now, as the IT Society president for 2011, a legitimate question that I ask myself is what the IT Society can do in order to foster the growth of our field. I believe that our strongest priority is to invest in the future generations of researchers, i.e., in our Ph.D. students, and at the same time to offer visibility to such outstanding students in order to compete for academic and leading industrial research positions around the world. In this respect, the Annual School of Information Theory is a strategic asset to create a vibrant scientific community, and to offer an occasion to doctoral students to get in touch with each other and with leading researchers in our field, in a convivial and more relaxed environment than a conventional conference. The success of this initiative is striking, and I'd like to thank Roberto Padovani for his generous gift that allows the IT Society to support a Padovani Lecturer at the school. Also, it is very comforting to notice that several sister initiatives have been taking place in IEEE regions other than North America. Similarly, the Distinguished Lecturers Program can become a great showcase of our discipline beyond the conventional venues of ISIT and ITW. Finally, I would like to endorse the commitment of IT Society members in serving as liaisons with respect to various initiatives and committees cutting across IEEE and beyond, around themes of great potential interest. There are several other initiatives that we may undertake, as for example contributing to the debate inside IEEE to define a suitable "open publication" model, without hurting our source of revenue due to IEEEXplore.

I would like to close by encouraging all of you to get more involved in the Society. The IT Society's most valuable resource is its outstanding and highly committed members, and we are always in need of dedicated and energetic volunteers to generate new ideas and bring them to life. Board of Governors meetings are open to all, so feel free to attend any or all if you wish, to propose a new initiative, to see how the Society is governed, or just to participate to the discussion. As usual, 2011 BoG meeting agendas are posted on the IT Society website in advance of each meeting. You can also join Society mailing lists to be informed about recent events or to participate in ongoing discussions. Alternatively, you can contact me (at caire@usc.edu) or the other officers with your ideas, thoughts, or concerns. Thanks again for entrusting me with the Presidency of this great Society.

IT Society Members Win IEEE Medals and Awards

2011 IEEE Medals

IEEE Richard W. Hamming Medal

Toby Berger
Cornell University and University of Virginia

"For contributions to Information Theory, including source coding and its applications."

IEEE Jack S. Kilby Signal Processing Medal

Ingrid Daubechies
Princeton University

"For pioneering contributions to the theory and applications of wavelets and filter banks."

IEEE Alexander Graham Bell Medal

Arogyaswami J. Paulraj
Stanford University

"For pioneering contributions to the application of multiantenna technology to wireless communication systems."

2011 IEEE Technical Field Awards

IEEE Koji Kobayashi Computers and Communications Award

For outstanding contributions to the integration of computers and communications.

Thomas J. Richardson
Qualcomm Flarion Technologies

Rüdiger Urbanke
Ecole Polytechnique Fédérale de Lausanne (EPFL)

"For developing the theory and practice of transmitting data reliably at rates approaching channel capacity."

IEEE Eric E. Sumner Award

For outstanding contributions to communication theory

H. Vincent Poor
Princeton University

"For pioneering contributions to multiple-access communications."

IEEE Kiyo Tomiyasu Award

For outstanding early to mid-career contributions to technologies holding the promise of innovative applications.

Moe Z. Win
Massachusetts Institute of Technology

"For fundamental contributions to highspeed reliable communications using optical and wireless channels"

2011 IEEE Donald G. Fink Prize Paper Award

Andreas F. Molisch
University of Southern California

Larry J. Greenstein
Rutgers University-WINLAB

Mansoor Shafi
Telecom New Zealand

For their paper entitled "Propagation Issues for Cognitive Radio," Proceedings of the IEEE, Vol. 97, No. 5, May 2009

Society Awards

2009 Best Paper Award in "Signal Processing and Coding for Data Storage" of the IEEE Communications Society

"Rank Modulation for Flash Memories," by Anxiao (Andrew) Jiang, Robert Mateescu, Moshe Schwartz and Jehoshua Bruck, in IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2659–2673, June 2009

2010 IEEE Communications Society Young Researcher Award for the Europe, Middle East and Africa region

João Barros
University of Porto

2010 Newly Elevated IEEE Fellows

Mark Bell
Purdue University

for contributions to signal design and processing in radar and communication systems

Holger Boche
Technical University of Munich

for contributions to signal processing and multi-user wireless communications

Marco Chiani
University of Bologna

for contributions to wireless communication systems

Keith Chugg
University of Southern California

for contributions to adaptive and iterative data detection and decoding

Tolga Duman
Arizona State University

for contributions to coding and modulation for wireless, recording and underwater acoustic channels

Elza Erkip
Polytechnic Institute of NYU

for contributions to multi-user and cooperative communications

Dennis Goeckel
University of Massachusetts Amherst

for contributions to wireless communication systems and networks

Piyush Gupta
Bell Labs, Alcatel-Lucent

for contributions to wireless networks

Yunghsiang Han
National Taipei University

for contributions to decoding techniques

Robert Heath
The University of Texas at Austin

for contributions to multiple antenna wireless communications

Thomas Kolze
Broadcom Corporation

for contributions to physical layer architecture in communication systems

Ioannis Kontoyiannis
Athens University of Economics and Business

for contributions to data compression

Kwang Bok Lee
Seoul National University

for contributions to high-speed wireless communication systems

Rainer Martin
Ruhr University

for contributions to speech enhancement for mobile communications and hearing aids

Dharmendra Modha
IBM Almaden Research Center

for contributions to cognitive computing and caching algorithms

Hermann Ney Aachen
RWTH University

for contributions to statistical language modeling, statistical machine translation, and large vocabulary speech recognition

Erik Ordentlich
Hewlett-Packard Laboratories

for contributions to universal algorithms and data compression

Shivendra Panwar
Polytechnic Institute of NYU

for contributions to design and analysis of communication networks

Balaji Prabhakar
Stanford University

for contributions to network theory and algorithms

James Ritcey
University of Washington

for contributions to bit-interleaved coded modulation with iterative decoding

William Ryan
University of Arizona

for contributions to channel coding for reliable data transmission and storage

Jawad Salehi
Sharif University of Technology

for contributions to fundamental principles of optical code division multiple access

Anna Scaglione
University of California

for contributions to filterbank precoding for wireless transmission and signal processing for cooperative sensor networks

Yossef Steinberg
Technion Israel Institute of Technology

for contributions to information theory

Mitchell Trott
Hewlett Packard Laboratories

for contributions to wireless communication

Vinay Vaishampayan
AT&T Labs

for contributions to error-resilient compression systems

Emanuele Viterbo
DEIS-Università della Calabria

for contributions to coding and decoding for wireless digital communications

Li-Chun Wang
National Chiao Tung University

for contributions to cellular architectures and radio resource management in wireless networks

Xiaolin Wu
McMaster University

for contributions to image coding, communication and processing

Hirosuke Yamamoto
The University of Tokyo

for contributions to source coding and information-theoretic securec oding

Roy Yates
Rutgers University-WINLAB

for contributions to wireless network resource allocation

In Memoriam
Rudolf Ahlswede 1938–2010

I. Csiszár, N. Cai, K. Kobayashi, and U.Tamm

Rudolf Ahlswede, a mathematician, one of the truly great personalities of Information Theory, passed away on December 18, 2010 in his house in Polle, Germany, due to a heart attack. He is survived by his son Alexander. His untimely death, when he was still very actively engaged in research and was full with new ideas, is an irrecoverable loss for the IT community.

Rudolf Ahlswede

Ahlswede was born on September 15, 1938 in Dielmissen, Germany. He studied Mathematics, Philosophy and Physics in Göttingen, Germany, taking courses, among others, of the great mathematicians Carl Ludwig Siegel and Kurt Reidemeister. His interest in Information Theory was aroused by his advisor Konrad Jacobs, of whom many students became leading scientists in Probability Theory and related fields.

In 1967 Ahlswede moved to the US and became Assistant Professor, later Full Professor at Ohio State University, Columbus. His cooperation during 1967–1971 with J. Wolfowitz, the renowned statistician and information theorist, contributed to his scientific development. Their joint works included two papers on arbitrarily varying channels (AVCs), a subject to which Ahlswede repeatedly returned later.

His first seminal result was, however, the coding theorem for the (discrete memoryless) multiple-access channel (MAC). Following the lead of Shannon's Two-Way Channel paper, this was one of the key results originating Multiuser Information Theory (others were those of T. Cover on broadcast channels and of D. Slepian and J. Wolf on separate coding of correlated sources), and it was soon followed by an extension to two-output MAC's, requiring new ideas. Also afterwards, Ahlswede continued to be a major contributor to this research direction, in collaboration with J. Körner (visiting in Columbus in 1974) and later also with other members of the Information Theory group in Budapest, Hungary. In addition to producing joint papers enriching the field with new results and techniques, this collaboration also contributed to the Csiszár-Körner book where several ideas are acknowledged to be due to Ahlswede or have emerged in discussions with him.

In 1975 Ahlswede returned to Germany, accepting an offer from Universität Bielefeld, a newly established "research university" with low teaching obligations. He was Professor of Mathematics there until 2003, and Professor Emeritus from 2003 to 2010. For several years he devoted much effort to building up the Applied Mathematics Division, which at his initiative included Theoretical Computer Science, Combinatorics, Information Theory, and Statistical Physics. These administrative duties did not affect his research activity. He was able to develop a strong research group working with him, including visitors he attracted as a leading scientist, and good students he attracted as an excellent teacher. In the subsequent years Ahlswede was heading many highly fruitful research projects, several of them regularly extended even after his retirement which is quite exceptional in Germany. The large-scale interdisciplinary project "General Theory of Information Transfer" (Center of Interdisciplinary Research, 2001–2004) deserves special mentioning. It enabled him to pursue very productive joint research with many guests and to organize several conferences. An impressive collection of new scientific results obtained within this project was published in the book "General Theory of Information Transfer and Combinatorics" (Lecture Notes in Computer Science, Springer, 2006).

During his research career Ahlswede received numerous awards and honours. He was recipient of the Shannon Award of the IEEE IT Society in 2006, and previously twice of the Paper Award of the IT Society (see below). He was member of the European Academy of Sciences, recipient of the 1998/99 Humboldt-Japan Society Senior Scientist Award, and he received honorary doctorate of the Russian Academy of Sciences in 2001. He was also honored by a volume of 50 articles on the occasion of his 60'th birthday (Numbers, Information and Complexity, Kluwer, 2000.)

Ahlswede's research interests included also other fields of Applied and Pure Mathematics, such as Complexity Theory, Search Theory (his book "Search Problems" with I. Wegener is a classic), Combinatorics, and Number Theory. Many problems in these disciplines that aroused Ahlswede's interest had connections with Information Theory, and shedding light on the interplay of IT with other fields was an important goal for him. He was likely the first to deeply understand the combinatorial nature of many IT problems, and to use tools of Combinatorics to solve them.

In the tradition of giants as Shannon and Kolmogorov, Ahlswede was fascinated with Information Theory for its mathematical beauty rather than its practical value (of course, not underestimating the latter). In the same spirit, he was not less interested in problems of other fields which he found mathematically fascinating. This is not the right place to discuss his (substantial) results not related to IT. We just mention the celebrated Ahlswede-Daykin "Four Functions Theorem" having many applications in Statistical Physics and in Graph Theory, and the famous Ahlswede-Khachatrian "Complete Intersection Theorem". The latter provided the final solution of a problem of Paul Erdös, which had been very long-standing even though Erdös offered $500 – for the solution (Ahlswede and Khachatrian collected). For more on this, and also on combinatorial results of information theoretic interest, see his book "Lectures on Advances in Combinatorics" with V. Blinovsky (Springer, 2008).

Even within strict sense Information Theory, Ahlswede's contributions are too wide-ranging for individual mentioning, they extend

as far as the formerly exotic but now highly popular field of Quantum Information Theory. Still, many of his main results are one of the following two kinds.

On the one hand, Ahlswede found great satisfaction in solving hard mathematical problems. Apparently, this is why he returned again and again to AVCs, proving hard results on a variety of models. By his most famous AVC theorem, the (average error) capacity of an AVC either equals its random code capacity or zero. Remarkably, this needed no hard math at all, "only" a bright idea, the so-called elimination technique (a kind of derandomization). He was particularly proud of his solution of the AVC version of the Gelfand-Pinsker problem about channels with non-causal channel state information at the sender. To this, the elimination technique had to be combined with really hard math. Another famous hard problem he solved was the "zero excess rate" case of the Multiple Descriptions Problem (the general case is still unsolved).

On the other hand, Ahlswede was eager to look for brand new or at least little studied models, and was also pleased to join forces with coauthors suggesting work on such models. His most frequently cited result (with Cai, Li and Yeung), the Min-Cut-Max-Flow Theorem for communication networks with one source and any number of sinks, belongs to this category. So do also his joint results with Csiszár on hypothesis testing with communication constraints, and with Dueck on identification capacity, receiving the Best Paper Award of the IT Society in 1988 and 1990. Later on, Ahlswede has significantly broadened the scope of the theory of identification, for example to quantum channels (with Winter). Further, a two-part joint paper with Csiszár provides the first systematic study of the concept of common randomness, both secret and non-secret, relevant, among others, for secrecy problems and for identification capacity. The new kind of problems studied in these papers support Ahlswede's philosophical view that the real subject of information theory should be the broad field of "information transfer", which is currently unchartered and only some of its distinct areas (such as Shannon's theory of information transmission and the Ahlswede-Dueck theory of identification) are in view. Alas, Rudi is no longer with us, and extending information theory to cover such a wide scope of yet unknown dimensions will be the task of the new generation.

Frederick Jelinek 1932–2010

Toby Berger, Terrence L. Fine, and Sanjeev Khudanpur

Frederick Jelinek, a post-WW2 teenage Czechoslovakian émigré to the United States, abandoned his childhood dream of becoming a lawyer because he felt a native Czech speaker would be unable to develop the command of English requisite for that profession. Ironically, Jelinek later became the long term manager of the Continuous Speech Recognition Group at IBM Yorktown, where he made profound contributions to both the theory and the practice of automatic English speech recognition systems. Moreover, he held professorial positions at Cornell University before his IBM years and at Johns Hopkins University in his post-IBM career. At the Cornell School of Electrical Engineering in the 1960's, Fred built a strong group of students and faculty members, together with whom he placed Cornell squarely on the map of major centers of research on information theory and error control coding; also, he authored his first book, *Probabilistic Information Theory*. In 1993 Professor Jelinek joined the Hopkins faculty as the Director of the newly formed Center for Language and Speech Processing. He elevated that Center to arguably the premiere institution concerned with the science and engineering of the language-technology interface. Also, he wrote his second book, *Statistical Methods for Speech Recognition.* Designed as a graduate-level monograph for students working in automated language processing research, it was pervasively informed by the landmark work conducted by his Group at IBM.

Frederick Jelinek

Born in Kladno, Czechoslovakia, on November 18, 1932, he was named Bedřich Jelinek. His sister and he briefly enjoyed a pleasant childhood. However, the rise of Nazi Germany led to his father's death in a concentration camp. The oppressive postwar Soviet occupation of Czechoslovakia prompted his mother to emigrate to New York with her children, at which point his given name was changed to Frederick. Fred studied engineering at City College of New York and received a scholarship that permitted him to complete his undergraduate studies at MIT. In the mid 1950's he made his first return to Czechoslovakia, where a childhood friend who had become a filmmaker introduced him to Milena Tobolova, an aspiring screenwriter who in time became Fred's wife. Jerome Wiesner, a Science Advisor to Presidents Eisenhower, Kennedy and Johnson who later became President of MIT, is said to have acted on behalf of Fred and Milena by requesting Soviet Premier Nikita Kruschchev expedite permission for Milena to emigrate to the U.S.A. Whether or not this was the catalyst will never be known, but Milena was granted the permission, and Fred and Milena soon after married. They raised two children, Hannah and William. Milena continues to teach film making at Columbia University and to be an active member of the Czechoslovakian film making community.

Fred's inclination toward linguistic research dates back at least to his MIT days. Indeed, he claimed that one of his principal reasons for joining the Cornell faculty was that he would have a chance to work with the eminent linguist, Charles Hockett. He was pleasantly impressed by the EE Chair, Henry Booker, who offered him an assistant professorship immediately following the presentation of his colloquium. Before Fred's recruitment, the faculty at Cornell interested in information theory and communications were the eminent Jacob Wolfowitz in Math and Henry McGaughan

and Nicholas DeClaris in EE. Fred was Cornell EE's first full-time information theorist. He wanted to know "how the information universe worked." He had notions of how different areas of technology were connected to information theory, particularly computational complexity and the role of early computers in communication processing.

Fred propelled the growth of a group at Cornell with interests related to information theory. Tom Gaarder, a student of Norm Abramson, was hired in 1965, but left to join Abramson in Hawaii in 1967. Fred recruited Terry Fine who started in Fall 1966. Neil J.A. Sloane had been a graduate student in EE working on a dissertation in neural networks that was motivated by Frank Rosenblatt's pioneering work at Cornell. Fred was Neil's advisor of record in EE. When Neil finished in 1967 he was hired as an assistant professor in EE by Fred and Terry. Neil created the first course in coding theory at Cornell and began his monumental handbook of integer sequences Toby Berger was recruited in 1968 from Raytheon, where he had already begun work on his classic "Rate Distortion Theory." There was a weekly meeting dubbed the Information, Communications and Decision Theory "Syndicate". It consisted of Fred, Terry, Neil, Toby and their students engaging in an evening of research presentations and socializing. Elwyn Berlekamp enticed Neil to Bell Labs in mid-1969. Upon Fred's departure for IBM in 1972, Toby became Cornell's lead information theorist. The last information-theoretic hire made while Fred was at Cornell was that of Tom Cover's student, Patrick Bergmans. Patrick served from 1972–74, when he returned to his native Belgium and pursued a multifaceted career in academia, as a printing industry entrepreneur, and eventually as a Xerox executive. In addition to Neil, Fred advised several Cornell EE PhD students including Frank Huband (Exec. Dir. ASEE), Ken Schneider (Telebyte founder and CEO), John Anderson (Professor at McMaster and Lund), and Hen-Suh Park (Korea Telecom Industry Association).

Most of Fred's early publications were in the IEEE Transactions on Information Theory. Particularly notable among them was *Tree Encoding of Memoryless Time-Discrete Sources with a Fidelity Criterion*, *IEEE* Trans. IT-15:5, 584–590, 1969, which won the IT Group's Outstanding Paper Award. Fred was elected to the Administrative Committee of the IT Group and rose to its presidency in 1977; his Cornell hires Berger and Fine each also served as IT Group presidents. Jelinek played a key role in the organization of ISIT77 held at Cornell, an event that attracted media attention because, over NSA objections, papers on cryptography were for the first time scheduled and presented publicly.

Fred spent a sabbatical at the T.J. Watson Research Labs of IBM. This experience predisposed him to join and then lead the new group on Speech Recognition that was forming there in 1972 under Joe Raviv's leadership. When Raviv left to create the IBM laboratory in Haifa, Fred became leader of this group. Cornell gave Fred the maximum leave of two years before requiring that he return. Electing to remain at IBM, Fred led a group that paved a new path in speech recognition that is still state-of-the-art. It was a rare and happy confluence between the circle of information-theoretic ideas that appealed to Fred and a path-setting highly successful approach to this problem.

Early in his IBM days Fred published the now-classic paper, *Continuous speech recognition by statistical methods*, in the 1976 Proceedings of the IEEE. The ideas set forth therein, although considered

From L to R, Patrick Bergmans, Toby Berger, Fred Jelinek, and Terry Fine at the IEEE International Symposium on Information Theory in Ashkelon, Israel, Summer 1973.

heretical by certain linguistics experts of that era, have become the foundation of virtually every practical system that addresses the continuous speech recognition problem. Fred and his co-authors Lalit Bahl, John Cocke and Joe Raviv are widely known for their short but highly influential paper, *Optimal decoding of linear codes for minimizing symbol error rate*, IEEE Trans IT, 284-287, 2003, which treated "the general problem of estimating the a posteriori probabilities of the states and transitions of a Markov source observed through a discrete memoryless channel." Known ever since as the BCJR algorithm for hidden Markov chains, it has enjoyed widespread application not only to automatic speech recognition but also in such diverse fields as stock market analysis and error correcting codes. Indeed, some members of Fred's IBM lab subsequently founded hugely successful hedge funds, perhaps by applying the BCJR algorithm. Steve Wicker tells us that "the BCJR algorithm (the J is for Jelinek) is a critical element in Turbo decoding. There is thus a little bit of Fred in every 3G cell phone on the planet."

At Hopkins Fred not only spearheaded the development of CSLP's technical staff but also was instrumental in developing a strong graduate student and postdoc presence. Through a steadily increasing program of internships and the venerated Johns Hopkins Workshops on Language and Speech, he expanded CSLP into a vehicle for synergistic research with budding researchers and their mentors from institutions around the country and around the world. Steve Young, in his commentary *Frederick Jelinek 1932–2010: The Pioneer of Speech Recognition Technology*, SLTC Newsletter, November 2010, sums up the breadth and depth of Fred Jelinek's contributions as follows:

> He was not a pioneer of speech recognition; he was the pioneer of speech recognition. His contribution has been recognized by many awards. He received the IEEE Signal Processing Society award in 1998 and the IEEE Flanagan Award in 2005. He received the ISCA Medal for Scientific Achievement in 1999 and he was made an inaugural ISCA Fellow in 2008. He was awarded an Honorary Doctorate at Charles University of Prague in 2001 and he was elected to the National Academy of Engineering in 2006….Fred Jelinek was an inspiration to our community. He will be sorely missed by all who knew him.

At the event *Remembering and Celebrating the Life of Frederick Jelinek* held at Johns Hopkins on November 6, 2010, many people from all

the stages of Fred's personal and professional lives gathered to share their memories of Fred and his works. We end this tribute to Fred with remarks made there by Noah Smith, an assistant professor of computer science at Carnegie Mellon. We feel Noah's remarks, printed here with his permission, succinctly and effectively capture much of the essence of Fred Jelinek's personality and spirit:

> When my career started, a little more than a decade ago, at a summer workshop here, Fred was there. He offered subtle guidance. More, he paid attention. Fred made time for people who wanted to learn, no matter how green they were. He never held back disagreement or skepticism, but he was open to persuasion, especially if you could back up your argument with data. He took pleasure in seeing us stand up earnestly to his earnest challenges. In questioning, he sought understanding, and he did so with reckless abandon and not a hint of self-consciousness. He took the work seriously and never compromised on the science, but he didn't take himself too seriously. He understood the social side of science and the value of sitting around a table with colleagues and taking pleasure in their company. Fred was living proof that senior scholars can stay engaged till the end. After I graduated, and he found me a job, he still called me from time to time to check on how things were going.
>
> I still have his unmistakable voice on my office answering machine. "Noah: It's Fred. Call me back." There is no one like him. Even on the answering machine, his voice fills the room.
>
> Good role models are hard to come by, and rarer still is a role model you seek to imitate without realizing that it's happening. Our field, our academic family, will not be the same without him, but his legacy as a complete scholar will be passed down for many generations to come.

Joseph Ovseyevich 1916–2010

Vadim Stefanuk and Martin Hellman

It is with sadness that we report the death of our friend and colleague, Joseph Ovseyevich, on December 13, 2010, at the age of 94. Prof. Ovseyevich served as the Scientific Director of the Institute for Problems of Information Transmission of the Russian Academy of Sciences from the Institute's inception in 1961. From its Russian name, *Institut Problemy Peredachi Informatsii,* the Institute is known as IPPI. Ovseyevich played an important role in establishing IPPI as one of the world's leading information theory research organizations. Well known contributors to our field who worked at IPPI or were influenced by it include Roland Dobrushin, Rafail Khasminskii, Gregori Margulis, Mark Pinsker, Vadim Stefanuk, Albert Shiryaev, Yuri Shtarkov, Boris Tsybakov, Rom Varshamov, Akiva Yaglom, Viktor Ziablov, and Kamil Zigangirov.

Joseph Ovseyevich

Due to Ovseyevich's efforts IPPI became a leading scientific organization in Russia, known for its achievements in areas ranging from biology to linguistics to telecommunications, and of course, information theory. Professor Ovseyevich was known in the Academy of Sciences not only for his talent in leading IPPI at a scientific level, but also for his deep interest in the people residing there. His door was always open, and many at IPPI valued his thoughtful advice and support on personal as well as scientific matters. Those who knew him, both in Russia and abroad, valued him as a reliable and trustworthy friend.

Professor Ovseyevich played a key role in organizing a number of International Symposia on Information Theory, sponsored by the Soviet (and later Russian) Academy of Sciences. These were often held back-to-back with similarly named IEEE Information Theory Symposia, a practice that facilitated attendance by Western researchers as well as spotlighting the research results of some prominent Soviet researchers who were unable to travel abroad. This was part of his larger effort to build bridges between IPPI and the IEEE Information Theory Society.

Yosif Abramovich "Joseph" Ovseyevich was born in Yaroslavl city in 1916, and in 1923 moved to Moscow. He graduated from Moscow Institute of Communication Engineering in 1940. His scientific career was interrupted by the Second World War, during which he served in the infantry. After the Nazis were defeated, he left the Red Army as a highly decorated officer with the rank of Major. In 1946 he returned to scientific activity within the Soviet Academy of Sciences. He received his Ph.D. in 1954, and his Doctor of Science (Dr.Sc.) in 1973, after defending his thesis titled "The Methods of Information Transmission in Analog Networks."

His further scientific interests were devoted to information theory, mainly to the theoretical problems of the throughput of real radio channels, linear distortions and their correction. Professor Ovseyevich published over 70 papers.

His outside interests included poetry and playing piano. In spite of poor health in recent years, he continued visiting IPPI, even taking part in a number of scientific meetings there.

On December 16, 2010, a last tribute was paid to Professor Ovseyevich at the Presidium of the Russian Academy of Sciences. Everyone present stressed that he will be remembered as an outstanding person of courage, leadership, intelligence and warmth. Our colleagues at IPPI have committed themselves to maintain the high standards of scientific and personal life embodied in Professor Joseph Ovseyevich. He will be missed.

Raymond J. Solomonoff 1926–2009

Peter Gács and Paul M. B. Vitányi

Ray Solomonoff, the first inventor of some of the fundamental ideas of Algorithmic Information Theory, died in December, 2009. His original ideas helped start the thriving research areas of algorithmic information theory and algorithmic inductive inference. His scientific legacy is enduring and important. He was also a highly original, colorful personality, warmly remembered by everybody whose life he touched. We outline his contributions, placing it into its historical context, and the context of other research in algorithmic information theory.

1. Introduction

Raymond J. Solomonoff died on December 7, 2009, in Cambridge, Massachusetts. He was the first inventor of some of the fundamental ideas of Algorithmic Information Theory, which deals with the shortest effective description length of objects and is commonly designated by the term "Kolmogorov complexity."

In the 1950s Solomonoff was one of the first researchers to introduce probabilistic grammars and the associated languages. He championed probabilistic methods in Artificial Intelligence (AI) when these were unfashionable there, and treated questions of machine learning early on. But his greatest contribution is the creation of Algorithmic Information Theory.

In November 1960, Solomonoff published the report [14] presenting the basic ideas of Algorithmic Information Theory as a means to overcome serious problems associated with the application of Bayes's rule in statistics. His findings (in particular, the invariance theorem) were mentioned prominently in April 1961 in Minsky's symposium report [8]. (Andrei N. Kolmogorov, the great Russian mathematician, started lecturing on description complexity in Moscow seminars about the same time.)

Solomonoff's objective was to formulate a completely general theory of inductive reasoning that would overcome shortcomings in Carnap's [1]. Following some more technical reports, in a long journal paper in two parts he introduced "Kolmogorov" complexity as an auxiliary concept to obtain a universal a priori probability and proved the invariance theorem that, in various versions, is one of the characteristic elements of Algorithmic Information Theory [16,17]. The mathematical setting of these ideas is described in some detail below.

Solomonoff's work has led to a novel approach in statistics leading to applicable inference procedures such as the minimal description length principle. Jorma J. Rissanen, credited with the latter, relates that his invention is based on Solomonoff's work with the idea of applying it to classical statistical inference [10,11].

Since Solomonoff is the first inventor of Algorithmic Information Theory, one can raise the question whether we ought to talk about "Solomonoff complexity". However, the name "Kolmogorov complexity" for shortest effective description length has become well entrenched and is commonly understood. Solomonoff's publications apparently received little attention until Kolmogorov started to refer to them from 1968 onward. Says Kolmogorov, "I came to similar conclusions [as Solomonoff], before becoming aware of Solomonoff's work, in 1963–1964" and "The basic discovery, which I have accomplished independently from and simultaneously with R. Solomonoff, lies in the fact that the theory of algorithms enables us to eliminate this arbitrariness by the determination of a 'complexity' which is almost invariant (the replacement of one method by another leads only to the addition of a bounded term)"

Solomonoff's early papers contain in veiled form suggestions about randomness of finite strings, incomputability of Kolmogorov complexity, computability of approximations to the Kolmogorov complexity, and resource-bounded Kolmogorov complexity.

Kolmogorov's later introduction of complexity was motivated by information theory and problems of randomness. Solomonoff introduced algorithmic complexity independently and earlier and for a different reason: inductive reasoning. Universal a priori probability, in the sense of a single prior probability that can be substituted for each actual prior probability in Bayes's rule was invented by Solomonoff with Kolmogorov complexity as a side product, several years before anybody else did.

A third inventor is Gregory J. Chaitin, who formulated a proper definition of Kolmogorov complexity at the end of his paper [2].

For a more formal and more extensive study of most topics treated in this paper, we recommend [7].

2. The Inventor

Ray Solomonoff published a scientific autobiography up to 1997 as [23]. He was born on July 25, 1926, in Cleveland, Ohio, in the United States. He studied physics during 1946-1950 at the University of Chicago (he recalls the lectures of E. Fermi). He obtained a Ph.B. (bachelor of philosophy) and a M.Sc. in physics. He was already interested in problems of inductive inference and exchanged viewpoints with the resident philosopher of science at the University of Chicago, Rudolf Carnap, who taught an influential course in probability theory.

From 1951-1958 he held half-time jobs in the electronics industry doing math and physics and designing analog computers.

In 1956, Solomonoff was one of the 10 or so attendees of the Dartmouth Summer Research Conference on Artificial Intelligence, at Dartmouth College in Hanover, New Hampshire, organized by

M. Minsky, J. McCarthy and C.E. Shannon, and in fact stayed on to spend the whole summer there. (This meeting gave AI its name.) There Solomonoff wrote a memo on inductive inference.

McCarthy had the idea that given every mathematical problem, it could be brought into the form of "given a machine and a desired output, find an input from which the machine computes that output." Solomonoff suggested that there was a class of problems that was not of that form: "given an initial segment of a sequence, predict its continuation." McCarthy then thought that if one saw a machine producing the initial segment, and then continuing past that point, would one not think that the continuation was a reasonable extrapolation? With that the idea got stuck, and the participants left it at that.

Also in 1956, Ray circulated a manuscript of "An Inductive Inference Machine" at the Dartmouth Summer Research Conference on Artificial Intelligence, and in 1957 he presented a paper with the same name at the IRE Convention, Section on Information Theory, a forerunner of the IEEE Symposium on Information Theory. This partially used Chomsky's paper [3] read at a Symposium on Information Theory held at MIT in September 1956. "An Inductive Inference Machine" already stressed training sequences and using previous solutions in solving more complex problems. In about 1958 he left his half-time position in industry and joined Zator Company full time, a small research outfit located in some rooms at 140 1/2 Mount Auburn Street, Cambridge, Massachusetts, which had been founded by Calvin Mooers around 1954 for the purpose of developing information retrieval technology. Floating mainly on military funding, Zator Co. was a research front organization, employing Mooers, Solomonoff, Mooers's wife, and a secretary, as well as at various times visitors such as Marvin Minsky. It changed its name to the more martial sounding Rockford Research (Rockford, Illinois, was a place where Mooers had lived) around 1962. In 1968, the US Government reacted to public pressure (related to the Vietnam War) by abolishing defense funding of civil research, and Rockford foundered. Being out of a job, Solomonoff left and founded his own (one-man) company, Oxbridge Research, in Cambridge in 1970, and has been there ever since, apart from spending nine months as research associate at MIT's Artificial Intelligence Laboratory, the academic year 1990-1991 at the University of Saarland, Saarbruecken, Germany, and a more recent sabbatical at IDSIA, Lugano, Switzerland.

It is unusual to find a productive major scientist that is not regularly employed at all. But from all the elder people (not only scientists) we know, Ray Solomonoff was the happiest, the most inquisitive, and the most satisfied. He continued publishing papers right up to his death at 83.

In 1960 Solomonoff published [14], in which he gave an outline of a notion of universal a priori probability and how to use it in inductive reasoning (rather, prediction) according to Bayes's rule. This was sent out to all contractors of the Air Force who were even vaguely interested in this subject. In [16,17], Solomonoff developed these ideas further and defined the notion of enumeration, a precursor of monotone machines, and a notion of universal a priori probability based on his variant of the universal monotone machine. In this way, it came about that the original incentive to develop a theory of algorithmic information content of individual objects was Solomonoff's invention of a universal a priori probability that can be used as a priori probability in applying Bayes's rule.

Solomonoff's first approach was based on Turing machines with markers that delimit the input. This led to awkward convergence problems with which he tried to deal in an ad-hoc manner. The young Leonid A. Levin (who in [27] developed his own mathematical framework, which became the source of a beautiful theory of randomness), was told by Kolmogorov about Solmonoff's work. He added a reference to it, but had in fact a hard time digesting the informalities; later though, he came to appreciate the wealth of ideas in [16]. Solomonoff welcomed Levin's new formalism with one exception: it bothered him that the universal a priori probability for prediction is a semimeasure but not a measure (see below). He continued to advocate a normalization operation keeping up a long technical argument with Levin and Solovay.

In 2003 he was the first recipient of the Kolmogorov Award by The Computer Learning Research Center at the Royal Holloway, University of London, where he gave the inaugural Kolmogorov Lecture. Solomonoff was a visiting Professor at the CLRC. A list of his publications (published and unpublished) is at http://world.std.com/~rjs/pubs.html.

3. The Formula

Solomonoff's main contribution is best explained if we start with his inference formula not as he first conceived it, but in the cleaner form as it is known today, based on Levin's definition of apriori probability [27]. Let T be a computing device, say a Turing machine. We assume that it has some, infinitely expandable, internal memory (say, some tapes of the Turing machine). At each step, it may or may not ask for some additional input symbol from the alphabet $\{0, 1\}$, and may or may not output some symbol from some finite alphabet Σ. For a finite or infinite binary string p, let $T(p)$ be the (finite or infinite) output sequence emitted while not reading beyond the end of p. Consider the experiment in which the input is an infinite sequence of tosses of an independent unbiased coin. For a finite sequence $x = x_1 \ldots x_n$ written in the alphabet Σ, let $M_T(x)$ be the probability that the sequence outputted in this experiment begins with x. More formally, let $T^{-1}(x)$ be the set of all those binary sequences p that the output string $T(p)$ contains x as a prefix, while if p' is a proper prefix of p then $T(p')$ does not output x yet. Then

$$M_T(x) = \sum_{p \in T^{-1}(x)} 2^{-|p|}, \tag{1}$$

where $|p|$ is the length of the binary string p. The quantity $M_T(x)$ can be considered the *algorithmic probability* of the finite sequence x. It depends, of course, on the choice of machine T, but if T is a universal machine of the type called *optimal* then this dependence is only minor. Indeed, for an optimal machine U, for all machines T there is a finite binary r_T with the property $T(p) = U(r_T p)$ for all p. This implies $M_U(x) \geq 2^{-|r_T|} M_T(x)$ for all x. Let us fix therefore such an optimal machine U and write $M(x) = M_U(x)$. This is (the best-known version of) Solomonoff's *apriori probability*.

Now, Solomonoff's prediction formula can be stated very simply. Given a sequence x of experimental results, the formula

$$\frac{M(xy)}{M(x)} \tag{2}$$

assigns a probability to the event that x will be continued by a sequence (or even just a symbol) y. In what follows we will have

opportunity to appreciate the theoretical attractiveness of the formula: its prediction power, and its combination of a number of deep principles. But let us level with the reader: it is incomputable, so it can serve only as an ideal embodiment of some principles guiding practical prediction. (Even the apriori probability $M(x)$ by itself is incomputable, but it is at least approximable by a monotonic sequence from below.)

4. First, Informal Ideas

Scientific ideas of great originality, when they occur the first time, rarely have the clean, simple form that they acquire later. Nowadays one introduces description complexity ("Kolmogorov" complexity) by a simple definition referring to Turing machines. Then one proceeds to a short proof of the existence of an optimal machine, further to some simple upper and lower bounds relating it to probability and information. This a highly effective, formally impeccable way to introduce an obviously interesting concept.

Inductive inference is a harder, more controversial issue than information and randomness, but this is the problem that Solomonoff started with! In the first papers, it is easy to miss the formal definition of complexity since he uses it only as an auxiliary quantity; but he did prove the machine independence of the length of minimal codes.

The first written report seems to be [14]. It cites only the book [1] of Carnap, whose courses Solomonoff attended. And Carnap may indeed have provided the inspiration for a probability based on pure logical considerations. The technical report form allowed the gradual, informal development of ideas.

The work starts with confining the considerations to one particular formal representation of the general inference problem: predicting the continuations of a finite sequence of characters. Without making any explicit references, it sets out to combine two well-studied principles of inductive inference: Bayesian statistics and the principle that came to be known (with whatever historic justification) as "Occam's Razor". A radical version of this principle says that we should look for a shortest explanation of the experimental results and use this explanation for prediction of future experiments. In the context of prediction, it will be therefore often justified to call descriptions *explanations*.

Here is the second paragraph of the introduction:

> Consider a very long sequence of symbols – e.g., a passage of English text, or a long mathematical derivation. We shall consider such a sequence of symbols to be "simple" and have high a priori probability, if there exists a very brief description of this sequence – using, of course, some sort of stipulated description method. More exactly, if we use only the symbols 0 and 1 to express our description, we will assign the probability 2^{-n} to a sequence of symbols, if its shortest possible binary description contains n digits.

The next paragraph already makes clear that what he will mean by a short "description" of a string x: a program of a general-purpose computer that outputs x.

The combination of these three ingredients: *simplicity, apriori probability, universal computer* turned out to have explosive power, forming the start of a theory that is far from having exhausted its potential now, 50 years later. This was greatly helped by Kolmogorov's independent discovery that related them explicitly to two additional classical concepts of science: *randomness* and *information*.

There is another classical principle of assigning apriori probabilities that has been given a new interpretation by Solomonoff's approach: *Laplace's principle of indifference*. This says that in the absence of any information allowing to prefer one alternative to another, all alternatives should be assigned the same probability. This principle has often been criticized, and it is indeed not easy to delineate its reasonable range of applicability, beyond the cases of obvious symmetry. Now in Solomonoff's theory, Laplace's principle can be seen revived in the following sense: if an outcome has several possible formal descriptions (interpreted by the universal monotonic machine), then *all descriptions of the same length are assigned the same probability*.

The rest of the report [14] has a groping, gradual nature as it is trying to find the appropriate formula for apriori probability based on simplicity of descriptions.

The problems it deals with are quite technical in nature, that is it is (even) less easy to justify the choices made for their solution on a philosophical basis. As a matter of fact, Solomonoff later uses (normalized versions of) (2) instead of the formulas of these early papers. Here are the problems:

1) Machine dependence. This is the objection most successfully handled in the paper.

2) If we assign weight 2^{-n} to binary strings of length n then the sum of the weights of all binary strings is infinite. The problem is dealt with in an ad-hoc manner in the report, by assigning a factor $(1-\epsilon)^k$ to strings of length k. Later papers, in particular Solomonoff's first published paper [16] on the subject, solve it more satisfactorily by using some version of definition (1): on monotone machines, the convergence problem disappears.

3) We should be able to get arbitrary conditional probabilities in our Bayesian inference, but probability based on shortest description leads to probabilities that are powers of two. Formula (2) solves this problem as simply as it solved the previous one, but the first publication [16] did not abandon the ad-hoc approach of the technical report yet either, summing up probabilities for all continuations of a certain length (and taking the limit).

4) There are principles of induction suggesting that not only minimal descriptions (explanations) should be considered. Formula (2) incorporates all descriptions in a natural manner. Again, the ad-hoc approach, extending the sum over all descriptions (weighted as above), still is also offered in [16].

It remained for later researchers (Kolmogorov, Levin) to discover that – in certain models (though not on monotonic computers) even to within an additive constant – asymptotically, the logarithm of the apriori probability obtained this way is the same as the length of the shortest description. Thus, a rule that bases prediction on shortest explanations is not too different from a rule using the prediction fitting "most" explanations. In terms of the monotone machines, this relation can be stated as follows. For a string x, let $Km(x)$ be the length of the shortest binary string

that causes the fixed optimal monotonic machine to output some continuationo f x. Then

$$Km(x) - 2\log Km(x) \leq -\log M(x) \leq Km(x). \quad (3)$$

The paper [16] offers yet another definition of apriori probability, based on a combination of all possible computable conditional probabilities. The suggestion is tentative and overly complex, but its idea has been vindicated by Levin's theorem, in [27], showing that the distribution $M(x)$ dominates all other "lower semicomputable semimeasures" on the set of infinite sequences. (Levin did not invent the universal semimeasure $M(x)$ as response to Solomoff's work, but rather as a natural technical framework for treating the properties of complexity and randomness.) Here, the *semimeasure* property requires, for all x, the inequalities $M(x) \geq \sum_{b\in\Sigma} M(xb)$, while $M(\Lambda) \leq 1$ for the empty string Λ. Lower semicomputability requires that $M(x)$ is the limit of an increasing sequence of functions that is computable in a uniform way. A computable measure is certainly also a lower semicomputable semimeasure. The dominance property distinguishes Solomonoff's apriori probability among all lower semicomputable semimeasures. Levin's observation is crucial for all later theorems proved about apriori probability; Solomonoff made important use of it later.

The paper [17] considers some simple applications of the prediction formulas, for the case when the sequence to be predicted is coming from tossing a (possibly biased) coin, and when it is coming from a stochastic context-free grammar. There are some computations, but no rigorous results.

5. The Prediction Theorem

Solomonoff wrote an important paper [18] that is completely traditional in the sense of having a non-trivial theorem with a proof. The result serves as a justification of the prediction formula (2). What kind of justifications are possible here? Clearly, not all sequences can be predicted successfully, no matter what method is suggested. The two possibilities are:

1) Restrict the kind of sources from which the sequences may be coming, to a still sufficiently wide class.

2) Show that in an appropriate sense, your method is (nearly) as good as any other method, in some wide class of methods.

There is a wealth of research on inference methods considering a combination of both kinds of restriction simultaneously, showing typically that for example if a sequence is generated by methods restricted to a certain complexity class then a successful prediction method cannot be restricted to the same class.

Solomonoff's theorem restricts consideration to sources $x_1x_2\ldots$ with some computable probability distribution P. Over a finite alphabet Σ, let $P(x)$ denote the probability of the set of all infinite sequences starting with x, further for a letter b of the alphabet denote $P(b\,|\,x) = P(xb)/P(x)$. The theorem says that the formula $M(b\,|\,x_1\ldots x_n)$, gets closer and closer to the conditional probability $P(b\,|\,x_1\ldots x_n)$ as n grows – closer for example in a mean square sense (and then also with P-probability 1). This is better than any classical predictive strategy can do. More explicitly, the value

$$S_n = \sum_{x:|x|=n-1} \sum_{b\in\Sigma} P(x)(M(b\,|\,x) - P(b\,|\,x))^2$$

is the expected error of the squared probability of the nth prediction if we use the universal M instead of the unknown P. Solomonoff showed $\sum_{n=1}^{\infty} S_n < \infty$. (The bound is essentially the complexity $K(P)$, of P, so it is relatively small for simple distributions P. There is no bound when P is not even computable.) Hence the expected squared error can be said to degrade faster then $1/n$ (provided the expectation is "smooth").

The set of *all* computable distributions is very wide. Consider for example a sequence $x_1x_2\ldots$ whose even-numbered binary digits are those of π, while its odd-numbered digits are random. Solomonoff's formula will converge to 1/2 on the odd-numbered digits. On the even-numbered digits, it will get closer and closer to 1 if b equals the corresponding digit of π, and to 0 if it does not. By now, several alternative theorems, and amplifications on this convergence property have appeared: see for example [7,5].

The proof relies just on the fact that $M(x)$ dominates all computable measures (even all lower semicomputable semimeasures). It generalizes therefore to any family of measures that has a dominating measure – in particular, to any countable family of measures.

Despite the attractiveness of the formula, its incorporation of such a number of classical principles, and the nice form of the theorem, it is still susceptible to a justified criticism: the formula is in a different category from the sources that it predicts: those sources are computable, while the formula is not ($M(xy)/M(x)$ is the ratio of two lower semicomputable functions). But as mentioned above, the restrictions on the source and on the predictor cannot be expected to be the same, and at least Solomonoff's formula is brimming with philosophical significance.

The topic has spawned an elaborate theory of prediction in both static and reactive unknown environments, based on universal distributions with arbitrary loss bounds (rather than just the logarithmic loss) using extensions and variations of the proof method, inspiring information theorists such as Thomas M. Cover [4]. An example is the book by Marcus Hutter [5]. A related direction on prediction and Kolmogorov complexity, using various loss bounds, going by the name of "predictive complexity", in a time-limited setting, was introduced by Vladimir G. Vovk (see [26] and later works).

We noted that Solomonoff normalized his universal apriori distributions, in order to turn them into regular probability distributions. These normalizations make the theory less elegant since they take away the lower semicomputability property: however, Solomonoff never gave them up. And there is indeed no strong argument for the semicomputability of $M(x)$ in the context of prediction. In about 1992, Robert M. Solovay proved that every normalization of the universal a priori semimeasure to a measure would change the relative probabilities of extensions by more than a constant (even incomputably large) factor. In a recent paper with a clever and appealing proof, Solomonoff [25] proved that if we predict a computable measure with a the universal a priori semimeasure normalized according to his prescription, then the bad changes a la Solovay happen only

with expectation going fast to 0 with growing length of the predicted sequence.

6. Universal Search

It was not until 1978, that Ray Solomonoff started to pay attention to the emerging field of computational complexity theory. In that year, Leonid Levin arrived in Boston, and they became friends. Levin had discovered NP problems around 1970, independently from Stephen Cook, and had shown the completeness of a small number of NP problems (independently of Richard Karp). For our present purpose, an NP problem is best viewed as a *search problem*. It is defined with the help of a *verification predicate* $V(x, w)$, where x is the *instance*, w is a potential *witness*, and $V(x, w)$ is true if and only if the witness is accepted. We can assume that $V(x, w)$ is computable in time linear in the size $|x|$ of the instance x (in an appropriate computation model, see later). The problem is to decide for a given instance x whether there is any witness w, and if yes, to find one. As an example, consider the problem of finding a description of length l that computes a given string x within time t on some fixed machine U. Let $x = U^t(p)$ mean that machine U computes x in time t from program p. The instance of the problem could be the string 0^l10^t1x, and the verifier $V(0^l10^t1x, p)$ would just check whether $|p| \leq l$ and $U^t(p) = x$.

Levin's paper [6] announces also a theorem that has no counterpart in the works of Cook and Karp: the existence of an algorithm that finds a witness to an NP-complete problem in time optimal to within a multiplicative constant. Theoretically, this result is quite interesting: from then on, one could say that the question has not been *how* to solve any NP problem efficiently, only *what* is the complexity of Levin's algorithm. If there is a theorem that it works in time $g(|x|)$, then of course also the problem of whether there is any witness at all becomes decidable in time $g(|x|)$.

Levin's paper gave no proof for this theorem (a proof can be found now, for example, in [7]). There is a natural, approximate idea of the proof, though. What is special about an NP problem is that once a potential witness is guessed, it is always possible to check it efficiently. Therefore it does not harm much (theoretically, that is as long as we are willing to tolerate multiplicative constants) a good solution algorithm $A(x)$ if we mix it with some other ones that just make wild guesses. Let $\rho_1, \rho_2, \ldots$ be any computable sequence of positive numbers with $\sum_i \rho_i \leq 1$. We could list *all* possible algorithms $A_1, A_2, \ldots$, in some order, and run them *simultaneously*, making a step of algorithm A_i in a fraction ρ_i of the time. At any time, if some algorithm A_i proposes a witness we check it. In this way, if any algorithm A_i finds witnesses in time $g(|x|)$ then the universal algorithm finds it in time $\rho_i^{-1}g(|x|)$: this is what is meant by optimality within a multiplicative constant.

In order to actually achieve the multiplicative constant in his theorem, Levin indicated that the machine model U has to be of a "random access" type: more precisely, of a type introduced by Kolmogorov and Uspensky and related to the "pointer machine" of Schönhage. He also introduced a variant of description complexity $Kt(w) = \min_{t,\, z:U^t(z)=w} |z| + \log t$ in which a penalty of size $\log t$ is built in for the running time t of the program z outputting the sequence w on the universal machine U. A more careful implementation of Levin's algorithm (like the one given later by Solomonoff) tries the candidate witnesses w essentially as ordered by their complexity $Kt(w)$.

Up to now, Levin's optimal algorithm has not received much attention in the computational complexity literature. In its present form, it does not seem practical, since the multiplicative constant ρ_z^{-1} is exponential in the length of the program z. (For time bounds provable in a reasonable sense, Hutter reduced the multiplicative constant to 5, but with a tremendous additive constant [7]. His optimal algorithm depends on the formal system in which the upper bounds are proved.) But Solomonoff appreciated it greatly, since in computing approximations to his apriori probability, this seems still the best that is available. He gave detailed implementations of the optimal search (giving probably the first written proof of Levin's theorem), in its application to computing algorithmic probability [19,21]. These did not result in new theorems, but then Solomonoff had always been more interested in practical learning algorithms. In later projects (for example [22]) aimed at practical prediction, he defines as the *conceptual jump size* CJS of the program z the quantity t_z/p_z, where p_z is some approximation to the apriori probability of z, and t_z is its running time. The logarithm of the conceptual jump size and Levin's $Kt(w)$ are clearly related.

7. Training Sequences

Solomonoff continued to believe in the existence of a learning algorithm that one should find. He considered the approach used for example in practical speech recognition misguided: the algorithm there may have as many as 2000 tunable real number parameters. In the 1990s, he started a company to predict stock performance on a scientific basis provided by his theories. Eventually, he dropped the venture claiming that "convergence was not fast enough."

In a number of reports: [13, 15, 20, 22, 9, 24], universal search as described above is only a starting point for an array of approaches, that did not lead to new theorems, but were no less dear to Ray's heart for that. What we called "program" above can alternatively be called a "problem solving technique", or a "concept". This part of Ray's work was central for him; but the authors of the present article are closer to mathematics than to the experimental culture of artificial intelligence, therefore the evaluation poses challenges for them. We hope that the AI community will perform a less superficial review of this part of the oevre than what we can offer here.

Learning proceeds in stages, where each stage includes universal search. The conceptual jump size CJS introduced above (see [9]) continues to play a central role. Now, "probability" is used in the sense of the probability assigned by the best probabilistic model we can find in the available time for the given data. There is also an update process introducing more and more complex concepts. The concepts found useful on one stage are promoted to the status of primitives of a new language for the next stage, allowing to form more complex composite concepts (and goals). They are combined in various ways, assigning preliminarily just product probability to the composite concept. If a composite concept proves applicable with a probability beyond this initial value, it will be turned it into a new building block (with a corresponding larger probability). In this way, one hopes to alleviate the problem of excessively large multiplicative constants of universal search (see [21]).

Ray did not limit inductive inference to a model where a learner is presented a stream of experimental results. He realized that in practice, a lot of learning happens in a much more controlled situation, where there is a "teacher" (or several). Now, *supervised learning* is a well-studied set of models: in this, a teacher

provides answers to some set of questions that the learner can ask. In Solomonoff's model, the teacher also *orders* the questions in increasing conceptual jump size, facilitating thereby the above concept-building process. Already the report [13] sketches a system designed to recognize more and more complex patterns, as it is being fed a sequence of examples of gradually increasing complexity.[1] Ray spent many years working out some examples in which a learning algorithm interacts with a training sequence. The examples were of the type of learning a simple language, mainly the language of arithmetic expressions. By now, there are systems in AI experimenting with learning based on universal optimal search: see Schmidhuber in [12] and other works.

We are not aware of any *theoretical* study that distinguishes the kind of knowledge that the teacher can transmit directly from the one that the student must relearn individually, and for which the teacher can only guide: order problems by complexity, and check the student answers. The teacher may indeed be in conscious possession of a network of concepts and algorithms, along with estimates of their "conceptual jump size", and we should assume that she communicates to the student directly everything she can. (The arithmetic algorithms, Ray's main example, can certainly be fed into a machine without need for learning.) But it appears that in typical realistic learning, the directly, symbolically transferable material is only a very incomplete projection of the mental models that every pupil needs to build for himself.

References

[1] Rudolf Carnap. *Logical Foundations of Probability*. University of Chicago Press, 1950.

[2] Gregory J. Chaitin. On the length of programs for computing binary sequences, II. *Journal of the ACM*, 16:145–159, 1969.

[3] Noam Chomsky. Three models for the description of language. *IRE Trans. Inform. Theory*, 2(3):113–124, September 1956.

[4] Thomas M. Cover. Universal gambling schemes and the complexity measures of Kolmogorov and Chaitin. In J. K. Skwirzynski, editor, *The Impact of Processing Techniques on Communication*, pages 23–33. Martinus Nijhoff, 1985. Stanford University Statistics Department Technical Report # 12, 1974.

[5] Marcus Hutter. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer-Verlag, Berlin, 2005.

[6] Leonid A. Levin. Universal sequential search *problems. Problems of Inform. Transm.*, 9(3):255–256, 1973.

[7] Ming Li and Paul M. B. Vitányi. *Introduction to Kolmogorov Complexity and its Applications (Third edition)*. Springer Verlag, New York, 2008.

[8] Marvin L. Minsky. Problems of formulation for artificial intelligence. In R. E. Bellman, editor, *Proceedings of the Fourteenth Symposium in Applied Mathematics*, pages 35–45, New York, 1962. American Mathematical Society.

[9] Wolfgang Paul and Raymond J. Solomonoff. Autonomous theory building systems. In P. Bock, M. Loew, and M. Richter, editors, *Neural Networks and Adaptive Learning*, pages 1–13, Schloss Reisenburg, 1990.

[10] Jorma J. Rissanen. A universal prior for integers and estimation by minimal description length. *Annals of Statistics*, 11(2):416–431, 1983.

[11] Jorma J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, London, U.K., 1989.

[12] Jürgen Schmidhuber. Optimal ordered problem solver. *Machine Learning*, 54:211–254, 2004.

[13] Raymond J. Solomonoff. An inductive inference machine. In *IRE Convention Record, Section on Information Theory*, pages 56–62, New York, 1957. Author's institution: Technical Research Group, New York 3, N.Y.

[14] Raymond J. Solomonoff. A preliminary report on a general theory of inductive inference. Technical report ZTB-138, Zator Company, Cambridge, MA, 1960.

[15] Raymond J. Solomonoff. Training sequences for mechanized induction. In M. Yovits, editor, *Self-organizing systems*, 1961.

[16] Raymond J. Solomonoff. A formal theory of inductive inference I. *Information and Control*, 7:1–22, 1964.

[17] Raymond J. Solomonoff. A formal theory of inductive inference II. *Information and Control*, 7:225–254, 1964.

[18] Raymond J. Solomonoff. Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Transactions on Information Theory*, IT-24(4):422–432, July 1978.

[19] Raymond J. Solomonoff. Optimum sequential search. Technical report, Oxbridge Research, Cambridge, MA, 1984.

[20] Raymond J. Solomonoff. Perfect training sequences and the costs of corruption – a progress report on inductive inference research. Technical report, Oxbridge Research, Cambridge, MA, 1984.

[21] Raymond J. Solomonoff. The application of algorithmic probability to problems in artificial intelligence. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, Advances in Cognitive Science, AAAS Selected Symposia, pages 473–491, North-Holland, 1986. Elsevier.

[22] Raymond J. Solomonoff. A system for incremental learning based on algorithmic probability. In *Proceedings of the Sixth Israeli Conference on Artificial Intelligence, Computer Vision and Pattern Recognition*, pages 515–527, Tel Aviv, 1989.

[23] Raymond J. Solomonoff. The discovery of algorithmic probability. *Journal of Computer System Sciences*, 55(1):73–88, 1997.

[24] Raymond J. Solomonoff. Progress in incremental machine learning. Technical Report 03-16, IDSIA, Lugano, Switzerland, 2003. Revision 2.0. Given at NIPS Workshop on Universal Learning Algorithms and Optimal Search, Dec. 14, 2002, Whistler, B.C., Canada.

[25] Raymond J. Solomonoff. The probability of "undefined" (nonconverging) output in generating the universal probability distribution. *Information Processing Letters*, 106(6):238–246, 2008.

[26] Vladimir G. Vovk. Prediction of stochastic sequences. *Problems of Information Transmission*, 25:285–296, 1989.

[27] Alexander K. Zvonkin and Leonid A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25(6):83–124, 1970.

[1] Marvin Minsky considers that the practical potential of the pattern recognition algorithms in this work of Ray still has not received the attention it deserves.

The Historian's Column

Anthony Ephremides

As the time of the next International Symposium on Information Theory (ISIT) in St. Petersburg is fast approaching, and as I am heavily involved in its organization, I could not help reflecting upon the history of the relations between our Society and the Russian community of Information Theory. It has been a long and turbulent history and I have reminisced about it repeatedly over the years in this column. It has been, of course, seriously influenced and affected by the political realities over the last sixty years and it has been fascinating, and stressful but, ultimately, its evolution has been determined by the intense and deep interest of both sides in the field of Information Theory.

The legacy of Kolmogorov's thinking has dominated the development of the field in Russia. There have been legions of outstanding contributions by Russian scientists (the likes of Varshamov, Tsybakov, Ziablov, Zigangirov, Koshelev, Khabatiansky, Yaglom, Gelfand, and many others). But there have been two giants (Dobrushin and Pinsker) who have stood out and cast a long shadow on the field. Their scientific work has been complemented also by their boldness and independence during the dark Soviet times. It is in part a tribute to their pride and courage that we finally hold the coveted ISIT in Russia.

The first few (and legendary) contacts between the two communities started in the 1960's and there is a great deal of lore that surrounds them. From the involvement of alleged KGB agents to machinations inside the regime's apparatus, these first contacts are replete with (almost) romantic strife for breaking the barriers that kept the scientists of both sides apart. The 1973 Symposium on Information Theory that was organized by the Russians in Tallinn, Estonia, just before the IEEE ISIT in Ashkelon, Israel, drew heavy participation by IEEE Information Theorists (mostly from the United States) and marked the beginning of a dramatic series of events of hit-and-miss contacts that eventually culminated with the upcoming ISIT in St. Petersburg. I still have vividly in my mind the picture of Kamil Zigangirov riding the trolley-bus with us in Tallinn. I still hear the plea of an ordinary citizen outside our hotel to buy my "Polaroid" sunglasses. And I still recall the fascination of our Russian colleagues with David Middleton who stood larger than life as he filled slide after slide with long equations and multiple integrals as he was expanding on the physics of Noise and its relationship to Information Theory.

The Tallinn encounter was quickly followed by the famous Moscow workshop in December 1975. This is when I had the honor and pleasure of meeting Dobrushin and Pinsker up close and having discussions with them. This is when the drama of getting a visa for Adrian Segall unfolded with almost cinematographic complexity and unpredictability. This is when we experienced the warm Russian hospitality alongside fabulous buffets of caviar, vodka, and untold numbers (and quantities) of libations and toasts. This is also when we experienced glimpses of the regime's brutality. I still remember Bassalygo's efforts to procure tickets for performances at the Bolshoi. And I still remember the visit to Boris Godunov's grave in Zagorsk, the impressive and forbidding silhouette of the hotel Ukraina, the snow-packed streets, and the efforts of Lee Davisson to buy a fur hat (while Koshelev was offering to me his own as a present!).

Years of spotty contacts followed, including the tumultuous organization of a workshop in New York State, The role of our Swedish colleagues (and, in particular, Rolf Johannesson) in keeping the contacts alive has been invaluable. The organization of the biannual Swedish-USSR workshops that alternated between Sweden and the USSR kept the flame alive and provided plenty of opportunity for interaction and even joint work. The unforgettable first such workshop in 1985 took place in Graenna. I also attended the one in Gotland in 1989 and the one in Moscow in 1991. This last one took place as the deep transformations inside Russia were already underway. I recall the flowing torrents of Limonskaya Vodka and the valiant efforts of Verdu and myself to find tickets to the Bolshoi from scalpers in the streets of Moscow. And I simply can never forget the gracious hosting of a superb dinner at a Georgian restaurant on the ground floor of an apartment building by Mark Pinsker or another fabulous luncheon at a new (at the time) restaurant hosted by Ilya Dumer. And this is when I first met my "future" colleague Sasha Barg.

A lot has happened since then. Some of the drama's protagonists have passed away. Others emigrated to the United States and Europe. Some have remained in their old posts at the Institute for Problems for Information Transmission (known as IPPI) and have provided valuable links of continuity. When I see Nikita Vvedenskaya who is still in the trenches at IPPI, I feel that time has stood still as she, along with others, remains a stalwart pillar of presence of Information Theory in this historic Institute.

In a few months a new chapter will open in St. Petersburg. The rough edges have been smoothened, the difficulties have been forgotten, and the painful memories have faded. What remains is a feeling of respect, mutual understanding, and commitment to the field of Information Theory. As we gather in the beautiful Imperial city with its splendid palaces, canals, theaters, and tons of culture and History, we should also take note of the fact that the general co-chair of the Symposium is none other than Vladimir (Volodja) Blinovsky, who is Mark Pinsker's son. What a befitting tribute to his father's immense contributions and to the spirit of continuity and cooperation in the global evolution of our field. Next to the shores of the Baltic Sea in the beautifully redecorated Hotel Park Inn (previously known as Pribaltiskaya) we will be looking East and West as the world of Information Theory will convene to celebrate and worship a timeless field that keeps marking History every year for over six decades.

Let me take this opportunity to invite you and urge you to come to St. Petersburg and let me add my welcome as General Co-Chair. We owe a great deal of gratitude to many colleagues in the vast Russian Federation for their perseverance, dedication, and hospitality, as well as their voluminous and lasting contributions to the field.

Teaching Lossless Data Compression

Sergio Verdú

Most courses on information theory, particularly those patterned after Cover and Thomas [1], cover the algorithmic side of lossless data compression, most notably Huffman, arithmetic and Lempel-Ziv codes. I like to do it at the advanced-undergraduate level and it is great fun to teach. However, how we describe and analyze those algorithms is not the purpose of this column. Instead, expanding on one of the items in my Shannon Lecture, I will discuss the teaching of the fundamental limits of lossless data compression.

Although there are other source coding setups, such as (Tunstall) variable-to-fixed coding, the conventional canon found in most textbooks, and taught in most information theory courses, deals with two separate source coding paradigms:

1) Variable-length symbol-by-symbol lossless compression;

2) Fixed-length (or fixed-to-fixed) almost-lossless compression.

The centerpieces of the fundamental limits of symbol-by-symbol lossless compression are a converse result that states that the code-lengths of any uniquely decodable code must satisfy the Kraft inequality, and an achievability result that guarantees the existence of a prefix code with any set of code-lengths that satisfies the Kraft inequality. Therefore, non-prefix codes do not offer the prospect of increased efficiency. The converse result readily leads to the conclusion that the average length of a variable-length symbol-by-symbol uniquely decodable code cannot be smaller than the source entropy, a result commonly, and wrongly, attributed to Shannon. A formula for the minimal average length is not known, but an algorithm to compute it is indeed known (provided the alphabet is finite) since it is achieved by the Huffman code. Moreover, since the achievability result guarantees that a code exists whose code-lengths are equal to the log-reciprocal-probabilities (aka "ideal code-lengths") rounded-up, we can upper bound the minimal average length of a binary symbol-by-symbol lossless code by the entropy plus one bit.

In contrast to the analysis of variable-length codes, the conventional analysis of fixed-length almost-lossless compression is asymptotic in nature and does not just deal with averages: instead, by invoking Shannon's notion of typicality and the law of large numbers, it shows that as long as the coding rate exceeds the entropy and the source is memoryless, vanishing error probability is achievable. It is also a good idea for the teacher to show the converse of this statement. See, for example, the proof of [2, Theorem 1.1.1].

Some natural questions that the student may ask:

- Arithmetic codes and Lempel-Ziv codes are strictly lossless and have variable length, but they are not symbol-by-symbol codes. Why did we restrict the analysis of variable-length codes to symbol-by-symbol codes?

- Are fixed-length almost-lossless codes used in practice?

- Why did we limit our analysis of the fundamental limits to memoryless sources? Aren't sources in the real world redundant because of their memory?

- If symbol-by-symbol codes cannot exploit memory, in what applications are they used?

- Attaining good average length at the expense of large variance, or large probability of exceeding a given threshold, may not be desirable. For variable-length lossless codes, is there anything beyond the minimal average length that we can analyze?

- There are about 2^{922} possible tweets (twitter messages are limited to no more than 140 characters). How many bits are required to compress 95% of the time?

First, let us address the issue of memory. Yes, if we do not exploit the memory in discrete sources such as text and digital photographs we do a really lousy job at compressing. But compression of redundant memoryless sources is much more common than one may think, primarily because through various reversible transformations such as run-length encoding (e.g. fax), linear transforms (e.g. JPEG) and the Burrows-Wheeler transform (e.g. bzip) the redundancy in memory is shifted to non-equiprobability of the first-order distribution. In those settings, symbol-by-symbol codes, and in particular Huffman codes find plenty of applications. But as far as the fundamental limits, I feel remiss if I end my coverage with memoryless sources. After all, Shannon gave the the asymptotic equipartition property [3, Theorem 3] not just for memoryless sources but for Markov sources. Pedagogically, after treating the memoryless case, it is not too difficult to prove the asymptotic equipartition property for stationary ergodic sources using Markov approximation: either by McMillan's original approach (see Gallager [4]) or the Algoet-Cover sandwich method [1].

Next, let us revisit the traditional restriction of the analysis of strictly lossless codes to symbol-by-symbol codes. It is not as severe as it sounds because we can always think of super-symbols, each encompassing m consecutive symbols of the original source. Then, the minimal average length is the entropy of m-words plus at most 1. This trick is more useful conceptually than algorithmically, since the computational complexity skyrockets with m. Algorithmically, a preferable way to deal with memory is to abandon the symbol-by-symbol paradigm altogether and use arithmetic coding or Lempel-Ziv codes. As far as the analysis of the ultimate efficiency, we can go one step further and dispense with super-symbols altogether by viewing the whole file to be compressed as one symbol. But here's the thing: we can get better efficiency than what the conventional analysis taught in textbooks predicts! To fix ideas, consider the twitter question. Suppose we knew the probabilities of all possible tweets; what would be the best lossless data compressor? Would it be a Huffman code for a source with an alphabet of 2^{922} "symbols"? The average length of the Huffman-coded version will be equal to the entropy of the tweet distribution plus at

most 1 bit. But we can do better: list all the tweets in decreasing probabilities and encode them with the binary strings:

$$\{\varnothing, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \ \ldots\}$$

This is the best code not just in the sense of minimizing the average, but in the much stronger sense of maximizing the cdf of the length at every point. Therefore, the abscissa at which the cdf of the length reaches 0.95 is the answer to the foregoing question. How much better than the Huffman code is the optimal variable-length code? Its average length is in fact *upper* bounded by the entropy (the ideal code-lengths are not so ideal after all) and admits the expression

$$\sum_{k=1}^{\infty} \mathbb{P}[X \geq 2^k]$$

where the source realizations have been relabeled so that the integer $X = \ell$ is the ℓth most probable outcome. According to recent results [5], for memoryless sources of length n and entropy H, the minimal average length of a fixed-to-variable code behaves as $nH - (1/2)\log n + O(1)$, rather than $nH + O(1)$ for a super-symbol Huffman code or an arithmetic code. (The symbol-by-symbol Huffman code would be even worse: $n(H + g)$ for $0 \leq g < 1$.) Wait. Wasn't the entropy a sacrosanct limit that no variable-length code could beat? That is true for symbol-by-symbol uniquely decodable codes; however, once the super-symbol becomes the whole file to compress we should realize that the prefix condition (or the uniquely decodable condition) is superfluous. But then, how do we know where the compressed file ends? Just think of a file stored in a hard disk. Does it satisfy the prefix condition (in some humongous tree) so it can tell us by itself where its "ends"? No. In fact, it will probably be stored in many chunks identified by a directory of pointers. Admittedly, the optimal non-prefix variable-length code is easier said than done unless the source model is very simple; moreover, if the file to compress is long enough, then the penalty for imposing the extraneous prefix condition on the encoded file gets amortized. Still, in the short run and in particular if we also pay attention to variance in addition to average, there are efficiencies to reap.

And finally, let us address why we study almost-lossless n-to-k codes. In those codes, one of the binary output k-strings signals that the input realization cannot be handled by the compressor. Although all practical data compressors are strictly lossless and therefore variable-length, we can easily turn an almost-lossless code into a lossless code by, for example, substituting the special output k-string by the input string. Conversely, we can easily turn a variable-length code into an almost-lossless fixed-length code. Another important motivation for teaching them is that not only do almost-lossless source codes provide the reason to study the asymptotic equipartition property, but they are the indispensable gateway to many other results in information theory, such as the source-channel separation theorem, the Slepian-Wolf theorem and the random binning proof method, ubiquitous in multiuser information theory. Furthermore, linear channel coding is the dual problem to linear fixed-to-fixed source coding; in fact, excellent data compressors can be built based on modern sparse-graph codes [6]. These are all good motivations to teach almost lossless compression, but in fact I would argue that the main reason is a very simple fact (proof left to the reader) that has gone unrecognized because of the traditional fixation with Kraft-inequality compliant variable-length codes: Regardless of the source, the minimal probability of error of an n-to-k code is equal to the probability that the length of the optimal variable-length code for $X_1, \ldots, X_n$ is greater than or equal to k. Therefore, the analysis (asymptotic or not) of the fundamental limits of fixed-to-fixed data compression is equivalent to the analysis of the fundamental limits of fixed-to-variable data compression.

Sergio Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, Email: verdu@princeton.edu

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

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

[3] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul.–Oct. 1948.

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

[5] W. Szpankowski and S. Verdú, "Minimum Expected Length of Fixed-to-Variable Lossless Compression without Prefix Constraints," *IEEE Trans. on Information Theory*, to appear.

[6] G. Caire, S. Shamai, A. Shokrollahi and S. Verdú, "Fountain Codes for Lossless Data Compression," *Algebraic Coding Theory and Information Theory*, A. Ashikhmin, A. Barg, I. Duursma, Eds., DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 68, pp. 1–20, American Mathematical Society, 2006.

Can Structure Beat Random? – The Story of Lattice Codes

Plenary talk presented at the 2010 IEEE International Symposium on Information Theory, Austin, Texas.

Ram Zamir,
EE - Systems, Tel Aviv University, ISRAEL
E-mail: zamir@eng.tau.ac.il

Abstract

At the birth of information theory, Shannon surprised the communication world with the concept of random coding, which he used for proving the ultimate limits of his theory. This powerful tool is, however, non-constructive. In Shannon's words: "An attempt to obtain a good approximation to ideal coding by following the method of the proof is generally impractical... related to the difficulty of giving an explicit construction for a good approximation to a random sequence." A practical substitute to random coding are structured codes (one example of which - the Hamming code – appeared already in Shannon's paper from 1948). Multiterminal information theory provides us now with a new surprise: for some distributed coding problems structured codes seem to be better than random codes! This summary of my ISIT 2010 plenary talk discusses how lattice codes are used in Gaussian multiterminal settings, and the intuition they provide for the question in the title.

I. Motivation

It is not hard to detect the few differences between the two faces in Fig. 1. Once detected, it is also not too hard to describe them with just a few words. But would a few words be sufficient if the two faces were described by two *separate* observers?

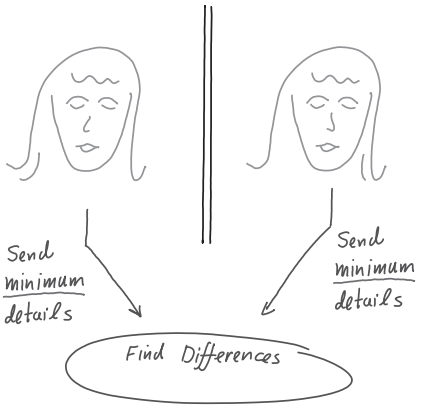


Fig. 1 Find (and communicate) the differences.

An information-theoretic analogue of this question is the "two help one"problem of Fig. 2, which was proposed in a seminal paper from the late 70's by Körner and Marton [25]. They showed that if one wishes to reconstruct the modulo-two sum of two correlated binary sources from their independent encodings, then linear coding seems to be better than random coding.

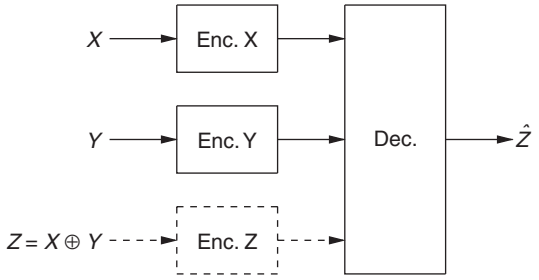


Fig. 2 The Körner-Marton configuration.

Specifically, the Körner-Marton (KM) setup consists of a binary doubly symmetric source (X, Y), and an "error" variable $Z = X \oplus Y$ indicating when X and Y are different, i.e., $\Pr(Z = 1) = \Pr(X \neq Y) = \theta$. The goal is to encode the sources X and Y separately such that Z can be reconstructed losslessly. If coordination between the encoders were allowed, then they could compute the XOR sequence $Z_1, \ldots, Z_n$ and encode it at a rate of $H(Z)$. Via a "genie aided" argument, Körner and Marton showed that in the uncoordinated case, the sum rate required is at least

$$R_x + R_y \geq 2H(Z). \tag{1}$$

Furthermore, this sum rate can be achieved by a *linear code*: each encoder transmits the syndrome of the observed source relative to a good linear binary code for a BSC with crossover probability θ.

A common technique in proving direct coding theorems in information theory is the use of a *random* code, induced by some *single-letter* formula. In an attempt to find such a formula for the problem in Fig. 2, Körner and Marton examined a "natural" extension for the solution of the "one help one" problem [1], [46]; the resulting achievable rates satisfy, [25, appendix]

$$R_x + R_y \geq H(X, Y). \tag{2}$$

These rates correspond to Slepian-Wolf encoding of X and Y [9],[1] and are clearly strictly contained in (1) (since $H(X, Y) = 1 + H(Z)$ in (2) is greater than $2H(Z)$ for $\theta \neq 1/2$). Thus, the "natural" random binning solution for the "two help one" problem is suboptimal, and inferior to structured (linear) coding.

Does this mean that any random coding scheme (i.e., single-letter solution) would be suboptimal for the "two help one" problem? Instead of dealing with that directly, we turn to structured (lattice) coding in the Euclidean space, with the hope to get further intuition about this issue in multi-terminal Gaussian setups.

II. Why Lattices?

Lattices form effective arrangements of points in space for various geometric and coding problems, e.g., sphere covering and packing, quantization, and signaling for the additive white Gaussian-noise

[1]It can also be derived from the Berger-Tung achievable region [3] for distributed lossy coding of X and Y with one reconstruction $\hat{Z}$ under the distortion measure $d(X, Y, \hat{Z}) \triangleq X \oplus Y \oplus \hat{Z}$.

(AWGN) channel [6], [16], [10]. The best lattice for each problem may be different. Nevertheless, as the dimension goes to infinity, there exist lattices which tend to be "perfect" for all problems.

In the context of this talk, lattices serve as a bridge from the low dimensions of common modulation techniques (PCM, PAM, QAM) to the large dimensions of coded modulation schemes, or to the infinite dimension of Shannon's theory. They also provide an "algebraic" binning scheme for some Gaussian side information problems [52], [13]. Moreover, recent developments in the area of Gaussian network information theory, [35], [36], [26], [40], [41], [38], indicate that lattices are sometimes even better than their random coding counterparts!

III. Lattice Definitions and Figures of Merit

An n-dimensional lattice Λ is defined by a set of n basis vectors $g_1, \ldots, g_n$ in $\mathbb{R}^n$. The lattice Λ is composed of all integer combinations of the basis vectors, i.e.,

$$\Lambda = \{\lambda = G \cdot \mathbf{i} : \mathbf{i} \in \mathbb{Z}^n\}, \tag{3}$$

where $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, and the $n \times n$ *generator matrix* G is given by $G = [g_1 \mid g_2 \mid \ldots \mid g_n]$. When G is the unit matrix, we obtain the integer lattice $\mathbb{Z}^n$. Thus, Λ in (3) can be written also as $G\mathbb{Z}^n$. Note that the zero vector is always a lattice point, and that G is not unique for a given Λ. See [6] as an excellent background.

A few important notions are associated with a lattice. The nearest neighbor quantizer $Q_\Lambda(\cdot)$ is defined by

$$Q_\Lambda(\mathbf{x}) = \lambda \in \Lambda \quad \text{if } \|\mathbf{x} - \lambda\| \leq \|\mathbf{x} - \lambda'\| \ \forall\, \lambda' \in \Lambda \tag{4}$$

where $\|\cdot\|$ denotes Euclidean norm, and ties are broken in a systematic manner. The fundamental Voronoi region of Λ is the set of points in $\mathbb{R}^n$ closest to the zero codeword, i.e., $\mathcal{V}_0 = \{\mathbf{x} : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}$. The Voronoi region associated with each $\lambda \in \Lambda$ is the set of points $\mathbf{x}$ such that $Q_\Lambda(\mathbf{x}) = \lambda$, and is given by a shift of $\mathcal{V}_0$ by λ.

Other fundamental regions $\mathcal{P}_0$ exist which generate a lattice partition of the form $\{\lambda + \mathcal{P}_0\}_{\lambda \in \Lambda}$, and a corresponding lattice quantizer

$$Q_{\Lambda, \mathcal{P}_0}(\mathbf{x}) = \lambda \quad \text{if } \mathbf{x} \in (\lambda + \mathcal{P}_0) \tag{5}$$

For example, the fundamental parallelotope $\{G\alpha : 0 \leq \alpha_i < 1,\ i = 1 \ldots n\}$ amounts to transforming the unit cube (the fundamental region of $\mathbb{Z}^n$) by the generator matrix G. Nevertheless, the volume of all fundamental regions of Λ is the same, and is given by $|\det(G)| \triangleq V_\Lambda$.

The modulo-Λ operation w.r.t. the lattice Λ and some assumed fundamental region $\mathcal{P}_0$ in (5) is defined as

$$\mathbf{x} \mod_{\mathcal{P}_0} \Lambda = \mathbf{x} - Q_{\Lambda, \mathcal{P}_0}(\mathbf{x}) \tag{6}$$

which is also the quantization error of $\mathbf{x}$ with respect to Λ.

The two most well studied figures of merit of a lattice are its packing radius and covering radius, illustrated in Fig. 3. Here we will focus on two other figures of merit which have more of an engineering flavor: the *normalized second moment*, which is a measure of goodness for quantization, and the *volume to noise ratio*, which is a measure of goodness for AWGN channel coding.

Mean-squared error (MSE) quantization: The second moment σ^2_Λ of a lattice is defined as the second moment per dimension of a uniform distribution over the fundamental Voronoi region $\mathcal{V}_0$,

$$\sigma^2_\Lambda = \frac{1}{V_\Lambda} \cdot \frac{1}{n} \int_{\mathcal{V}_0} \|\mathbf{x}\|^2 d\mathbf{x}. \tag{7}$$

A dimensionless figure of merit of a lattice quantizer with respect to the MSE distortion measure is the normalized second moment (NSM)

$$G(\Lambda) = \frac{\sigma^2_\Lambda}{V_\Lambda^{2/n}}. \tag{8}$$

The minimum possible value of $G(\Lambda_n)$ over all lattices in $\mathbb{R}^n$ is denoted by G_n. The normalized second moment of a sphere, denoted by G^*_n, approaches $1/(2\pi e)$ as the dimension n goes to infinity. The isoperimetric inequality implies that $G_n > G^*_n > 1/(2\pi e)$ for all n. We also have $G_n \leq G_1 = G(\mathbb{Z}) = 1/12$.

The operational significance of this figure of merit comes from classical results in high-resolution quantization theory. It is also useful in the context of constellation shaping, as we shall see in Sec. V. A result due to Poltyrev which appeared in [50] states that the sequence G_n achieves the sphere lower bound, i.e.,

$$\lim_{n \to \infty} G_n = \frac{1}{2\pi e}. \tag{9}$$

Another result in [50] is that the quantization noise of a lattice achieving G_n is "white", i.e., the covariance matrix of a uniform distribution over $\mathcal{V}_0$ is given by $\sigma^2_\Lambda \cdot I$, where I is the identity matrix.

Coding for the unconstrained AWGN Channel: The AWGN channel is given by the input/output relation

$$Y = X + Z \tag{10}$$

where Z is i.i.d. Gaussian noise of variance σ^2_z. We denote by $\mathbf{Z}$ an i.i.d. vector of length n of noise random variables.

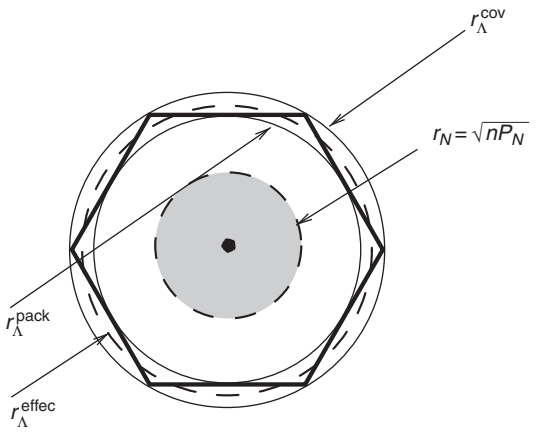


Fig. 3 The fundamental Voronoi region and its packing radius, covering radius and effective radius (radius of the sphere having the same volume). Packing and covering efficiencies are measured by the corresponding ratios.

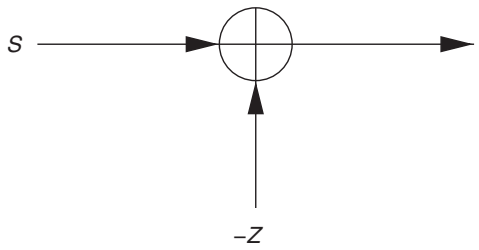


Fig. 4 Equivalent additive-noise channel of a dithered lattice quantizer. Z is independent of the input S, and uniform over the fundamental region $\mathcal{P}_0$ of Λ.

The notion of lattices which are good for AWGN coding may be defined using Poltyrev's [42] definition of capacity per unit volume of *unconstrained* channels, allowing to separate the "granular" properties of the lattice as a good channel code from the issue of shaping (to meet the power constraint). The probability of decoding error in this setup is the probability that the noise leaves the Voronoi region of the transmitted lattice point

$$P_e = \Pr\{\mathbf{Z} \notin \mathcal{V}_0\}. \tag{11}$$

The volume-to-noise ratio (VNR) of a lattice at probability of error P_e is defined as the dimensionless number

$$\mu(\Lambda, P_e) = \frac{V_\Lambda^{2/n}}{\sigma_z^2} \tag{12}$$

where σ_z^2 is such that (11) is satisfied with equality [17]. Note that for fixed P_e, the VNR is invariant to scaling of the lattice. The minimum possible value of $\mu(\Lambda, P_e)$ over all lattices in $\mathbb{R}^n$ is denoted by $\mu_n(P_e)$. The VNR of a sphere is denoted $\mu_n^*(P_e)$. Since a sphere supports the isotropic vector $\mathbf{Z}$ better than any shape of the same volume (see the *sphere bound* of [17]), we have $\mu_n(P_e) > \mu_n^*(P_e) > 2\pi e$, where the second inequality holds for all sufficiently small P_e, and $\mu_n^*(P_e) \to 2\pi e$ as $n \to \infty$, for all $P_e > 0$. It follows from Poltyrev (see also [16], [17], [21]) that the sequence of minimum possible VNRs asymptotically achieves this lower bound:

$$\lim_{n\to\infty} \mu_n(P_e) = 2\pi e, \quad \text{for all } 0 < P_e < 1. \tag{13}$$

In fact, simultaneous goodness in *both* senses (9) and (13) above is asymptotically possible.

Theorem 1. [10] *There exists a sequence Λ_n of lattices of increasing dimension n, which satisfies*

$$G(\Lambda_n) \to \frac{1}{2\pi e} \quad \text{and} \quad \mu(\Lambda_n, P_e) \to 2\pi e.$$

It is also shown in [10] that these lattices achieve the Minkowski and Rogers bounds for sphere packing and covering, and the Poltyrev exponent of the unconstrained AWGN channel.

IV. Dithered Quantization

In quantization theory (as well as in some non-linear processing systems) the term "dithering" corresponds to intentional randomization, aimed to improve the perceptual effect of the quantization, e.g. to reduce "blocking" effects in picture coding. Dithered quantization is also an effective means to guarantee a desired distortion level, independent of the source statistics.

We say that $\mathbf{U}$ is a "subtractive dither" if it is known at both the encoder and the decoder (i.e., it is a common randomness), the encoder adds it to the source vector $\mathbf{s}$ prior to the quantization, while the decoder subtracts it from the quantized value, so the overall reconstruction is $Q_\Lambda(\mathbf{s}+\mathbf{U}) - \mathbf{U}$. Addition and subtraction of $\mathbf{u}$ before and after quantization amounts to shifting the quantizer by $-\mathbf{u}$. Since the lattice quantizer $Q_\Lambda(\cdot)$ is periodic in space, a random shift $\mathbf{U}$ which is uniform over the lattice period makes the quantization error uniform as well.

Theorem 2. [50], [55] *Let $\mathbf{U}$ be uniform over the fundamental region $\mathcal{P}_0$ of the lattice quantizer (5). Then, the quantization error $Q_{\Lambda,\mathcal{P}_0}(\mathbf{s}+\mathbf{U}) - \mathbf{U} - \mathbf{s}$ is uniform over $-\mathcal{P}_0$, the reflection of $\mathcal{P}_0$, independent of the source vector $\mathbf{s}$.*[2]

Equivalently, $(\mathbf{s}+\mathbf{U}) \bmod_{\mathcal{P}_0} \Lambda$ is uniform over $\mathcal{P}_0$ for any $\mathbf{s}$, a result termed the "Crypto Lemma" by Forney [18].

As a corollary of Theorem 2 and (7), the mean-squared distortion of a Voronoi dithered quantizer (4) is equal to the lattice second moment:

$$\frac{1}{n} E\|Q_\Lambda(\mathbf{s}+\mathbf{U}) - \mathbf{U} - \mathbf{s}\|^2 = \sigma_\Lambda^2 \tag{14}$$

independent of the source vector $\mathbf{s}$.

In high-resolution quantization theory it is common to approximate the quantization process as adding (independent) noise to the source. Theorem 2 shows that for dithered quantization this model is exact at *any resolution*. See Fig. 4.

IV.A. Entropy Coded Dithered Quantization

The next theorem makes the connection to an additive-noise channel even stronger. Assume that for given source statistics, the lattice quantizer output is losslessly "entropy" coded, conditioned on the dither value. That is, each lattice point is mapped into a binary word of variable length, such that the average code length is approximately equal to the conditional entropy of the quantizer output. We call such a combination of a lattice quantizer and optimum lossless encoding an Entropy-Coded Dithered Quantizer (ECDQ).

Theorem 3. [48] *The average code length of the ECDQ, i.e., the conditional entropy of the dithered lattice quantizer, is equal to the mutual information in the equivalent additive-noise channel of Fig. 4:*

$$H(Q_\Lambda(\mathbf{S}+\mathbf{U}) \,|\, \mathbf{U}) = I(\mathbf{S}; \mathbf{S}-\mathbf{U}) \tag{15}$$

The mutual information formula above resembles the expression for Shannon's rate-distortion function [9]: $R(D) = \inf_{E\{(S-\hat{S})^2\} \le D} I(S;\hat{S})$ This formal resemblance leads to a universal bound on the loss of the ECDQ.

Theorem 4. [57], [48] *For any source $\mathbf{S}$, the redundancy of the ECDQ above the rate-distortion function under a squared error distortion measure is at most*

[2]Thm. 2 still holds if $\mathbf{U}$ is replaced by a "generalized dither", i.e., any vector $\tilde{\mathbf{U}}$ such that $(\tilde{\mathbf{U}} \bmod \Lambda)$ is uniform over $\mathcal{P}_0$ [55].

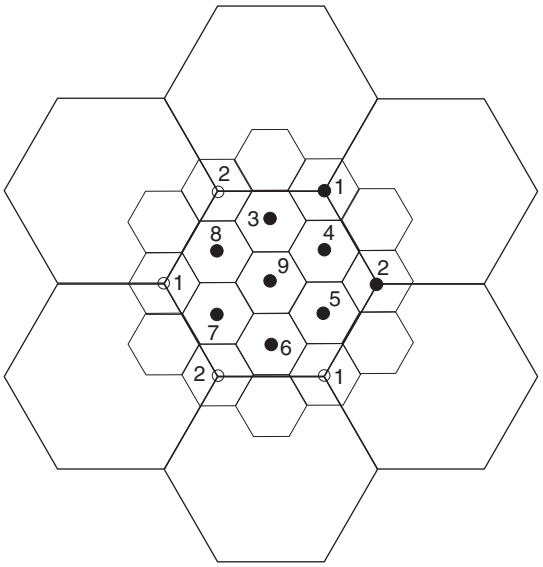


Fig. 5 Nested lattices: special case of self similar lattices.

$$H(Q_\Lambda(\mathbf{S}+\mathbf{U})\,|\,\mathbf{U}) - R(D) \le \frac{1}{2} + \frac{1}{2}\log(2\pi e G(\Lambda)) \qquad (16)$$

bits, and it is only $(1/2)\log(2\pi e G(\Lambda))$ *in the limit as D goes to zero (i.e., at high-resolution conditions).*

Divergence of dither from Gaussianity: The second term on the right hand side above can be interpreted as the divergence (or "Kullback-Leibler distance") of the dither distribution from AWGN:

$$\frac{1}{2}\log(2\pi e G(\Lambda)) = \frac{1}{n} D(\mathbf{U}\,||\,\mathbf{U}^*) \qquad (17)$$

where $\mathbf{U}^*$ is a zero-mean i.i.d. Gaussian vector with $\mathrm{Var}(U_i) = \sigma_\Lambda^2$ for all i, and where $D(\cdot\|\cdot)$ denotes divergence [9], [50]. Thus, for lattices which are good for quantization, i.e., $\lim_{n\to\infty} G(\Lambda_n) = 1/(2\pi e)$, the divergence of the dither from Gaussianity (17) goes to zero, so the equivalent channel of Fig. 4 becomes an AWGN channel.

IV.B. Filtered ECDQ

Consider the equivalent additive-noise channel model in Fig. 4. As discussed earlier, for any finite dimension the noise of optimal quantization lattices is *white* [50]. If the second order statistics of the source are also known, then we can use Wiener linear estimation principles to reduce the overall MSE in reconstructing the source $\mathbf{S}$.

If the source is white, then the Wiener filter is a simple scalar coefficient β at the output of the equivalent channel. For such a source the reconstruction becomes $\hat{\mathbf{S}} = \beta[Q_\Lambda(\mathbf{S}+\mathbf{U}) - \mathbf{U}]$, where $\beta = \sigma_S^2/(\sigma_S^2+\sigma_\Lambda^2)$, and the overall distortion $D = E\,\|\hat{\mathbf{S}} - \mathbf{S}\|^2$ decreases from σ_Λ^2 to $D = (1/\sigma_S^2 + 1/\sigma_\Lambda^2)^{-1}$. This reduction in distortion of the "post-filtered" ECDQ allows us to improve the bound of Thm. 4 in the Gaussian source case.

Theorem 5. [49] *For a Gaussian source with variance* σ_S^2*, the redundancy of the post-filtered ECDQ over the rate-distortion function* $R^*(D) = 1/2\log(\sigma_S^2/D)$ *is at most*

$$H(Q_\Lambda(\mathbf{S}^*+\mathbf{U})\,|\,\mathbf{U}) - R^*(D) \le \frac{1}{2}\log(2\pi e G(\Lambda)) \qquad (18)$$

for all *distortion levels* $0 < D \le \sigma_S^2$.

See [49] for the extension of this concept to sources with *memory* using pre/post-filters.

Note that the output scaling factor β is smaller than one for the entire distortion range $0 < D \le \sigma_S^2$. Since the reconstruction $\hat{\mathbf{S}}$ belongs to $\beta\Lambda$ (up to a shift due to the dither), it follows that the decoding lattice $\beta\Lambda$ is a "deflated" version of the encoding lattice Λ. More on the meaning of this encoding-decoding "mismatch" in the next section.

V. Voronoi Codebooks

As Information Theory shows us, coding for Gaussian sources and channels should be done using "Gaussian codebooks". That is, the codewords should be selected from a Gaussian generating distribution. The number of codewords is determined by the target rate, while the generating distribution is white, and its variance is equal to the source variance - in source coding, and to the transmitter power – in channel coding. The resulting codebook in $\mathbb{R}^n$ (n being the code dimension) has roughly uniformly distributed codewords over a *sphere*. Can we replace a Gaussian codebook by a lattice code?

In the ECDQ system discussed above, the codebook was the whole (unbounded) lattice and *not shaped* to fit the source variance. The lack of shaping is compensated for by entropy coding, which amounts to "soft" shaping: the lattice points which fall inside the typical (spherical) source region get a shorter binary representation, and dominate the coding rate, while the contribution of the points outside this region is negligible. A similar situation occurs in channel coding with *probabilistic shaping* [28], or alternatively, in unconstrained channels [42]. In fixed-length source coding or power-constrained channel coding, however, the codebook must be bounded.

In this section we describe a lattice codebook, whose codewords and shaping region both have a lattice structure. The construction is based on the notion of nested lattices, [51], [52], [11], [13], which has its roots in de Buda's spherical lattice codes [4], [30] and Forney's Voronoi constellations [14], [15], and owe its development to the search for structured binning schemes for side information problems; see the next section.

V.A. Nested Lattices

A pair of n-dimensional lattices $(\Lambda_1,\ \Lambda_2)$ is called nested if $\Lambda_2 \subset \Lambda_1$, i.e., there exists corresponding generator matrices G_1 and G_2, such that

$$G_2 = G_1 \cdot \mathbf{J}\,,$$

where $\mathbf{J}$ is an $n \times n$ integer matrix whose determinant is greater than one. We call Λ_1 the *fine lattice* and Λ_2 the *coarse lattice*. The cell volumes of Λ_1 and Λ_2 satisfy

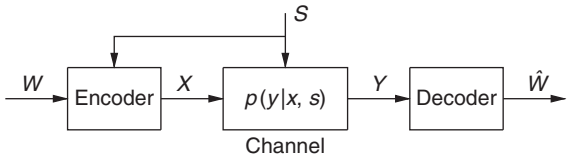


Fig. 6 A channel with side-information at the transmitter.

$$V_{\Lambda_2} = |\det(\mathbf{J})| \cdot V_{\Lambda_1}. \tag{19}$$

We call $\sqrt[n]{|\det(\mathbf{J})|} = \sqrt[n]{V_{\Lambda_2}/V_{\Lambda_1}}$ the *nesting ratio*.

Fig. 5 shows nested hexagonal lattices with $\mathbf{J} = 3 \cdot I$, where I is the 2×2 identify matrix. This is an example of the important special case of *self similar lattices*, where Λ_2 is a scaled – and possibly reflected or rotated – version of Λ_1.

For some fundamental region $\mathcal{P}_{0,2}$ of Λ_2, the points of the set

$$\Lambda_1 \bmod \Lambda_2 \triangleq \Lambda_1 \cap \mathcal{P}_{0,2} \tag{20}$$

are called the *coset leaders* of Λ_2 relative to Λ_1; for each $v \in \{\Lambda_1 \bmod \Lambda_2\}$ the shifted lattice $\Lambda_{2,v} = v + \Lambda_2$ is called a coset of Λ_2 relative to Λ_1. It follows that there are $V_{\Lambda_2}/V_{\Lambda_1} = |\det(J)|$ different cosets.

If $\mathcal{P}_{0,2}$ in (20) is the fundamental Voronoi region $\mathcal{V}_{0,2}$ of Λ_2, then we obtain a Voronoi constellation [14],[15]. In the example of Fig. 5, the Voronoi constellation consists of the bold points. A parallelepiped region $\mathcal{P}_{0,2}$ is preferable, however, if we wish to simplify coset enumeration [56].

Dithered Voronoi codebook: A dithered Voronoi codebook consists of all shifted fine lattice points $\lambda + \mathbf{u}$, for $\lambda \in \Lambda_1$, inside the Voronoi region of the coarse lattice Λ_2, i.e.,

$$(\mathbf{u} + \Lambda_1) \bmod \Lambda_2 \tag{21}$$

where the dither $\mathbf{u}$ is an arbitrary vector in $\mathbb{R}^n$ to be specified later. (For $\mathbf{u} = 0$ and $\mathcal{P}_{0,2} = \mathcal{V}_{0,2}$ this is the set of relative coset leaders in (20).) The size of this codebook is $V_{\Lambda_2}/V_{\Lambda_1}$ (independent of $\mathbf{u}$), so the associated coding rate is

$$R = \frac{1}{n}\log_2(V_{\Lambda_2}/V_{\Lambda_1})$$

bits per dimension.

Existence of good nested lattices: The existence of a sequence of good pairs of nested lattices, where one of the lattices (the fine one or the coarse one) is good for AWGN channel coding, while the other lattice is good for source coding under mean-squared distortion, is addressed in [10]. See [27] for an extension. The key to proving the existence of such lattices is to consider an appropriate *random ensemble* of lattices. An ensemble based on *generalized construction A* was defined in [32], while the Minkowski-Hlawka-Siegel ensemble is considered in [42], [21], [56].

V.B. Achieving the AWGN Channel Capacity

We now show an efficient coding scheme for the AWGN channel $Y = X + Z$ of (10) using a pair of nested lattices $\Lambda_2 \subset \Lambda_1$. In this scheme Λ_2 (the coarse lattice) is used for *shaping* while Λ_1 (the fine lattice) is used for *coding*.

Let the dither $\mathbf{U}$ be uniform over a fundamental region of Λ_2 (or a generalized dither as mentioned earlier), and let $\mathbf{v}$ be any codeword (or coset leader) in $\Lambda_1 \bmod {}_{\mathcal{P}_0}\Lambda_2$, with $\bmod \Lambda_2$ w.r.t. a "convenient" enumeration fundamental region $\mathcal{P}_0$.

To transmit the message $\mathbf{v}$, the encoder outputs

$$\mathbf{X} = (\mathbf{v} + \mathbf{U}) \bmod {}_{\mathcal{V}_0}\Lambda_2$$

with $\bmod \Lambda_2$ now performed w.r.t. the fundamental Voronoi region $\mathcal{V}_0$. By (14) we have that $E\{\|\mathbf{X}\|^2\} = \sigma^2_{\Lambda_2}$. Thus if we chose a lattice with second moment $\sigma^2_{\Lambda_2} = P$, then each codeword satisfies the power constraint (on the average with respect to the dither).

The decoder first linearly estimates the vector $\mathbf{v}$ by

$$\hat{\mathbf{Y}} = \alpha\mathbf{Y} - \mathbf{U} \tag{22}$$

(where $0 < \alpha \leq 1$ is a coefficient to be determined later). Then, it quantizes $\hat{\mathbf{Y}}$ to the nearest fine lattice point, and identifies its coset leader in the codebook. The decoded message is thus $\hat{\mathbf{V}} = Q_{\Lambda_1}(\hat{\mathbf{Y}}) \bmod {}_{\mathcal{P}_0}\Lambda_2$. This is equivalent to

$$\hat{\mathbf{V}} = \left[\alpha \cdot Q_{\Lambda_1/\alpha}\left(\mathbf{Y} - \frac{\mathbf{U}}{\alpha}\right)\right] \bmod {}_{\mathcal{P}_0}\Lambda_2, \tag{23}$$

i.e., to decoding with respect to the inflated lattice Λ_1/α. (Note the resemblance to the deflated lattice $\beta\Lambda$ in Sec. IV.B.)

The equivalent channel from the codeword $\mathbf{v}$ to the modulo estimation vector $\tilde{\mathbf{Y}} = \hat{\mathbf{Y}} \bmod {}_{\mathcal{P}_0}\Lambda_2$ is called a *modulo-lattice transformation* [11]. The distributive law of the modulo operation[3] and Thm. 2 imply:

Theorem 6. (Effective modulo- Λ additive-noise channel) [11] The channel from $\mathbf{v}$ to $\tilde{\mathbf{Y}}$ is equivalent in distribution to the modulo additive-noise channel

$$\tilde{\mathbf{Y}} = (\mathbf{v} + \mathbf{Z}_{\text{eff}}) \bmod {}_{\mathcal{P}_0}\Lambda_2$$

where the effective noise is given by

$$\mathbf{Z}_{\text{eff}} = [\alpha\mathbf{Z} + (1-\alpha)\mathbf{U}'] \bmod {}_{\mathcal{P}_0}\Lambda_2 \tag{24}$$

and where $\mathbf{U}'$ *is uniform over* $\mathcal{V}_0$ *and independent of* $\mathbf{v}$ *and* $\mathbf{Z}$.

Note that the *effective (additive) noise* $\mathbf{Z}_{\text{eff}}$ is a weighted combination of two components: AWGN and a dither component, where the latter is called "self noise" because it comes from the coarse lattice.

For a modulo additive-noise channel, a uniform input $\mathbf{V} \sim \text{Unif}(\mathcal{P}_0)$ maximizes the mutual information $I(\mathbf{V};\tilde{\mathbf{Y}})$, which becomes $\log(V_{\Lambda_2}) - h(\mathbf{Z}_{\text{eff}})$. The optimum α is thus the one that minimizes the entropy of the effective noise.[4] As the lattice dimension increases, the self noise $\mathbf{U}'$ and therefore the effective

[3] $((a \bmod {}_{\mathcal{V}_0}\Lambda_2) + b) \bmod {}_{\mathcal{P}_0}\Lambda_2 = (a + b) \bmod {}_{\mathcal{P}_0}\Lambda_2$.

[4] For rates below capacity, a smaller α would give better error performance [31], [44].

noise $\mathbf{Z}_{\text{eff}}$ become closer to a Gaussian distribution (in the divergence sense (17)), in which case minimizing entropy amounts to minimizing variance. Thus the optimum α becomes the Wiener coefficient $\alpha = \sigma^2_{\Lambda_2}/(\sigma^2_{\Lambda_2} + N) = P/P + N$, and the resulting noise variance is the MMSE solution

$$\text{Var}(\mathbf{Z}_{\text{eff}}) = \frac{PN}{P+N}. \qquad (25)$$

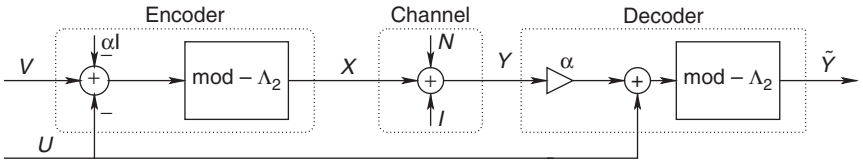


Fig. 7 Lattice-strategies for the dirty-paper channel.

Due to the dither, the error probability is *identical* for all codewords (as reflected by the equivalent modulo-additive channel of Thm. 6), and is equal to

$$P_e = \Pr\{\mathbf{Z}_{\text{eff}} \notin \mathcal{V}_{0,1}\}. \qquad (26)$$

Thus, by the definition of the the VNR (12), if we target some P_e, the volume of the fine lattice cell must be $V_{\Lambda_1} \approx [\mu(\Lambda_1, P_e) \cdot \text{Var}(\mathbf{Z}_{\text{eff}})]^{n/2}$ or larger, where we assumed a Gaussian $\mathbf{Z}_{\text{eff}}$.[5] On the other hand, the power constraint implies that the volume of the coarse cell is $V_{\Lambda_2} = [P/G(\Lambda_2)]^{n/2}$ or smaller. For the MMSE solution (25), we thus get a coding rate of

$$R = \frac{1}{n}\log\left(\frac{V_{\Lambda_2}}{V_{\Lambda_1}}\right) \approx \frac{1}{2}\log\left(\frac{P/G(\Lambda_2)}{\mu(\Lambda_1, P_e)\text{Var}(\mathbf{Z}_{\text{eff}})}\right) \qquad (27)$$

$$= C - \frac{1}{2}\log(G(\Lambda_2) \cdot \mu(\Lambda_1, P_e)) \qquad (28)$$

where $C = 1/2\log(1 + P/N)$ is the AWGN channel capacity.

The capacity loss in (28) is approximately the NSM-VNR cross product of the lattice pair. To reduce this loss, we need the coarse lattice to be a "good" quantizer, while the fine lattice should be a "good" AWGN channel code, both in the sense of Sec. III. For such a good pair of nested lattices $G(\Lambda_2) \to 1/2\pi e$ and $\mu(\Lambda_1, P_e) \to 2\pi e$ as $n \to \infty$, so the system approaches the AWGN channel capacity. An analysis of the error exponent of Voronoi codebooks can be found in [31], [44].

V.C. Achieving the Gaussian RDF

A dual construction of a *Voronoi quantizer* achieving the quadratic-Gaussian (QG) rate-distortion function can be designed along similar lines. Again, the NSM-VNR cross product of the lattice pair – now with the roles of Λ_1 and Λ_2 switched relative to (28) – will determine the rate loss of the system. The coarse lattice should therefore be a "good" AWGN channel code, while the fine lattice should be a "good" quantizer [54].

VI. Side-Information Problems

Classical Information Theory deals with point-to-point communication, where a single source is transmitted over a channel to a single destination. In a distributed situation there may be more than one (possibly correlated) sources, hence more than one encoder, and/or more destinations, hence more than one channel output and decoder. The simplest situation, which captures much of the essence in the problem, are sources and channels with side information.

In the source version of the problem – solved by Wyner and Ziv [47] – a source S is encoded knowing that a correlated signal J is available at the decoder (but not at the encoder). In the Gaussian case, we assume that $S = J + Q$, where Q is a white Gaussian source independent of J.

The channel version of the problem was solved by Gelfand and Pinsker in [19]. It assumes that the input to a state-dependent channel is encoded knowing the channel states non-casually. The decoding is done solely based on the channel output, without having access to the channel states. In the special case known as the "dirty paper" channel (DPC), or the *Costa problem*, the input-output relation is $Y = X + I + Z$, where I is an interference signal known at the encoder, and Z (the unknown noise) is AWGN [8].

An interesting feature of Gaussian side-information problems is that their information-theoretic solutions amount to complete elimination of the effect of the partially known signals J and I.

For the DPC problem, a simple variation on the Voronoi modulation and decoding system of Sec. V.B achieves the same coding rate as in (28), where now $C = (1/2)\log(1 + P/N)$ denotes the "clean" AWGN channel capacity [11], [39]. The main change is the subtraction of the scaled interference αI modulo the coarse lattice – see Fig. 7. (For a scalar-lattice solution for the *causal* DPC problem – see [11], [45].) The Gaussian Wyner-Ziv (WZ) problem is solved by a similar variation on the Voronoi quantization scheme of Sec. V.C [52]. In both DPC and WZ variations, the cosets of Λ_2 relative to Λ_1 (20) replace the *random bins* of the classical solutions of [19], [47].

A nice benefit of the dithered lattice approach is that the known parts (J and I) can be arbitrary signals, i.e., they do not even need to have a stochastic model. Yet, if J and I *are* random, then they can play the role of the dither, so common randomness becomes unnecessary.

See [22] for a *modulo lattice modulation* (MLM) scheme for *joint* source-channel coding with side-information using a *single* shaping lattice.

VII. Gaussian Networks

There are many ways in which side-information paradigms can enter general multi-terminal networks. The obvious cases are the broadcast channel, in which the (joint) encoder may view the transmission to one terminal as side-information for the transmission to the other terminals. Similarly, in multi-terminal coding of correlated sources, the (joint) decoder may view the reconstruction of one source as side information for the reconstruction of the other

[5] This assumption is true for high SNR (implying $\alpha = 1$), or high dimension and a "good" coarse lattice (to make the self-noise component "Gaussian enough"). Furthermore, the effective noise $\mathbf{Z}_{\text{eff}}$ is in fact more favorable than Gaussian noise for sufficiently small P_e, so the Gaussian approximation provides a lower bound on the rate of the system.

sources. In both these cases, the side-information is concentrated in the "relevant" terminal in the network. Indeed, in the QG case, it is easy to figure out how to replace the standard information-theoretic "random binning" technique by a lattice-based solution. This solution uses the the lattice-WZ and lattice-DPC schemes of Sec. VI as building blocks [52]. As in section VI, the main motivation for such a lattice scheme is the complexity reduction (and perhaps the intuition) gained by a structured solution.

A more interesting situation, however, occurs when side information is *distributed* among more than one terminal. Surprisingly, it turns out that in some distributed linear network topologies, the lattice-based system *outperforms* the random-binning solution. Moreover, in some cases it is in fact optimal! Apparently, the linearity of the network in these scenarios favors linear (rather than random) binning, as we already saw in the binary Körner-Marton problem.

VII.A. The Gaussian Körner-Marton Problem

Krithivasan and Pradhan [26] extended the Körner-Marton problem of Fig. 2 to the QG case. Suppose X and Y are positively correlated Gaussian sources, say, $Y = X + N$ where N is independent of X, and the decoder wants to reconstruct their difference N with some mean-squared distortion D. As they show, near-optimal performance can be achieved if each source is *lattice-WZ* encoded, where the coarse lattice – tuned to match the variance of the difference N – is *identical* at both encoders. The decoder subtracts the two encodings, modulo the coarse lattice, to isolate the desired (quantized) difference signal.

Unlike the original "lossless" KM setup, however, the lattice scheme does not match the "genie aided" outer bound; for $\sigma_x^2 \gg \sigma_n^2$, it loses 3dB in distortion (one bit in the sum rate) due to the accumulation of two independent quantization noises. Yet, at least for high rates this is still better than a "standard" random binning solution *a la* Berger-Tung [3], which (implicitly) encodes both sources X and Y just to transmit their difference.

VII.B. The Dirty Multiple Access Channel

We next consider what seems to be the "dual" of the Körner-Marton problem: a generalization of the Gaussian dirty-paper problem to a multiple access setup, as illustrated in Fig. 8. There are two additive-interference signals, one known to each transmitter but none to the receiver.

It is shown in [40] that the rates achievable using Costa's binning scheme (induced by his auxiliary random variables) vanish in the limit when the interference signals are strong. In contrast, if both

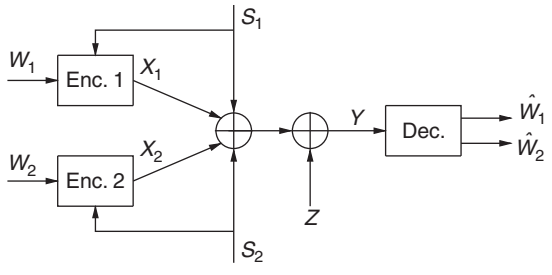


Fig. 8 Doubly dirty MAC.

encoders apply lattice-DPC using the *same* shaping (coarse) lattice Λ_s, then the sum interference is concentrated on Λ_s. The equivalent channel seen by the receiver is thus a MAC version of the modulo-additive channel of Thm. 6, and the sum rate is positive *independent* of the interferences.

Furthermore, [40] gives an outer bound for the capacity region of the dirty MAC for arbitrarily strong interferences, which is strictly smaller than the clean MAC capacity region. Lattice-DPC of large dimension meets this outer bound for some cases, in particular for *imbalanced* power constraints, as well as in the limit of high SNR [40].[6]

VII.C. The Loss of Single-Letter Characterization

Costa's binning scheme is derived from a Gaussian single-letter formula. It fails on the dirty-MAC because, unlike for lattice-binning, the sum of two independent bins (from the two users) results in a "bad" codebook. A similar phenomena occurs in the Gaussian Körner-Marton problem: the *difference* of two independent bins, each one generated by a Gaussian single-letter expression, results in a "bad" codebook. Are there better single-letter formulas for these two problems?

We conjecture that the best single-letter formula for the dirty MAC in the limit of strong interference and high SNR is given in terms of a one-dimensional lattice [40], [41]. The resulting rate loss is thus the "shaping gain" $(1/2)\log(2\pi e/12) \approx 0.254$ bits, i.e., the divergence from Gaussianity of a scalar dither (17). For a binary version of the dirty MAC, it is shown in [41] that the capacity loss of the best known single-letter formula is ~0.2 bits.

VII.D. Lattice Network Coding

In a standard packet switching network, nodes act as routers – they wish to find the best route for a packet under the current conditions. If the inflow to a node is higher than its output capacity, then some of the packets need to be discarded. The idea of network coding is that a bottleneck node can "combine" together packets rather than choose which one to pass on and which one to discard. If the final destination gets enough such "combinations" (from different routes), then it can resolve the ambiguity and decode all the transmitted packets reliably.

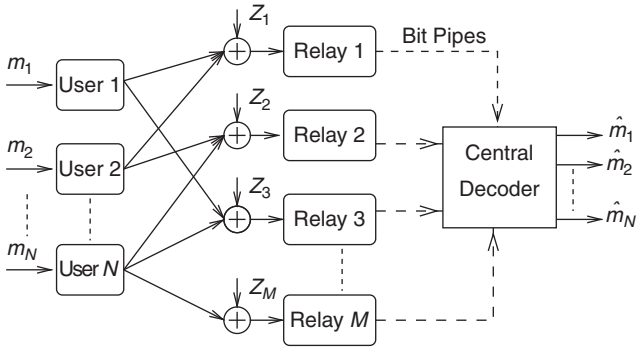


Fig. 9 A multi-relay multi-user network scenario.

[6]The loss w.r.t. the outer bound in the balanced case is similar to the 3dB loss in the Gaussian KM; it amounts to doubling the "self noise" component in (24), hence the "1" in the AWGN channel capacity formula reduces to some number $1 > \gamma > 1/2$ [40], [37].

The focus of most research on network coding has been on *linear* coding schemes [29]. In theory, though, any mapping at the nodes which is overall information preserving would work, as long as the network is lossless. In particular, random binning at the nodes is information preserving with high probability [20]. However, when extending the network coding idea to *noisy* networks, the structure of the code is essential to avoid *noise accumulation* and loss of capacity.

Specifically, consider the Gaussian relay network proposed in [36], depicted in Fig. 9, where N users wish to communicate with a destination (central decoder) through a layer of $M \geq N$ relays. Each relay receives some weighted (by the fading coefficients) linear combination of the transmitted signals corrupted by AWGN. Thus, the different signals at the relay input are already "combined" by the network. Relaying this combination as is (say, in some analog or compressed form) means that the noise will be forwarded to the final receiver as well. On the other hand, requiring the relay to decode all its input signals *separately* (as a MAC receiver) means a waste of capacity. See, e.g., [2].

It has been shown recently how to use lattice codes for ("physical-layer") network coding in the presence of Gaussian noise [35], [34], [36], [37]. If all the users use the *same* coding (fine) lattice, then the relay can decode an integer linear combination of the codewords (a lattice point which is close to the received signal), thus removing the channel noise before forwarding the decoded point to the final receiver. A particularly insightful example is that of the *two-way relay*, where each user computes its intended message from its own message and the message-sum it gets from the relay [35],[34].

A framework for treating non-Gaussian noise and non-additive channels is proposed in [12].

VII.E. Interference Alignment

A similar idea applies for the suppression of interference in a multi-node interference channel (IC). One of the interesting observations of the recent years is the idea of interference alignment [5]: a channel aware transmission system can make the effective number of interferers seen by each receiver equal to one. Thus, effectively, the multi-node IC is no worse than the classic double-node IC!

The original idea was to align the interference in the time domain, and it used linear transformations [33]. An alternative approach, based on alignment in the amplitude domain, was proposed in [38]. This approach fits very naturally into the lattice framework.

Consider the many-to-one interference channel of Fig. 10. Assuming the interference path gains of users 2 to L are identical, and that these users use the same coding (fine) lattice Λ_I, the equivalent channel seen by user 1 is similar to that seen in the dirty MAC of Sec. VII.B: the interference signals are all *concentrated* on the points of a *single* lattice Λ_I. Thus, in effect, user 1 experiences a single interferer. Furthermore, using an "estimate-and-modulo"

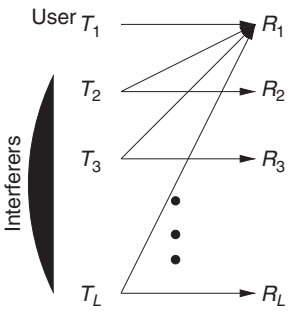


Fig. 10 Many-to-one interference alignment.

receiver as in Thm. 6, user 1 sees an equivalent modulo-Λ_I channel. Thus, it can achieve a rate of

$$R_1 = \frac{1}{2}\log\left(\min\left\{\frac{\sigma^2_{\Lambda_I}}{N}, \frac{P+N}{N}\right\}\right)$$

for large lattice dimension, corresponding to a full capacity in the strong interference regime.

VIII. Open Questions

On the practical side, lattice (or alternatively, linear trellis) codes with good performance and low encoding and decoding complexity are essential to make this theory attractive. New design approaches, e.g., [43], may be of interest.

The linear structure of the lattice plays a crucial role in the distributed lattice coding schemes presented in Sec. VII. For a proper operation, we need to align the lattice codes both in time and in amplitude. Yet in all the examples we considered, only one of the component codes of the system – either the shaping or the coding lattice – must be aligned. The other code does not even need to be a lattice! See Table 1. Other examples are of interest.

Random coding schemes – based on traditional single-letter (i.i.d.) solutions – seem to fail in these setups. For example, as discussed in Sec. VII.C, the loss of single-letter characterization in the Gaussian dirty MAC setup is conjectured to be $(1/2)\log(2\pi e/12) \approx 0.254$ bits.

Does structure really beat random? Note that proving the existence of good lattices requires random coding arguments [4], [42], [32], [10], [21], [56]. Also, our analysis of the lattice coding schemes assumes common randomness in the form of a dither. A question thus remains, if the failure of the traditional random coding approach is due to inappropriate single-letter solutions, or to its inherent weakness. We believe the latter to be true.[7]

The success of lattices in these setups hinges upon a good match between the linearity of the code and the linearity of the source or channel network. Can we go beyond the linear case?

Acknowledgement

My ISIT 10 talk was based on past and present work with Meir Feder, Gregory Poltyrev, Toby Berger, Shlomo Shamai, Uri Erez, Simon Litsyn, Dave Forney, Yuval Kochman and Tal Philosof. Thanks are also due to Bobak Nazer for helpful comments on the manuscript, and to Anatoly Khina for his help with the figures.

Table 1. Which component to align.

	Shaping (coarse) lattice	Coding (fine) lattice
Gaussian Korner Marton	aligned	–
dirty MAC	aligned	–
Lattice network coding	–	aligned
Interference alignment	–	aligned

[7]Nevertheless, this should not be seen as a discouraging fact, but rather as an indication for new directions and opportunities!

References

[1] R. Ahlswede and J. Körner. Source coding with side information and a converse for the degraded broadcast channel. *IEEE Trans. Information Theory*, vol. 21, pp. 629–637, 1975.

[2] S. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow: A deterministic approach. to appear in *IEEE Trans. Info. Theory* 2011, see http://arxiv.org/abs/0906.5394.

[3] T. Berger. Multiterminal Source Coding. New York: In G. Longo, editor, the Information Theory Approach to Communications, Springer-Verlag,1977.

[4] R. de Buda. Some optimal codes have structure. *IEEE Jr. on Selected Areas in Comm.*, 7:893–899, Aug. 1989.

[5] V. R. Cadambe, S. A. Jafar. Interference Alignment and the Degrees of Freedom for the K User Interference Channel. *IEEE Trans. Info. Theory*, IT-54(8):3425–3441, Aug. 2008.

[6] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Springer-Verlag, New York, N.Y., 1988.

[7] J. H. Conway and N. J. A. Sloane. Voronoi regions of lattices, second moments of polytopes, and quantization. *IEEE Trans. Info. Theory*, IT-28:211–226, Mar. 1982.

[8] M.H.M. Costa. Writing on dirty paper. *IEEE Trans. Info. Theory*, IT-29:439–441, May 1983.

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

[10] U. Erez, S. Litsyn, and R. Zamir. Lattices which are good for (almost) everything. *IEEE Trans. Info. Theory*, IT-51:3401–3416, Oct. 2005.

[11] U. Erez, S. Shamai, and R. Zamir. Capacity and lattice strategies for cancelling known interference. *IEEE Trans. Info. Theory*, IT-51:3820–3833, Nov. 2005.

[12] U. Erez and R. Zamir. A modulo-lattice transformation for multiple-access channels. In *Electrical and Electronics Engineers in Israel, 2008 IEEE 25th Convention of,* Dec. 2008. Also at ITA, UCSD, February 2009.

[13] U. Erez and R. Zamir. Achieving 1/2 log(1+SNR) on the AWGN channel with lattice encoding and decoding. *IEEE Trans. Info. Theory*, IT-50:2293–2314, Oct. 2004.

[14] G. D. Forney Jr. and L. F. Wei. Multidimensional constellations-Part I: Introduction, figures of merit, and generalized cross constellations. *IEEE Jr. on Selected Areas in Comm.*, 7:877–892, Aug. 1989.

[15] G. D. Forney Jr. and L. F. Wei. Multidimensional constellations-Part II: Voronoi constellations. *IEEE Jr. on Selected Areas in Comm.*, 7:941–958, Aug. 1989.

[16] G. D. Forney Jr. On the duality of coding and quantizing. In *DIMACS Ser. Discr. Math. Theory Comp. Sci.*, volume 14, 1993.

[17] G. D. Forney Jr., M.D. Trott, and S.-Y. Chung. Sphere-bound-achieving coset codes and multilevel coset codes. *IEEE Trans. Info. Theory*, IT-46:820–850, May, 2000.

[18] G. D. Forney Jr. On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener. In *41st Annu. Allerton Conf. Communication, Control, and Computing*, pp. 430–439, Oct. 2003.

[19] S.I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 9, No. 1:19–31, 1980.

[20] T. Ho, M. Medard, R. Koetter, D.R. Karger, M. Effros, Jun Shi, B. Leong. A Random Linear Network Coding Approach to Multicast. *IEEE Trans. Information Theory* IT-52(10):4413–4430, Oct. 2006.

[21] A. Ingber, R. Zamir, M. Feder. The optimal density of infinite constellations for the Gaussian channel. on Arxiv.

[22] Y. Kochman and R. Zamir. Joint Wyner-Ziv/dirty-paper coding using analog modulo-lattice modulation. *IEEE Trans. Info. Theory*, IT 55(11): 4878-4889, Nov. 2009.

[23] Y. Kochman and R. Zamir. Analog matching of colored sources to colored channels. *IEEE Trans. Info. Theory*, to appear 2011.

[24] Y. Kochman, A. Khina, U. Erez, and R. Zamir. Rematch and forward for parallel relay networks. In *ISIT-2008, Toronto, ON*, pages 767–771, 2008.

[25] J. Körner and K. Marton. How to encode the modulo-two sum of binary sources. Em *IEEE Trans. Information Theory*, vol. IT-25, pp. 219–221, March 1979.

[26] D. Krithivasan and S. S. Pradhan. Lattices for distributed source coding: Jointly Gaussian sources and reconstructions of a linear function. Submitted to *IEEE Trans. Inform. Theory,* July 2007, arXiv:0707.3461, e-print.

[27] D. Krithivasan and S. S. Pradhan. A proof of the existence of good nested lattices, [Online]. Available: http://www.eecs.umich.edu/ techreports/systems/cspl/cspl-384.pdf

[28] F. R. Kschischang and S. Pasupathy, Optimal nonuniform signaling for Gaussian channels. *IEEE Trans. Info. Theory*, IT-39(3):913–929, May, 1993.

[29] Li, S.-Y.R., Yeung, R.W. and Ning Cai. Linear network coding. *IEEE Trans. Inform. Theory*, 49(2):371–381, Feb. 2003.

[30] T. Linder, C. Schlegel, and K. Zeger Corrected proof of de Buda's theorem *IEEE Trans. Inform. Theory*, 39(5):1735–1737, Sep. 1993.

[31] T. Liu, P. Moulin, and R. Koetter. On error exponents of modulo lattice additive noise channels. *IEEE Trans. Inform. Theory*, 52:454–471,Fe b.2006.

[32] H. A. Loeliger. Averaging bounds for lattices and linear codes. *IEEE Trans. Info. Theory,* 43:1767–1773, Nov. 1997.

[33] Maddah-Ali, M.A.; Motahari, A.S.; Khandani, A.K. Communication Over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis. *IEEE Trans. Info. Theory*, IT-54(8): 3457–3470, Aug. 2008.

[34] W. Nam, S.-Y. Chung, and Y. H. Lee, Capacity bounds for two-way relay channels, in International Zurich Seminar on Communications. In *IZS 2008, Zurich, Switzerland,* March 2008.

[35] K. Narayanan, M. P. Wilson, and A. Sprintson. Joint physical layer coding and network coding for bi-directional relaying. In *45th Annual Allerton Conference*, Monticello, IL, Sept., 2007.

[36] B. Nazer and M. Gastpar, Compute-and-forward: Harnessing interference with structured codes. *In Proceedings of ISIT 2008,* July 6–11,Toronto, Canada.

[37] B. Nazer and M. Gastpar, Reliable Physical Layer Network Coding to appear in *IEEE Proc.* March 2011.

[38] A. Parekh, G. Bresler, and D. Tse. The approximate capacity of the many-to-one and one-to-many gaussian interference channels. In *Proc. of 45th Allerton Conference (Monticello, IL)*, Sep. 2007.

[39] T. Philosof, U. Erez, and R. Zamir. Combined shaping and precoding for interference cancellation at low SNR. In *Proc. IEEE International Symposium on Information Theory, pp. 68, (Yokohama, Japan)*, June 2003.

[40] T. Philosof, A. Khisti, U. Erez, and R. Zamir, Lattice strategies for the dirty multiple access channel. In *Proc. of IEEE International Symposium on Information Theory*, Nice, France, June 2007. Also accepted for publication in the *IEEE Trans. Info. Theory.*

[41] T. Philosof, and R. Zamir, On the loss of single letter characterization: the dirty multiple access channel. *IEEE Trans. Info. Theory,* IT 55(6):2442–2454, June 2009.

[42] G. Poltyrev. On coding without restrictions for the AWGN channel. *IEEE Trans. Info. Theory,* IT–40:409–417, Mar. 94.

[43] Sommer, N., Feder, M. and Shalvi, O. Low-Density Lattice Codes. *IEEE Trans. Info. Theory*, IT-54(4):1561–1585, April 2008.

[44] C. Swannack, U. Erez, and G. W. Wornell. Reflecting on the AWGN error exponent. In 43rd Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois, Sept. 28–30, 2005.

[45] F. Willems. Signalling for the Gaussian channel with side information at the transmitter In *Proc. IEEE International Symposium on Information Theory, p. 348*, (Sorrento, Italy), June 2000.

[46] A. Wyner. On source coding with side information at the decoder. IEEE Trans. Information Theory, vol. IT-21, pp. 294–300, 1975.

[47] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Theory,* IT-22:1–10, Jan., 1976.

[48] R. Zamir and M. Feder. On universal quantization by randomized uniform / lattice quantizer. *IEEE Trans. Info. Theory*, pages 428–436, March 1992.

[49] R. Zamir and M. Feder. Information rates of pre/post filtered dithered quantizers. *IEEE Trans. Info. Theory,* pages 1340–1353, Sep. 1996.

[50] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Info. Theory*, pages 1152–1159, July 1996.

[51] R. Zamir and S. Shamai. Nested linear / lattice codes for Wyner-Ziv encoding. In *Proceedings of the Information Theory Workshop, Killarney, Ireland*, pages 92–93, June 1998.

[52] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Info. Theory*, IT-48:1250–1276, June 2002.

[53] R. Zamir, Y. Kochman, and U. Erez. Achieving the Gaussian rate distortion function by prediction. *IEEE Trans. Info. Theory,* IT-54:3354–3364, July 2008.

[54] R. Zamir. Lattices are everywhere. *Proc. of ITA* 2009, pp. 392–421, U San Diego CA, 8–13 Feb. 2009.

[55] R. Zamir. How to generate a simple dither. In *IEEE convention*, Eilat, Israel, Nov. 2010.

[56] R. Zamir. *Lattice Coding for Signals and Networks*. In preparation.

[57] J. Ziv. On universal quantization. *IEEE Trans. Info. Theory,* IT-31:344–347, May 1985.

Abelian Varieties in Coding and Cryptography

Plenary talk presented at the 2010 IEEE Information Theory Workshop, Dublin, Ireland.

Ian F. Blake
Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, British Columbia, Canada
Email: ifblake@ece.ubc.ca

***Abstract*—**Algebraic curves over a finite field have played a central role in both coding theory and cryptography over the past three decades. In coding theory the use of algebraic curves led to the discovery of asymptotically good codes whose parameters lie above the Varshamov-Gilbert bound in certain cases while in cryptography the use of elliptic curves led to public key cryptosystems that are more efficient, in some sense, for a given level of security than integer factorization based ones. It would seem natural that the use of higher dimensional varieties might lead to even better results for both applications. Such has not so far been the case in any dramatic way. The purpose of this talk is to review the situation on the use of Abelian varieties in these two areas.[1]

I. Introduction

The success enjoyed by the use of algebraic curves in both coding and cryptography is well documented. In coding theory the use of such curves led to the construction of codes whose parameters, asymptotically, exceeded the Varshamov-Gilbert bound. In cryptography, the use of elliptic curves led to cryptosystems whose security, as far as is known, is proportional to the square root of the group order, rather than the subexponential security of integer factorization or certain discrete logarithm problems. The dramatic success of the use of curves, in both domains, make it natural to consider the use of higher dimensional analogs, Abelian varieties, in the sense that the increased degrees of freedom might yield further dividends.

This work reviews the progress that has been made in this effort. The next section gives the necessary notation and background on Abelian varieties. The special case of Hermitian varieties is used for illustrative purposes. Section III reviews the progress on the use of higher dimensional varieties in the construction of codes with good parameters. Section IV considers the application of higher dimensional geometry to cryptography. Much of this section focuses on Jacobians and supersingular varieties. A few comments on the the subject are given in the final section. Excellent references for the work are [6], [17], [18].

To make the article accessible to a wide audience, and in the interests of space, an informal approach to some of the definitions has been taken, at times sacrificing precision for space.

II. Preliminaries on Abelian Varieties

Denote by $\mathbb{P}^m$ the set of projective $(m+1)$-tuples over the finite field $\mathbb{F}_q$:

$$\mathbb{P}^m = \{(x_0, \cdots x_m) \,|\, \text{nonzero scalar multiples identified}, x_i \in \mathbb{F}_q\}$$

where $\#\mathbb{P}^m = \pi_{m,q} = q^m + q^{m-1} + \cdots + 1$. Similarly denote by $\mathbb{A}^m$ the set of affine m-tuples:

$$\mathbb{A}^m = \{(x_1, \ldots, x_m \,|\, x_i \in \mathbb{F}_q\}, \quad \#\,\mathbb{A}^m = q^m.$$

Varieties may be either affine or projective. and it is convenient to limit the discussion to projective, with the affine case being similar. Let $f \in \mathbb{F}_q[x_0, x_1, \ldots, x_m] = \mathbb{F}_q[X]$ be a homogeneous irreducible polynomial (all monomials of same degree) and let

$$V_f = \{P \in \mathbb{P}^m \,|\, f(P) = 0\}.$$

V_f will also be referred to as a hypersurface. More generally [6] one defines a prime ideal I of polynomials in $\mathbb{F}_q[X]$ and define V_I as the set of common zeros of the polynomials of I. Then V is a projective variety if and only if $V = V_I$ for some prime ideal in $\mathbb{F}_q[X]$. For coding applications such varieties will be sufficient. The *dimension* of a variety is [6] is the supremum of the length of a chain of distinct irreducible closed subspaces. A curve then is a variety of dimension 1 and a higher dimensional variety has dimension greater than 1.

For cryptography one needs the ability to compute in the variety which will lead to Abelian varieties. The notion of an Abelian variety [6] as a higher dimensional analog of an algebraic curve is introduced in an informal manner. An algebraic group is a (projective) variety, $\mathcal{X}$, with regular maps

$$m\colon \mathcal{X} \times \mathcal{X} \to \mathcal{X} \text{ and } \iota\colon \mathcal{X} \to \mathcal{X}$$

of addition and inverse along with a neutral element satisfying the usual properties. It can be shown that such a law must be commutative and such projective algebraic groups are called Abelian varieties. However, if the variety is defined as above, say as the set of projective points which are the zeros of an $(m+1)$-variable homogeneous polynomial, it may not be obvious, in general, how to define the two regular maps above on the variety and hence obtain an Abelian variety.

A convenient approach for cryptographic work will be to construct Abelian varieties from absolutely irreducible smooth projective curves via their Jacobian in the following manner. Let $\mathcal{C}_q$ be a curve of genus g over $\mathbb{F}_q$ defined as the set of solutions over the algebraic closure of $\mathbb{F}_q$ of the polynomial $f(x,y) = 0$, $f \in \mathbb{F}_q[x,y]$ and let the set of $\mathbb{F}_q$ rational points on the curve be:

$$\mathcal{C}(\mathbb{F}_q) = \{P = (x,y) \in \mathbb{F}_q^2 \,|\, f(x,y) = 0, f \in \mathbb{F}_q[x,y]\}$$

A divisor D of $\mathcal{C}_q$ is a formal sum of the form

$$D = \sum_{P \in \mathcal{C}_q} m_P P,$$

[1]This note is a modified version of an invited talk at the Information Theory Workshop in Dublin, August, 2010.

where the m_P are integers and only a finite number of them are nonzero. The divisors form a group under the natural addition. Define the degree of the divisor D to be the sum $\sum_{P\in\mathcal{C}_q} m_P \in \mathbb{Z}$. The divisors of degree 0 form a subgroup D_0.

By a function on the curve $\mathcal{C}_q$ is meant a rational bivariate function of the form $h(x,y) = a(x,y)/b(x,y),\ a, b \in K[x,y]$ where $b(x,y)$ is not divisible by the equation of the curve. The function h has a finite number of poles and zeros on the curve. To such a function we associate a divisor $div(h)$ defined as

$$div(h) = \sum_{P\in\mathcal{C}_q} m_P P$$

where, if h has a zero at P, m_P is the degree of that zero, and if h has a pole at P, m_P is the negative of the degree at that pole. The $div(h)$ is clearly of degree 0 taking into account the poles or zeros at the point at infinity, $\mathcal{O}$, and such a divisor is called principal i.e. the divisor $D \in D_0$ is principal if there is a function $h \in K(x,y)$ such that $D = div(h)$. The set of all such principal divisors is a subgroup $\mathcal{P}$.

The factor group $J_{\mathcal{C}_q} = D_0/\mathcal{P}$ is referred to as the Jacobian variety of the curve $\mathcal{C}_q$. If the curve is of genus g the Jacobian is an Abelian variety of dimension g. There is a natural arithmetic in the factor group of cosets, which is informally described here for the case of *hyperelliptic curves*, perhaps the most important extension from the elliptic curve case. Define by $J_{\mathcal{C}(\mathbb{F}_q)}$ to be the set of $\mathbb{F}_q$ rational points of $J_{\mathcal{C}_q}$ i.e. those elements of $J_{\mathcal{C}_q}$ fixed by the q-power Frobenius map (see later).

A hyperelliptic curve $\mathcal{C}_q$ of genus g over a finite field $\mathbb{F}_q$ is a smooth projective curve which admits an affine equation of the form [9]

$$y^2 + h(x)y = f(x)$$

where f is a polynomial of degree $2g+1$ and h is a polynomial of degree at most g, $\mathbb{F}_q \in \mathbb{F}_q[x]$. Note that if the point (x,y) is on the curve $\mathcal{C}_q$ then the *opposite* or *symmetric* point $(x, -y-h(x))$ is also. It can be shown that in each class of $J_{\mathcal{C}(\mathbb{F}_q)}$ there is a unique divisor of the form $D = P_1 + \cdots + P_k - k\mathcal{O},\ k \le g$, referred to as a *reduced* divisor, such that for all $i \neq j$ P_i and P_j are not opposite points. Then there is a unique representation of D by two polynomials $[u,v]$, $\deg v < \deg u \le g$ and u is monic and divides $v^2 + hv - f$ [9], $u, v \in \mathbb{F}_q[x]$. Here, the roots of the polynomial u are the x-coordinates of the points P_i appearing in the reduced divisor. With such a representation the divisor is said to be *prime* if the polynomial u is irreducible over $\mathbb{F}_q$ and the degree of of the divisor will be taken as the degree of u. It can be shown that the divisor $D \in J_{\mathcal{C}(\mathbb{F}_q)}$ with representative $[u,v]$ is equal to the sum of the prime divisors $[u_i, v_i]$ where the u_i are the irreducible factors of u. These observations will be useful in discussing their application to cryptography in Section IV. Arithmetic on the Jacobian is that of addition of cosets in the factor group in terms of the coset representative [5].

The Jacobian of a curve (and especially hyperelliptic curves) of genus g is a useful technique for forming Abelian varieties in that such a construction comes with a natural addition. This observation will be of value in cryptography, discussed in Section IV.

A natural question is which Abelian varieties arise as Jacobian of curves? This is known as the Schottky problem which has been widely investigated.. For lower dimensional varieties it is often the case that a given variety is the Jacobian of some curve.

Denote by $\mathcal{A}_q$ an arbitrary Abelian variety of dimension g with defining equations over the finite field $\mathbb{F}_q$, and the group of elements of the variety over the closure of $\mathbb{F}_q$ and by $\mathcal{A}(\mathbb{F}_{q^s})$ the group of $\mathbb{F}_{q^s}$ rational elements and by $\#\mathcal{A}(\mathbb{F}_{q^s})$ the number of such elements. In this last notation the dependence of the variety defined over $\mathbb{F}_q$ is understood. A great deal is known on such quantities. Of prime importance is the zeta function of the variety which enumerates the number of elements in the group $\#\mathcal{A}(\mathbb{F}_{q^s})$. The zeta function for an Abelian variety is defined as:

$$Z_{\mathcal{A}}(t) = \exp\left(\sum_{s=1}^{\infty} N_s \frac{t^s}{s}\right)$$

where $N_s = \#\mathcal{A}(\mathbb{F}_{q^s})$. The zeta function of a smooth projective variety is rational and the numerator is the characteristic polynomial of the Frobenius endomorphism. The Frobenius endomorphism of $\mathcal{A}_q$ is

$$\pi: x \to x^q$$

and its characteristic function is denoted $P_{\mathcal{A}_q}(t)$ i.e. the monic polynomial equation of least degree the endomorphism satisfies at points of the variety. These functions encapsulate much information of the variety, some of which is summarized here.

It can be shown ([23]) the roots of this polynomial $\{\alpha_i\}$, are of the form $\zeta q^{1/2}$ for some $2g$-th root of unity ζ, or $|\alpha_i| = \sqrt{q}$ and $\alpha_i \cdot \alpha_{g+i} = q$. The size of the variety is given by

$$\#\mathcal{A}(\mathbb{F}_q) = P_{\mathcal{A}_q}(1) = \prod_{i=1}^{2g}(1-\alpha_i).$$

The following bounds result as a consequence

$$[(\sqrt{q}-1)^{2g}] \le \#\mathcal{A}(\mathbb{F}_q) \le [(\sqrt{q}+1)^{2g}]$$

and

$$|\#\mathcal{A}(\mathbb{F}_q) - (q^g+1)| \le 2gq^{g^{-\frac{1}{2}}} + 4^g q^{g-1}$$

and the size of the set of $\mathbb{F}_q$ -rational points of the variety of dimension g is $O(q^g)$.

In the case the variety arises as the Jacobian of a curve $\mathcal{C}_q$ over $\mathbb{F}_q$ then the number of points on the curve itself can be expressed in the form

$$\#\,\mathcal{C}(\mathbb{F}_{q^s}) = q^s + 1 - \sum_{i=1}^{2g}\alpha_i^s$$

while the number of elements of the resulting Jacobian of the curve is

$$\#\,J_{\mathcal{C}(\mathbb{F}_{q^s})} = \prod_{i=1}^{2g}(1-\alpha_i^s) \sim \mathcal{O}(q^{gs}).$$

An elliptic curve is of genus 1 and its Jacobian variety is in fact isomorphic to the elliptic curve itself, with group operation the usual curve addition. The number of points of the curve (over $\mathbb{F}_q$) determines the number of points on the curve over $\mathbb{F}_{q^s}$, $s > 1$.

Just as the Hasse-Weil bound is available for curves, states that

$$|\#\mathcal{X}(\mathbb{F}_q) - (q+1)| \leq 2g\sqrt{q}$$

the analogous bound for hypersurfaces, useful in coding, in terms of degrees (of the defining multivariate polynomials) is [16]:

for $\mathcal{X}(\mathbb{F}_q)$ a smooth non-degenerate (not contained in any linear subspace of $\mathbb{P}^m$) hypersurface of degree s (of the defining equation) in $\mathbb{P}^m$, $m \geq 2$ over $\mathbb{F}_q$, then

$$|\#\mathcal{X}(\mathbb{F}_q) - (q^{m-1} + \cdots + q + 1)| \leq b(s,m)q^{(m-1)/2} \qquad (1)$$

where $b(s,m) = ((s-1)/s)((s-1)^m - (-1)^m)$ when m is even and one less when m is odd.

A. Hermitian Varieties

Hermitian varieties form a nice example of the previous theory and certain results for the set of rational points on an Hermitian hypersurface are considered here.

For this subsection let $q = r^2$ for a prime power r. The m-th order Hermitian variety over $\mathbb{F}_q$, $\mathcal{X}_m(\mathbb{F}_q)$, is the set of $\mathbb{F}_q$ rational solutions to the equation

$$x_0^{r+1} + x_1^{r+1} + \cdots + x_m^{r+1} = 0. \qquad (2)$$

The number of solutions to this equation over $\mathbb{F}_q$ was first given by Bose and Chakravarti [3], [4] and is given by

$$\begin{aligned}\#\mathcal{X}_m(\mathbb{F}_{r^2}) &= r^{2m-2} + \cdots + r^2 + 1 + b(r+1)r^{m-1} \\ &= \frac{(r^{m+1} - (-1)^{m+1})(r^m - (-1)^m)}{(q-1)}\end{aligned}$$

It is noted this number meets the upper bound of equation (1) and hence the Hermitian hypersurfaces are maximal for all dimensions $m \geq 1$.

It is a simple matter to actually construct the solutions to the Hermitian surface, enumerated above, as was essentially done in [3], [4], [2]. It is also straightforward to derive the zeta function for the variety (e.g. [2]) as well as other information which will prove useful in their application.

III. Abelian Varieties in Coding Theory

A brief overview of codes from hypersurfaces is given here, drawn mainly from the excellent survey article [16] and the paper [11]. Only evaluation codes are considered i.e codes whose codewords are obtained by evaluating homogeneous polynomials of a certain degree over the points of a projective variety, $\mathcal{X}_q$. In particular let $\mathcal{S} \subseteq \mathcal{X}_q$ be a subset of the rational points over of the variety $\mathcal{X}_q$ (often $\mathcal{S}$ will be the set of all rational points). Denote by $\mathbb{F}_q[x_0, x_1, \ldots, x_m]_{\leq h}$ the set of homogeneous polynomials over $\mathbb{F}_q$ of degree $\leq h$ in the $(m+1)$ variables and by $\mathbb{F}_q[x_0, x_1, \ldots, x_m]_h = \mathcal{F}_{h,m}$ the set of such polynomials of degree exactly h. Such sets will be used for the projective codes (to be defined later).

Consider the code defined by the map from the space of homogeneous polynomials in $m+1$ variables to the vector space over $\mathbb{F}_q$.

To make it concrete, let $S = \{P_1, P_2, \ldots, P_n\}$, $\#S = n$, $S \subseteq \mathcal{X}_q$ be a subset of the set of points X on a variety $\mathcal{X}$:

$$c\colon \mathcal{F}_{h,m} \to F_q^{\#S}$$

$$f \mapsto (f(P_1), f(P_2), \ldots, f(P_n)).$$

The code is denoted by $C_h(X,S)$. For $h=1$, i.e. the set of linear forms, denote the code $C(X) = C_1(X,S)$. The dimension of $C_h(X,S)$ is the dimension of the space of homogeneous polynomials less the dimension of the kernel of the above map. To determine the minimum distance of the code, let

$$V(f) = \{P \in S \,|\, f(P) = 0\}$$

that is, the number of coordinate positions of the codeword corresponding to the homogeneous polynomial f that are zero. Then the minimum distance of the code is

$$d = \min_{f \neq 0, f \in \mathcal{F}_{h,m}} (n - |(V(f) \cap \mathcal{S}|)).$$

The following bound is obtained in [14]: for $\mathcal{X}$ a projective variety of dimension m and degree $s < q+1$, for $h=1$ the code $C_1(X,\mathcal{S})$ has minimum distance:

$$d \geq n - s(q^{m-1} + \cdots + q + 1).$$

Numerous other results are available for special types of surfaces and for linear and quadric forms. Two examples of the theory are briefly noted, those of the full projective codes and those from Hermitian varieties.

A. Projective Reed-Muller Codes

Since all the codes considered here can be viewed as punctured projective Reed-Müller codes where certain coordinate positions are dropped, a brief review of the information on such codes is noted, drawn mainly from Lachaud [15], [16] and [22].

The *projective* Reed-Müller codes are defined as follows. Using all the projective points in $\mathbb{P}^m$, a variety of dimension m, and the all the homogeneous polynomials of degree h (i.e. those of $\mathcal{F}_{h,m}$), the resulting code will be called the *projective* Reed-Müller code of order h and denoted $\mathcal{R}_q(h, \mathbb{P}^m)$ and for $h < q$ has the parameters ([15]):

$$\text{length} = \pi_{m,q}, \quad \text{dimension} = \binom{h+m}{h}$$

$$\text{distance} = (q-h+1)q^{m-1}.$$

An affine version of the codes can also be defined. The parameters of the two codes are compared in [15]. It might be argued the projective codes are slightly better, but they are quite close. One can also obtain expressions for the code parameters when $q \leq h \leq m(q-1)$ [22] (for values of h above this value the codes are trivial).

B. Codes from Hermitian Varieties

Much of the literature on codes from Hermitian varieties has focused on only a limited subset of parameters. For $m=3$ the code formed from linear forms ($h=1$) can be shown [16] to have the parameters

$$n = (r^2+1)(r^3+1), \quad k = 4, \quad d = r^5.$$

For $m = 3$ but higher degree h forms, the resulting code has the parameters

$$n = (r^2+1)(r^3+1), \quad k = \binom{4+h}{h}, \quad d \geq n - h(r+1)(r^2+1).$$

For general m the variety is of dimension m and the parameters of the code $C_1(X, \mathcal{S}) = C(X)$ [16] are:

$$n = X_m(\mathbb{F}_{r^2}), \quad k = m+1,$$

$$d = \begin{cases} r^{2m-1} - r^{m-1} & m \equiv 0 \pmod 2 \\ r^{2m-1} & m \equiv 1 \pmod 2 \end{cases}.$$

Much of the recent work on these codes has focused on quadric the case, $h = 2$.

While the two examples are interesting, as well as others in the literature, as far as the author is aware there are no examples of codes from hypersurfaces that exhibit significantly better parameters than those available from curves. Modular curves played a prominent role in producing asymptotically good codes but the author is unaware of work to extend these notions to higher dimensions.

IV. Abelian Varieties in Cryptography

To discuss the use of Abelian varieties in cryptography a slight diversion into the complexity of the discrete logarithm problems is needed.

Discrete logarithm based public key cryptosystems depend for their security on the complexity of finding discrete logarithms. In a multiplicative group, such as the group of nonzero elements of a finite field, $\mathbb{F}_p^*$ (nonzero integers mod the prime p), if α is a generator, the discrete logarithm problem (DLP) is: given α, $y = \alpha^x$, find x, the discrete logarithm of y. In an additive group, such as the set of points on an elliptic curve (under their natural addition), the DLP is: given a point P on the curve and $xP = P + P + \cdots P$, x times, find x. To maximize security for a given size group, such problems always take place in a large prime order subgroup. To simplify the discussion, this distinction is ignored.

The complexity of solving the DLP in these two structures is very different. To understand the difference the notion of smoothness is noted. In the first structure, an integer $b \in \mathbb{F}_p^*$ is called smooth with respect to a if all of its prime divisors are less than a. An arbitrary integer can be uniquely decomposed into its prime factors. This leads to an *index calculus* attack on the DLP. There are two parts to the attack. In the first part a *factor base* D_x is formed consisting of all primes less than a given integer x. The discrete logarithms of all elements of D_x are found by choosing random elements in $\mathbb{F}_p^*$ of the form α^a for known a. If this (as an integer) factors into D_x a relationship is found between a and the discrete logarithms of a subset of elements of D_x. If a sufficient number of independent relations is found, they can be solved to find the discrete logarithms of D_x. In the second part of the algorithm a given element whose logarithm is required is randomly perturbed in a certain manner until it factors into D_x allowing the logarithm to be found. By carefully choosing x (hence the size of D_x) to balance the work between the two parts of the algorithm, the number of operations required to solve the DLP can be shown to be

$$L_p(a, c) = \mathcal{O}(\exp(c \ln(p)^a \ln\ln(p)^{1-a})).$$

Such complexity is referred to as *subexponential*- for $a = 1$ it is exponential (in $\ln(p)$) and for $a = 0$ it is polynomial. This complexity is considerably less than that required for many systems that lack the notion of smoothness.

In the absence of the notion of smoothness, an algorithm such as the *baby-step-giant-step* (BSGS) algorithm, applicable to any cyclic group, is available and has complexity $\mathcal{O}(\sqrt{p})$. The distinction between the performance of these two algorithms has driven much cryptographic research over the past twenty five years. The use of elliptic curves in cryptography was first proposed in 1985 independently by Koblitz and Miller. There is a natural addition of points on an elliptic curve and a (additive) subgroup of prime order q is used for cryptographic applications. The reason that elliptic curves were proposed is the thought that no notion of smoothness for use in the DLP problem was likely, for any cryptographically useful curve. While a notion of smoothness has been applied to an infinite family of certain types elliptic curves [7], such curves are easily avoided in practice. Thus the best algorithm for solving the DLP for 'useful' elliptic curves is the BSGS (or similar algorithm) with complexity $\mathcal{O}(\sqrt{q})$.

The implication of these observations is that a discrete logarithm based cryptosystem in $\mathbb{F}_p^*$ with a 1024 bits prime ($\sim\log(p)$) has a similar security to an elliptic curve based system over a prime field of 163 bits. Similarly a discrete logarithm based system with a 3072 bit prime has a similar security to an elliptic curve system with a 256 bit prime.

The success of elliptic curves led to the thought that higher dimensional analogs might yield further dividends. Since there is no natural addition of points on a hyperelliptic curve, the construction of an Abelian (Jacobian) variety $J_{\mathcal{C}}(\mathbb{F}_q)$ from a curve $\mathcal{C}$ of genus g over $\mathbb{F}_q$ as discussed in Section II is used. Two applications of this theory to cryptography are mentioned here.

The notion of using a hyperelliptic curve for cryptography was proposed by by Koblitz [13] and we use the terminology introduced in Section II. Using a hyperelliptic curve of genus g leads to a Jacobian variety of dimension g and of size approximately q^g. Arithmetic in this variety is that of the classes of the factor group. It was noted in Section II that each class has a unique representative that can be expressed as a pair of polynomials $[u, v]$. Techniques for the addition of two classes, viewing $J_{\mathcal{C}(\mathbb{F}_q)}$ as an additive group, was first addressed by Cantor [5]. Since then the problem has been an active area of research and now effective algorithms exist to perform very efficient arithmetic on the Jacobian of a hyperelliptic curve.

It was perhaps originally presumed by many researchers that, as there was no index calculus attack for the discrete logarithm problem on elliptic curves, the same would hold true true for hyperelliptic systems. However, such was not the case. Adleman et al [1] gave an index calculus attack on hyperelliptic systems which was effective for large genus. Gaudry ([9]) showed the result that there is a notion of a smooth divisor in the Jacobian variety which can be used to construct an index calculus attack, much as for a discrete logarithm attack in finite fields, that is effective for smaller genus curves. He defined a divisor with representative $[u, v]$ in $J_{\mathcal{C}(\mathbb{F}_q)}$ to be

S-smooth if all its prime divisors are of degree at most S. Of course a 1-smooth divisor is one for which all the irreducible factors of the polynomial u are of degree 1. The factor base then is taken to be the set of all prime divisors of degree at most S. With this notion of smoothness he [9] proved the following:

Theorem 1: The algorithm (for determining discrete logarithms in the Jacobian of a hyperelliptic curve of genus g) requires $O(q^2 + g!q)$ polynomial time operations in $g \log q$ and if one considers a fixed genus g, the algorithm takes time $O(q^2 \log^\gamma q)$.

More recent work [10] has shown an algorithm of time complexity $O(q^{2-2/g})$ for small genus curves. For $g = 4$ the complexity of such an attack would be $O(q^{3/2})$ and for $g = 3$ the complexity would be $O(q^{4/3})$ while for a BSGS type attack they would be $O(q^2)$ and $O(q^{3/2})$ respectively, since the group order is $\sim q^g$. The argument is to not use a system with less than square root (of group order complexity) since one could use an elliptic curve system more effectively. This would seem to limit the use of such curves of genus 2 and 3. However genus 2 and 3 hyperelliptic/Jacobian variety systems have been studied extensively in the literature. For example, for a genus 2 hyperelliptic system, arithmetic is over $\mathbb{F}_q$ while the group order is $O(q^2)$ while for an equivalent elliptic curve system the group order would be $O(q)$. Thus for the same level of security one could use a field size $O(\sqrt{q})$ in comparison with the elliptic system, making the bit length of the arithmetic for the hyperelliptic system half as long. Such considerations might well be important for constrained environments and genus 2 and 3 curves appear to have very efficient implementations to the point where they are very competitive with elliptic curves.

A second area where Abelian varieties have had an impact on cryptography has been in the area of pairings for application to the numerous (and growing) number of pairing-based protocols. Only a brief mention of this work is given here. As for elliptic curve based systems it is possible to define pairings on $\mathcal{A}(\mathbb{F}_q)$ that map (for the Weil pairing), pairs of r-torsion points of $\mathcal{A}(\mathbb{F}_q)$ to the multiplicative subgroup of a finite field of appropriate order, say q^k, where k is the smallest extension for which this is possible (it is referred to as the embedding degree). Supersingular Abelian varieties are of interest here and these are isogenous to a power of a supersingular elliptic curve over the closure of $\mathbb{F}_q$. Simple Abelian varieties (i.e. those not isogenous to a product of lower degree varieties) are of interest. With a pairing for a supersingular elliptic curve, breaking the discrete logarithm problem has a complexity of the minimum of either the square root in the group order on the curve or that of using index calculus in the finite field, using the pairing. One thus balances carefully the group order in the curve with the complexity order from the finite field, which also depends on the embedding degree. However supersingular elliptic curves have a maximum embedding degree of 6. It is shown [8], [19], [20], [24] that supersingular Abelian varieties offer a much wider array of embedding degree choices and so should prove very useful for the implementation of such cryptosystems. A competing factor is that the complexity of computing the pairing increases as the embedding degree increases. Nonetheless, the wider range of choice of embedding degrees is a useful option in such protocols.

V. Conclusion

The review has considered recent work in coding and cryptography using higher dimensional varieties. It is difficult to draw conclusions why these applications have not enjoyed more success. In coding theory it is perhaps due to the lack of understanding as to how to choose the variety and construct effective subsets of points of the variety, along with a suitable set of evaluation polynomials, for use in the code definition, in order to obtain codes with good parameters. In cryptography the subject of hyperelliptic curves can be viewed in terms of the Jacobian variety, which has enjoyed success in a limited way. It would also be of interest to learn how to define point addition on an arbitrary variety, for use as a discrete logarithm problem in cryptography. In both coding and cryptography it is concluded there is much yet to be discovered on effectively using the higher dimensional varieties.

References

[1] L. Adleman, J. De Marais and M.-D. Huang, A sub-exponential algorithm for discrete logarithms over the rational subgroup of the Jacobians of large genus hyperelliptic curves over finite fields, Algorithmic Number Theory, ANTS-1, Lecture Notes in Computer Science, **877**, pp. 28–40, (1994).

[2] Ian F. Blake and Hamid Usefi, Rational points and codes on Hermitian hypersurfaces, in preparation.

[3] R.C. Bose and I.M. Chakravarti, Hermitian varieties in a finite projective space *PG(N, q)*, *Canad. J. Math.*, v. 18, pp. 1161–1182 (1966).

[4] R.C. Bose, On the application of finite projective geometry for deriving a certain series of balanced Kirkman arrangements, *Calcutta Math. Society*, pp. 341–354 (1958-59).

[5] D.G. Cantor, Computing in the Jacobian of an hyperelliptic curve, *Mathematics of Computation*, 48, pp. 95–101, (1987).

[6] H. Cohen and G. Frey, *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Boca Raton, FL, Chapman and Hall/CRC, 2006.

[7] C. Diem, On the discrete logarithm problem in elliptic curves, to appear, *Compositio Mathematica*.

[8] S. Galbraith, Supersingular curves in cryptography, ASIACRYPT 2001, *Lecture Notes in Computer Science*, No. 2248, C. Boyd (Ed.), pp. 495–513, (2001).

[9] P. Gaudry, An algorithm for solving the discrete log problem on hyperelliptic curves, EUROCRYPT 2000, *Lecture Notes in Computer Science*, No. 1807, B. Preneel (Ed.), pp. 19–34, (2000).

[10] P. Gaudry, E Thomé, N. Thériault and C. Diem, A double large prime variation for small genus hyperelliptic index calculus, *Math. Comp.*, **76**, no. 257, pp. 475–492 (2007).

[11] S.H. Hansen, Error-correcting codes from higher dimensional varieties, *Finite fields and their applications*, **7**, pp. 530–552 (2001).

[12] F.A. Izadi and K. Murty, Counting points on an Abelian variety over a finite field, INDOCRYPT 2005, *Lecture Notes in Computer Science*, No. 2904, T. Johansson and S. Maitra (Eds.), pp. 323–333, (2003).

[13] N. Koblitz, Hyperelliptic cryptosystems, *J. Cryptology*, **1**, pp. 139–150 (1989).

[14] G. Lachaud, Number of points of plane sections and linear codes defined on algebriac varieties, in: Arithmetic, Geometry and Coding, R. Pellikaan, M. Perret and S.G. Vladut (eds.), Walter de Gruyter: Berlin, pp. 77–104, 1996.

[15] G. Lachaud, The parameters of projective Reed-Müller codes, *Discrete Mathematics*, 81, pp. 217–221 (1990).

[16] J.B. Little, Algebraic geometry codes from higher dimensional varieties, arXiv:0802.2349v1 [cs.IT] 16 Feb 2008.

[17] V. Kuma Murty, *Abelian varieties*, CRM Monograph Series vol. 3, (1993).

[18] V. Kumar Murty, Abelian Varieties and cryptography, in: INDOCRYPT 2005, ed. C. Veni Madhavan, *Lecture Notes in Computer Science*, vol. 3797, pp. 1–15, 2005.

[19] K. Rubin and A. Silverberg, Using Abelian varieties to improve pairing-based cryptography, Journal of Cryptology, vol. 22, pp. 360–364, 2009.

[20] K. Rubin and A. Silverberg, Supersingular Abelian varieties in cryptology, in: CRYPTO 2002, ed. M. Yung, *Lecture Notes in ComputerS cience*, vol. 2442, pp. 336–353, 2002.

[21] A. Sboui, Second highest number of points of hypersurfaces in $\mathbb{F}_q^n$, Finite fields and their applications, **13**, pp. 444–449, (2007).

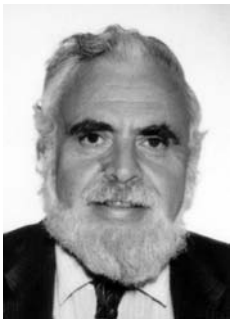
[22] Anders B. Sørensen, Rational points on hypersurfaces, Reed-Muller codes and algebraic geometric codes, Ph.D. Thesis, Aarhus, November, (1991).

[23] A. Weil, *Courbes algébriques et variétés Abéliennes*. Hermann, Paris (1971).

[24] H.J. Zhu, Group structures od elementary supersingular Abelian varieties over finite fields, *Journal of Number Theory*, vol. 81, pp. 292–309, 2000.

GOLOMB'S PUZZLE COLUMN™

Proofs With Integers as Exponents



Solomon W. Golomb

Each of the following problems has a solution using expressions of the form $\sum_{j=0}^{\infty} a_j x^j$ (or $\sum_{j=0}^{N} a_j x^j$ in Problems 2 and 4) where the coefficients a_j take only the values 0 or 1.

1) Find a function $f(x) = \sum_{j=0}^{\infty} a_j x^j$ such that $f(x)f(x^2) = 1/(1-x)$, valid in $|x| < 1$. (Equivalently, find a set S of non-negative integers such that every integer $n \geq 0$ has a unique representation $n = a + 2b$ with $a \in S$ and $b \in S$.)

2) Prove that, to pack an $a \times b \times c$ box exactly with $1 \times 1 \times n$ bricks, it is necessary (as well as sufficient) that the integer n divide (at least) one of the three integers a, b, or c.

3) For $k > 1$, it is possible to have k disjoint arithmetic progressions, $P_i = \{a_i n + b_i\}_{n=0}^{\infty}$, $1 \leq i \leq k$, whose union contains every positive integer exactly once (e.g., $\{2n+2\} \bigcup \{4n+1\} \bigcup \{4n+3\}$, with $k = 3$). Prove that this is impossible if all the a_i's are distinct.

4) Suppose that every integer k from 1 to $n^2 - 1$ has a unique representation of the form $k = a + b$, with $a \in A = \{0, a_1, a_2, \ldots, a_{n-1}\}$, $b \in B = \{0, b_1, b_2, \ldots, b_{n-1}\}$.

   4(a) Show, if n is prime, that the sets A and B of integers are unique (except for which is A and which is B).

   4(b) Show by example that if n is composite (e.g. for $n = 4$ and for $n = 10$), the sets A and B are not unique.

References

Problem 1 is based on problems 1 and 4 of the "Puzzles Column" in *Emissary*, published by the Mathematical Sciences Research Institute (MSRI), for Fall, 2010.

Problem 2, with the proof using exponents, is due to N.G. deBruijn.

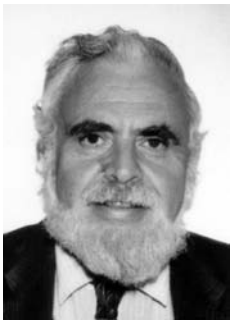
Problem 3, with proof via complex power series, is the Newman-Mirsky Theorem.

Problem 4 is a result I published in 1972, but was probably observed earlier.

GOLOMB'S PUZZLE COLUMN™

Derangements and Beyond Solutions

Solomon W. Golomb



From $d(n) = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!}$, it is clear that j divides all but the last $n-j+1$ terms. Thus,

1) $d(n) \equiv (-1)^n \frac{n!}{n!} \equiv (-1)^n \pmod{n}$.

2) $d(n) \equiv (-1)^n \left(\frac{n!}{n!} - \frac{n!}{(n-1)!} \right) \equiv (-1)^n (1-n) \equiv 0 \pmod{n-1}$.

3) $d(n) \equiv (-1)^n \left(\frac{n!}{n!} - \frac{n!}{(n-1)!} + \frac{n!}{(n-2)!} \right) \equiv (-1)^n (1 - n + (n^2 - n)) \equiv (-1)^n (n^2 - 2n + 1) \equiv (-1)^n (n(n-2)+1) \equiv (-1)^n \pmod{n-2}$.

A combinatorial proof of 2) is that the number of derangements on $(1, 2, 3, \ldots, n)$ that map 1 to k must be the same, by symmetry, for each k, $1 < k \leq n$; so $d(n)$ is a multiple of $n-1$.

4) $nd(n-1) + (-1)^n = n \sum_{k=0}^{n-1} (-1)^k \frac{(n-1)!}{k!} + (-1)^n = \sum_{k=0}^{n-1} (-1)^k \frac{n!}{k!} + (-1)^n \frac{n!}{n!} = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} = d(n)$.

5) The number of permutations on n objects with no 2-cycles is given, using inclusion/exclusion, by:

$$z(n) = n! - \binom{n}{2}(n-2)! + \frac{1}{2!}\binom{n}{2}\binom{n-2}{2}(n-4)! - \frac{1}{3!}\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}(n-6)! + \frac{1}{4!}\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}\binom{n-6}{2}(n-8)! - + \cdots = n! \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k}{2^k k!}.$$

6) Since $\sum_{k=0}^{\infty} \frac{(-1)^k 2^{-k}}{k!} = e^{-1/2}$, we find $z(n) \sim n!\sqrt{e}$ as $n \to \infty$. However, because the finite sum for $z(n)$ is truncated at $k = \lfloor n/2 \rfloor$, $z(n)$ is not the nearest integer to $n!/\sqrt{e}$.

7) A 3×3 "trerangement" is already a Latin square, and a 3×4 Latin rectangle can easily be completed (uniquely!) to a 4×4 Latin square. In fact, every $3 \times n$ Latin rectangle can be completed to an $n \times n$ Latin square.

8) $t(3) = 1$, $t(4) = 12$, $t(5) = 276$.

9) I am not aware of a general formula for $t(n)$, but let me know if you have found one.

10) Here are some anagrams as trerangements:

a. AELST	b. AEPRS	c. EIPRIST	d. EIMOPRS	e. AEGILNRT
LEAST	*PARSE*	*RIPEST*	*IMPOSER*	*ALERTING*
STEAL	*REAPS*	*STRIRE*	*PROMISE*	*INTEGRAL*
TALES	*SPEAR*	*TRIPES*	*SEMIPRO*	*TRIANGLE*

(Example e. was suggested by Donald Knuth.)

Other trerangements may be possible for a., b., and c.

ArXiv Postings Exceed 100/Month

In 2010 the number of articles posted to the Information Theory (cs.IT = math.IT) section of the arXiv preprint server exceeded a rate of 100 per month for the first time. This makes the IT section the second most active in both the CS and Math categories, following closely behind Discrete Mathematics/Combinatorics (cs.DM = math.CO), which had 1488 articles posted in 2010.

Since the initiation of the IT section in 2004, the number of articles posted has grown rapidly:

2004	2005	2006	2007	2008	2009	2010
41	331	365	614	791	962	1257

(This includes only primary articles, not cross-listed articles.)

On the other hand, an informal survey of the 81 articles that appeared in the November and December 2010 issues of the IEEE Transactions on Information Theory showed that only 36 (44%) had been posted previously on arXiv. "This shows that information theory is not yet like physics, where an article that has not been posted on arXiv effectively doesn't exist," said Joachim Rosenthal, who has been moderator of the math.IT section section since its inception. Madhu Sudan, moderator of the cs.IT section since inception, concurred: "We seem to be about halfway up the adoption S-curve, which is great, but we ought to keep striving for 100%."

In a related area, the IEEE Computer Systems Society has recently started a new arXiv section on Systems and Control (cs.SY). At the 2010 Conference on Decision and Control (CDC) in Atlanta, former IEEE Publications Vice-President John Baillieul and CS Society President Roberto Tempo organized a panel discussion on "E-print servers and traditional publishing," which included a segment on the IT Society's positive experience with arXiv. Past IT President G. David Forney, Jr. reported that "We have seen steady growth and noticeable benefits to our members from more rapid communication of new results, with no observable negative effects on our Transactions."

IEEE Information Theory Society Paper Award Call for Nominations

The Information Theory Society Paper Award shall be given annually for an outstanding publication in the fields of interest to the Society appearing anywhere during the preceding two **calendar years (2009–2010)**.

The purpose of this Award is to recognize exceptional publications in the field and to stimulate interest in and encourage contributions to fields of interest of the Society. The award consists of an appropriately worded certificate(s) and an honorarium of $1,000 for single author papers or $2,000 split equally among the authors of multiply authored papers.

Nomination Procedure (from the bylaws):

The Awards Subcommittee shall take into account

All nominations submitted in response to the open call for nominations in the last two years;

The nominations supplied by the Publications Committee in the last two years;

Any nomination that its members may want to submit for consideration.

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

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

Please send a brief rationale (limited to 300 words) for each nominated paper explaining its contribution to the field to the Society's First Vice President Muriel Medard (medard@mit.edu) by MARCH 15 2011.

Nominations for IEEE Medals and Recognitions

The IEEE is seeking nominations by 1 July 2011 for various IEEE Medals and Recognitions. For nomination guidelines and forms, visit http://www.ieee.org/awards.

Call for Papers

There will be a special session pertaining to neuroscience and information theory at the annual International Symposium on Information Theory (ISIT) 2011 in St Petersburg, Russia (http://www.isit2011.org/).

This is taking place due to

(i) increasing interest in the intersection of these disciplines

- a special issue on Information Theory in Molecular Biology and Neuroscience, IEEE Transactions on Information Theory, February 2010
- a special issue on Methods of Information Theory in Neuroscience Research, Journal of Computational Neuroscience, June 2010
- 5 years of special sessions on "Methods of Information Theory in Computational Neuroscience" at the Computational Neuroscience Annual Meeting (CNS)

(ii) a coincidence that provided an opportunity: the Computational Neuroscience Annual Meeting (CNS) will take place the week before ISIT, in Stockholm, Sweden (see here http://www.cnsorg.org/2011/ for more details). Moreover, a special session there, titled "Methods of Information Theory in Computational Neuroscience", will take place on Friday, July 29th. The link to the 2010 session is here: http://www.bionet.ee.columbia.edu/workshops/cns/methods10/methods10.html

The special session on neuroscience and information theory at ISIT will be of 3 hours in duration, consisting of 4 or 5 talks of 35-40 minutes in duration. Each researcher in computational/experimental neuroscience will give an overview of their work and will speak to the potential synergies with information theory. The detailed scheduling of this ISIT session is to be determined, but will most likely take place during Monday/Tuesday at ISIT - to provide overlap with the CNS conference.

The hope is that intrigued researchers on both sides of the disciplines might be able to use the same trip to Europe to take part in both conferences and further identify the interplay between both. Such computational /experimental neuroscientists with preliminary results that are appropriate for ISIT are encouraged to submit a paper by the standard ISIT Feb 15, 2011 deadline (http://www.isit2011.org/authors_call.php). Analogously, information theorists with neuroscience results that are appropriate for CNS are encouraged to submit an abstract to the CNS annual meeting by the Feb 14, 2011 deadline (http://www.cnsorg.org/2011/dates.shtml).

The links to the two conferences are here:

CNS: http://www.cnsorg.org/

ISIT: http://www.isit2011.org/

For anyone with further questions, feel free to contact

- Todd Coleman (colemant@illinois.edu)
- Aurel Lazar (aurel@ee.columbia.edu)


The Eighth International Symposium on Wireless Communication Systems

ISWCS 2011

Eurogress Aachen, Germany, November 6 - 9, 2011
URL: http://www.ti.rwth-aachen.de/iswcs2011

Organizing Committee

General Chair
Rudolf Mathar
RWTH Aachen University

Technical Program Chairs
Gerhard Kramer
Technische Universität München
Kwang-Cheng Chen
National Taiwan University

General Co-Chairs
Paul D. Mitchell
Rodrigo C. Lamare
The University of York, UK

Steering Committee
Gerhard Fettweis
Technische Universität Dresden
Boon Sain Yeo
SensiMesh, Singapore
Yuming Jiang
NTNU, Norway

Publicity Chair
Chin-Tser Huang
University of South Carolina

Special Session & Tutorial Chair
Anke Schmeink
RWTH Aachen University

Finance Chair
Meik Dörpinghaus
RWTH Aachen University

Publications Chair
Melanie Neunerdt
RWTH Aachen University

Technical Co-Sponsors

IEEE
IEEE COMMUNICATIONS SOCIETY
IEEE Information Theory Society
VDE ITG

Wireless Communications is at the center of a new and passionate era characterized by smart and flexible transceiver concepts, the convergence of systems and technologies, a transition towards all-IP networks and the development of technologies with a user-centric focus. In this context, the International Symposium on Wireless Communication Systems (ISWCS) is positioning itself as a recognised and dynamic forum for researchers and engineers from academia and industry to present and discuss original ideas and contributions in all fields related to mobile wireless communication systems.
The aim of this symposium is to present novel contributions in the form of tutorials, panel discussions, keynote speeches, technical papers, posters and testbed implementations. ISWCS'11 seeks to address and capture highly innovative and state-of-the-art research from academia, the wireless industry and standardization bodies. The scope of the conference includes a wide range of technical challenges encompassing wireless communications, quality of service support, wireless networking, signal processing, cross-layer air interface design, wireless broadband access and cooperative communication.

Topics of interest include but are not limited to

- Cooperative communication and relaying
- Cognitive radio
- Radio resource management
- Wireless access techniques
- Cross-layer air interface
- Information and communication theory
- Bio-inspired communications
- Signal processing, including VLSI architectures
- Estimation and detection
- Coding and modulation
- Network coding
- Compressive sensing
- MIMO communications
- Multi-carrier systems, OFDM
- Spread spectrum and UWB
- Antennas and propagation
- Self organizing networks
- Ad-hoc, mesh and sensor networks
- Mobility management and modeling
- QoS provisioning
- Wireless network architectures and technologies
- Smart grid communications
- Innovative services and applications
- Wireless privacy and security

Paper Submission Guidelines

Acceptance will be based on full papers. Accepted and registered papers will be published in the conference proceedings and made electronically available. Papers must be submitted via EDAS by May 30, 2011. Submission guidelines will be downloadable from the symposium web page.

Special Sessions and Tutorials

Proposals for tutorials and special sessions should provide a 200 word summary. For special sessions additionally the details of the invited papers are requested.

Important Dates

Submission deadline special sessions: Monday, May 2, 2011
Submission deadline tutorials: Monday, May 2, 2011
Submission deadline full papers: Monday, May 30, 2011
Notification of acceptance: Monday, August 1, 2011
Camera-ready version due: Monday, September 12, 2011

FORTY-NINTH ANNUAL ALLERTON CONFERENCE

ON COMMUNICATION, CONTROL, AND COMPUTING

September 28 – 30, 2011

Preliminary Call for Papers

The Forty-Ninth Annual Allerton Conference on Communication, Control, and Computing will be held from Wednesday, September 28 through Friday, September 30, 2011, at Allerton House, the conference center of the University of Illinois. Allerton House is located twenty-six miles southwest of the Urbana-Champaign campus of the University in a wooded area on the Sangamon River. It is part of the fifteen-hundred acre Robert Allerton Park, a complex of natural and man-made beauty designated as a National natural landmark. Allerton Park has twenty miles of well-maintained trails and a living gallery of formal gardens, studded with sculptures collected from around the world.

Papers presenting original research are solicited in the areas of communication systems, communication and computer networks, detection and estimation theory, information theory, error control coding, source coding and data compression, queueing networks, control systems, robust and nonlinear control, adaptive control, optimization, dynamic games, large-scale systems, robotics and automation, manufacturing systems, discrete event systems, intelligent control, multivariable control, computer vision-based control, learning theory, neural networks, VLSI architectures for communications and signal processing, and automated highway systems.

Plenary lecture: Professor Avi Wigderson of the Institute for Advanced Study, Princeton University, will deliver this year's plenary lecture. It is scheduled for Friday, September 30, 2011.

I**nformation for authors:** Regular papers suitable for presentation in twenty minutes are solicited. Regular papers will be published in full (subject to a maximum length of eight 8.5" x 11" pages, in two column format) in the Conference Proceedings.

For reviewing purposes of papers, a title and a five to ten page extended abstract, including references and sufficient detail to permit careful reviewing, are required.

Manuscripts must be submitted by **Wednesday, July 6, 2011**, following the instructions at the Conference website: http://www.csl.uiuc.edu/allerton/**.**

Authors will be notified of acceptance via e-mail by August 5, 2011, at which time they will also be sent detailed instructions for the preparation of their papers for the Proceedings.

Final versions of papers to be presented at the conference must be submitted electronically by September 30, 2011.

Conference Co-Chairs: Sean Meyn and Bruce Hajek

Email: *allerton@csl.uiuc.edu* URL: http://www.csl.uiuc.edu/allerton

COORDINATED SCIENCE LABORATORY AND THE
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

University of Illinois at Urbana-Champaign

ANNOUNCEMENT AND CALL FOR PAPERS

Information: http://icmcta.uab.cat or icmcta@deic.uab.cat

3 ICM TA

3rd International Castle Meeting on Coding Theory and Applications

Universitat Autònoma de Barcelona

We would like to invite you all to participate in the third International Castle Meeting on Coding Theory and Applications, which will be held in the Castell de Cardona, Barcelona. The previous two editions of this meeting took place in Castillo de la Mota, Medina del Campo, Valladolid in 1999 and 2008.

The Castell de Cardona is arguably the most important medieval fortress in Catalonia. It is situated on a hill overlooking the river valley of the Cardener and the town of Cardona. The fortress was initially constructed by Wilfred the Hairy in 886. It is in both the Romanesque and Gothic styles, and includes the so-called Sala Dorada and Sala dels Entresols. During the 14th century, the dukes of Cardona came from the most important family of the Crown of Aragon, which was second only to the royal house. Because of this, they were called "kings without crowns," as they had extensive territories in Catalonia, Aragon, and Valencia, and dynastic ties with Castile, Portugal, Sicily, and Naples. During the Spanish War of Succession, the Castell de Cardona was the last redoubt of the supporters of Charles VI of Austria, before being occupied by Bourbon troops, which supported Philip V.

TOPICS: Algebraic-geometric codes, codes and combinatorial structures, codes and graphs, group codes, network coding, quantum codes, turbo codes and applications based on coding theory. Other related topics can also be considered at the discretion of the scientific committee.

INVITED SPEAKERS:

Robert Calderbank	*Princeton University, USA*
Tuvi Etzion	*Technion IIT, Israel*
Marcus Greferath	*University College Dublin, Ireland*
Jennifer D. Key	*Clemson University, USA*

SUBMISSIONS: Those wishing to contribute a 20 minute talk are invited to submit a 4-6 page extended abstract. Further information about the electronic submission will be posted at the meeting web site. Submitted extended abstracts should be of sufficient detail to be reviewed by experts in the field.

FINAL PROCEEDINGS: The book of proceedings (with ISBN) will be published by *Servei de Publicacions UAB*. It will be provided to the participants during the meeting and it will contain the accepted extended abstracts guaranteed to be presented at the meeting.

FULL PAPERS: Authors of accepted abstracts will have the option to submit a full version. After a second thorough refereeing process, the best papers will appear in a special issue of the journal *Designs, Codes and Cryptography*.

LANGUAGE: The official language will be English.

DEADLINES:

Abstract submission by:	**May 1, 2011**
Notification of decision:	June 1, 2011
Opening of the registration:	June 2, 2011
Deadline for early registration:	June 16, 2011
Final version abstract submission by:	June 16, 2011
Full paper submission:	To be announced.

STEERING COMMITTEE:

Ángela Barbero	*UVa Valladolid, Spain*
Øyvind Ytrehus	*UiB Bergen, Norway*

ORGANIZING COMMITTEE:

Joaquim Borges	(co-chair of the Scientific Committee)
Cristina Fernández	(co-chair of the Local committee)
Jaume Pujol	(co-chair of the Local committee)
Josep Rifà	(chair of the Steering Committee)
Mercè Villanueva	(co-chair of the Scientific Committee)

SCIENTIFIC COMMITTEE:

Peter Beelen	*DTU, Denmark*
Maria Bras-Amorós	*URV Tarragona, Spain*
Juergen Bierbrauer	*Michigan Tech, USA*
Claude Carlet	*Univ. Paris VIII, France*
Gerard Cohen	*Univ. Paris VI, France*
Italo Dejter	*UPR Puerto Rico, USA*
Alexandros Dimakis	*USC California, USA*
Steven Dougherty	*Univ. Scranton Pennsylvania, USA*
Iwan Duursma	*UIUC Illinois, USA*
Cristina Fragouli	*EPF Lausanne, Switzerland*
Tom Høholdt	*DTU, Denmark*
Heeralal Janwa	*UPR Puerto Rico, USA*
San Ling	*NTU, Singapore*
Edgar Martínez-Moro	*UVa Valladolid, Spain*
Patric Östergard	*Aalto Univ., Finland*
Kevin Phelps	*AU Alabama, USA*
Ruud Pellikaan	*TUe Eindhoven, Netherlands*
Ángel del Río	*Univ. Murcia, Spain*
Eirik Rosnes	*UiB Bergen, Norway*
Patrick Solé	*CNRS-ENST Paris, France*
Emina Soljanin	*Bell Labs New Jersey, USA*
Faina Soloveva	*RAS Novosibirsk, Russia*
Alexander Vardy	*UCSD California, USA*
Jos Weber	*TU Delft, Netherlands*
Victor Zinoviev	*RAS Moscow, Russia*

LOCAL COMMITTEE:

Muhammad Bilal	*UAB Barcelona, Spain*
Bernat Gastón	*UAB Barcelona, Spain*
Jaume Pernas	*UAB Barcelona, Spain*
Lorena Ronquillo	*UAB Barcelona, Spain*
Albert Sánchez	*UAB Barcelona, Spain*

ITW 2011 - Paraty, Brazil

IEEE Information Theory Workshop - October 16-20, 2011

Call for Papers

The **2011 IEEE Information Theory Workshop** (ITW 2011) will take place on **October 16-20, 2011 in Paraty, Brazil**. Paraty is a quaint historical village, founded in 1597, located on the Brazilian coast, at about equal distances from Rio de Janeiro and São Paulo. It was the port from where gold was shipped to Portugal, during Brazil gold rush. The buildings and streets of Paraty have retained their colonial atmosphere and so have its beautiful surroundings with exuberant tropical forest and islands which provide excellent opportunities for easily accessible tours. The intimate town environment is conducive to rich exchanges among workshop participants. In addition to oral and poster technical sessions, the Workshop will feature keynote presentations, special sessions and panel discussions. It will be held at "Casa de Cultura", a 1754 restored building. Besides regular intercity bus service (4 hour trip), transfers will be arranged from and to São Paulo and Rio de Janeiro airports.

The scope of the Workshop includes, but is not limited to, the following topics:

- Graph-based codes and iterative decoding
- Physical-Layer security
- Distributed source and channel coding
- Coding for wireless systems
- Compressed sensing
- Codes, lattices and cryptography
- Multi-terminal information theory
- Quantum computing and coding
- Information Theory in Biology

There will be both invited and contributed sessions. Papers for the contributed sessions, up to five pages and following the webpage guidelines, are solicited.

Important Dates:

Paper Submission	Notification of Acceptance	Camera Ready Submission
May 1st, 2011	July 17, 2011	August 14, 2011

Plenary Speakers

Daniel Bernstein **Ezio Biglieri** **Kannan Ramchandran** **Neil Sloane**

Invited Sessions

Codes, Lattices and Cryptography
Organizer: Paulo Barreto

Graph-Based Codes and Iterative Decoding
Organizer: Daniel Costello

Coding Theory
Organizer: Frédérique Oggier

Multi-Terminal Information Theory
Organizer: Suhas Diggavi

Compressed Sensing
Organizer: Olgica Milenkovic

Physical-Layer Security
Organizer: Matthieu Bloch

Panel

New Perspectives for Information Theory and Coding

Chair : Sergio Verdú

For more information, please visit the workshop website at
http://www.fee.unicamp.br/itw2011

General Co-Chairs
Amin Shokrollahi
Valdemar C. da Rocha Jr.
Sueli I. R. Costa

Advisory Committee
Vinay Vaishampayan
Reginaldo Palazzo Jr.
Weiler A. Finamore
Ricardo Dahab

Technical Program Committee
João Barros
Max H. M. Costa
Jaime Portugheis
(Co-Chairs)

Matthieu Bloch
Holger Boche
Giuseppe Caire
Dariush Divsalar
Sam Dolinar
Michael Gastpar
Vivek Goyal
Frank Kschischang
Nicholas Laneman
Simon Litsyn
Angel Lozano
Muriel Médard
Thomas Mittelholzer
Chandra Nair
Prakash Narayan
Anderson Nascimento
Daniel Panario
Cecílio Pimentel
Bixio Rimoldi
Miguel Rodrigues
Nicolas Sendrier
Danilo Silva
Patrick Solé
Emina Soljanin
Yossef Steinberg
Leandros Tassiulas
Andrew Thangaraj
David Tse
Daniela Tuninetti
Han Vinck
Emanuele Viterbo
Stefan Wolf

Financial Co-Chairs
Charles C. Cavalcante
Marcelo S. Pinho

Publications
Renato Baldini Filho
Gustavo Fraidenraich

Publicity
Renato Lopes

Web
Franz Pietz

Conference Calendar

DATE	CONFERENCE	LOCATION	WEB PAGE	DUE DATE
March 23-25, 2011	45th Annual Conference on Information Sciences and Systems (CISS 2011)	Johns Hopkins University, MD	http://ciss.jhu.edu	Passed
April 10-15, 2011	2011 IEEE Conference on Computer Communications (INFOCOM 2011)	Shanghai, China	http://www.ieee-infocom.org	Passed
May 14-17, 2011	2011 IEEE Vehicular Technology Conference (VTC2011-Spring)	Budapest, Hungary	http://www.ieeevtc.org/vtc2011spring/	Passed
May 24-26, 2011	2011 The Sixth Conference on Theory of Quantum Computation, Communication and Cryptography (TQC 2011)	Madrid, Spain	http://gcc.ls.fi.upm.es/tqc2011	Passed
June 5-9, 2011	2011 IEEE International Conference on Communications (ICC 2011)	Kyoto, Japan	http://www.ieee-icc.org	Passed
June 20-22, 2011	2011 IEEE Communication Theory Workshop (CTW 2011)	Sitges, Catalonia, Spain	http://www.ieee-ctw.org	March 13, 2011
July 31-August 5, 2011	2011 IEEE International Symposium on Information Theory (ISIT 2011)	St. Petersburg, Russia	http://www.isit2011.info	Passed
September 28-30, 2011	49th Annual Allerton Conference on Communications, Control, and Computing	Monticello, IL	TBD	July 6, 2011
October 16-20, 2011	2011 IEEE Information Theory Workshop (ITW 2011)	Paraty, Brazil	http://www.fee.unicamp.br/itw2011	April 10, 2011
November 6-9, 2011	8th International Symposium on Wireless Communication Systems (ISWCS 2011)	Aachen, Germany	http://www.ti.rwth-aachen.de/iswcs2011/	May 30, 2011

Major COMSOC conferences: http://www.comsoc.org/confs/index.html