IEEE Information Theory Society Newsletter

Vol. 58, No. 4, December 2008 Editor: Daniela Tuninetti ISSN 1059-2362

Source Coding and Simulation

XXIX Shannon Lecture, presented at the 2008 IEEE International Symposium on Information Theory, Toronto Canada

Robert M. Gray

Prologue

A unique aspect of the Shannon Lecture is the daunting fact that the lecturer has a year to prepare for (obsess over?) a single lecture. The experience begins with the comfort of a seemingly infinite time horizon and ends with a relativity-like speedup of time as the date approaches. I early on adopted a few guidelines: I had (1) a great excuse to review my more than four decades of information theoretic activity and historical threads extending even further back, (2) a strong desire to avoid repeating the topics and content I had worn thin during 2006–07 as an intercontinental itinerant lecturer for the Signal Processing Society, (3) an equally strong desire to revive some of my favorite topics from the mid 1970s—my most active and focused period doing unadulterated information theory, and (4) a strong wish to add something new taking advantage of hindsight and experience, preferably a new twist on some old ideas that had not been previously fully exploited. The result was a new look at an old problem—the connections between classic Shannon source coding subject to a fidelity criterion and the "simulation problem" formulated as part of a proof of the source coding theorem for trellis encoding using a time-invariant trellis.

The problem and several questions are naturally motivated by the contiguous diagrams of the two systems in Figure 1, and the discussion provides a tour of operational distortion-rate functions, block and sliding-block coding, process distortion and distance measures, and some fundamental shared ideas of information theory and ergodic theory.

Source Coding and Simulation

Both systems shown in Figure 1 concern an information source $X = \{X_n; n \in \mathcal{Z}\}$, a discrete-time stationary ergodic random process described by a process distribution μ. The random variables X_n take values in an *alphabet* A_X, which might be discrete, continuous, or mixed. The focus here is on stationary and ergodic, but many of the results generalize to nonergodic sources (using the ergodic decomposition) and to asymptotically mean stationary sources (sources for which sample averages converge).

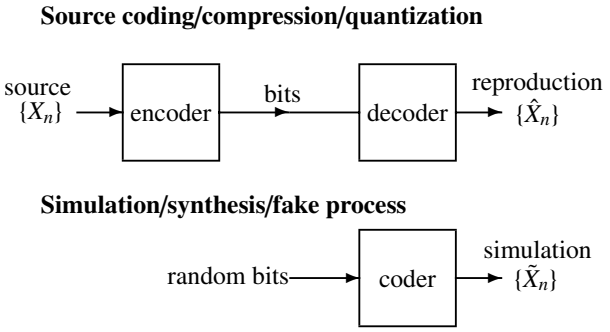


Figure 1: Source coding and simulation.

The goal of source coding [1] is to *communicate* or *transmit* the source through a constrained, discrete, noiseless communication or storage channel to a receiver, whose job is to reconstruct or reproduce the original source as accurately as possible given constraints on the channel. As is common, we will focus on the situation where the channel sends binary symbols or *bits* of information, say R bits per input symbol. For simplicity the special case of $R = 1$ will be emphasized here. The goal of simulation [2] is to *simulate* the source using a system of a particular structure, for example using a sequence of independent fair coin flips to drive a time-invariant or stationary coder to produce a random process with distribution closely matching that of the original source X in some precise sense. There are several reasons for considering a stationary operation on the input bits rather than some form of block coding:

- A stationary mapping of coin flips (or any iid source) will produce a process that is stationary and ergodic, as is the target source X we are trying to emulate. A block mapping will not be stationary or ergodic; it will be block stationary, but not necessarily block ergodic. Even if "stationarized" by a random start, the resulting process will retain in general periodicities not present in the original source.

continued on page 5

From the Editor

Daniela Tuninetti

Dear IT society members,

I trust you all had a good start of the fall semester. By the time this issue arrives on your desk, you will already be preparing finals and planning winter vacation. Although summer may appear long gone, this issue will bring back warm memories of ISIT 2008 in Toronto, Canada. I would like to express our thanks to co-chairs Frank R. Kschischang and En-hui Yang, and their team, for the organization of a great ISIT. Before I give you a taste of what you will find in this issue, I would also like to thank our president Dave Forney, whose term will end in December 2008, and welcome our new president, Andrea Goldsmith, whose term will start in January 2009. Please join me in congratulating Dave and welcoming Andrea.

This issue opens with a summary of Robert M. Gray's Shannon Lecture "Source Coding and Simulation." Then our regular columns by our president Dave Forney, our historian Anthony Ephremides, and our creative puzzle maker Sol Golomb will follow. You will then read first hand from Frank R. Kschischang and En-hui Yang about the major ISIT 2008 events.

You will find the reflections on the paper award announced at the Award Luncheon in Toronto. The winners of the 2008 IT Paper Award are two 2006 IT Transactions papers on Compressed Sensing by David Donoho, and by Emmanuel Candes and Terence Tao. The winner of the 2008 Joint IT / ComSoc Paper Award is a Communications Transactions paper on Accumulate-Repeat-Accumulate Codes by Aliazam Abbasfar, Dariush Divsalar and Kung Yao. Congratulations to all the authors on their paper award. The summaries of the paper awards will be followed by the summaries of the plenary talks: "Golay, Heisenberg and Weyl" by A. Robert Calderbank, "Building Quantum Computers" by Emanuel H. Knill, and "Randomness - A Computational Complexity View" by Avi Wigderson.

You will also enjoy an update on the activities of the student committee chaired by Aylin Yener, on the newly formed outreach committee chaired by Muriel Medard, and an account of the International Symposium on Advances in Communications on the occasion of Vijay K. Bhargava's 60th birthday. Congratulations Vijay on your birthday and all your achievements.

Last but not the least, you will find our NSF guest column by Program Manager Sirin Tekinay. This will be Sirin's last column as her term at NSF ended in September 2008. She did a terrific job as champion for our community.

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

Issue	**Deadline**
March 2009	January 10, 2009
June 2009	April 10, 2009

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

I may be reached at the following email address: danielat@uic.edu.

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

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

David Forney

It has been a great honor and privilege to be President of the IEEE Information Theory Society during 2008. It is a classy and successful organization, full of enjoyable colleagues, which always aims for the highest standards.

In January 1973, when I was Editor of the IEEE Transactions on Information Theory, I published the Editorial below in connection with a series of 25th-anniversary review articles. As a test of its claim of timelessness, on the occasion of the 60th anniversary of information theory, I take the liberty of publishing it again.

One of the Editor's few perquisites is the prerogative of publishing his own opinions from time to time, uninvited. I have been exposed to the cry that information theory is dead from the time I first entered graduate school at M.I.T., in 1961. Yet I think it will be evident from this series of papers, if not already from the reader's own experience, that a lot has happened in our field since 1961, perhaps more than happened before then (excluding Shannon). We are sometimes led astray by the unique history of information theory: created almost full grown from the brow of Zeus, it was one of the first disciplines to achieve the status of a fad, the passage of which era we can hardly regret. Now, as an increasingly mature field, it is not exempt (if it ever was) from the laws that govern the evolution of all sciences. New and basic ideas are always rare. The research frontier becomes ever more esoteric. Much work is done that does not speak to today's real needs, or the future's. Yet work proceeds that will have the cumulative effect over the course of time of completely changing the outlines of the field, its practice, and our understanding of it. Which work falls in which category is not always immediately clear.

Without great daring, one can make a few predictions, valid now or any time.

1) The field will not die, but it will change. It has the good fortune after all to be an engineering field, linked to technologies that are expanding with great vigor, and that yearly make feasible and economic ever more sophisticated kinds of information processing. It also has an intellectual beauty that will continue to attract some of the very best students.

2) Ten years from now the subjects of papers in this Transactions will be largely different from (and more abstruse than) today's; today's contributors will be complaining that the Transactions has become unintelligible and irrelevant.

3) Twenty-five years from now the present will be seen as a rather ignorant and uninformed time, full of misguided, naive, and trivial work; yet at the same time our successors will perceive in this era the laying of the foundations for the "information theory" of that day, and will envy us for working at a time when so much was fresh and new.

Call for Nominations

IEEE Information Theory Society Claude E. Shannon Award

This Shannon Award Committee decided to issue an open call for nominations for next year. The IEEE Information Theory Society Claude E. Shannon Award is given annually or *consistent and profound contributions to the field of information theory.*

Award winners are expected to deliver the Shannon Lecture at the annual IEEE International Symposium on Information Theory held in the year of the award.

NOMINATION PROCEDURE: Nominations and letters of endorsement must be submitted by March 1 to the current President of the IEEE Information Theory Society. (In 2009 the President will be Prof. Andrea Goldsmith, Packard 371, Stanford CA 94305, andrea@wsl.stanford.edu.) Please include:

Nominee:
1. Full name

2. Address, e-mail and telephone

3. Professional affiliation

4. Academic degrees (awarding institutions and dates)

5. Employment history

6. Principal publications (not more than ten).

7. Principal honors and awards

8. **Optional**. The nominee's CV may be submitted as an addendum to the nomination.

Nominator:
1. Full name

2. Address, Email and telephone

3. **Optional**. Endorser(s), names and addresses, e-mail addresses. (Letters of endorsement are not required. At most three may be submitted.)

Rational:
Discussion of how the nominee has made *consistent and profound contributions to the field of information theory* (not more than two pages).

continued on page 49

The Historian's Column

Anthony Ephremides

This column was written before the November election in the United States and amidst a gathering financial crisis that started enveloping the world. So, the urge to align my comments to the prevailing zeitgeist was too strong to resist. Thus, by the time you read this, either my predictions will have been confirmed or I will have missed the mark. In the latter case, because of the nature of my thoughts, nobody would care (including myself).

So, the bottom line of my prediction is that the storm that started during late September and went on through the month of October will subside by the time this column appears. This does not mean that there will not be serious lingering problems. It simply means that the global boat will have survived another (giant) wave. This prompts me to reminisce about similar events of the past from the perspective, of course, of our profession and our field.

Well, let us start with 1970-71. This is the year during which I was finishing my Ph.D. Dissertation and was actively looking for a job. The Dow-Jones Industrial index was hovering around 700 points. And there was gloom all around. The corridors where the graduate students unleashed their anxieties on each other were awash with rumors that there were no more academic jobs to be found. The economy was faltering under Richard Nixon's first term while an unpopular war was eroding the image of America in the rest of the world. The good old times were over. I was looking at my senior fellow students (like Stamatis Cambanis and Elias Masry) who had obtained academic job offers and I was thinking with envy that the door to academia had been shut behind them. You, lucky you! What would be our professional future? I still vividly recall talking to John Thomas (my advisor) and expressing my worries to him. And I remember his calm face and soft voice as he was reassuring me. "This is just one of those ups-and-downs," he had said.

After sleepless nights and several job applications it came as a pleasant surprise that there were invitations for interviews and subsequent job offers to several Universities. With a sigh of relief I sat down and luxuriated in deciding which of the four job offers to accept. I chose the University of Maryland (where I just completed 37 years).

My first few years saw additional ups-and-downs. Price controls were imposed by a Republican President. Those who viewed the measure as a step toward socialism had their weapons defused by the fact that the commander-in-chief was beyond suspicion of being a crypto-communist. There were rumors that promotions and tenure would be impossible from then on. The research dollars from the National Science Foundation were dwindling. The war was still going on. And then another big crisis erupted in the Middle East in 1973. And the first major gasoline crisis followed. We had to wait in line for as long as 30 minutes to fill up the tanks of our automobiles. The price of gasoline was pushing $1 per gallon (which at today's prices would be about $10 per gallon). The future did not look so good.

And then the war ended. The price controls were lifted. The gas lines disappeared. New research programs were launched. There were openings in academia. Industry was hiring. Coding Theory started being used out there. A wave of deregulation was introduced as Ronald Reagan was proclaiming "This is an energy-rich country". The notion of a "yuppie" emerged. A yuppie had oiled hairdos, thin-rim glasses, and donned yellow ties. The Wall Street wizards started proliferating. World Airways was the first low-cost airline that led the industry in destroying the notion of service to the passenger. But it lowered the ticket prices too. So the world started humming again. Cellular phones were born. There was euphoria in the stock market.

And then the first Iraq war erupted after Saddam Hussein's invasion of Kuwait. The economy started spiraling down after a brief up turn. Bush senior lost in 1992 to a little-known governor from Arkansas. The slogan of the times was: "It's the economy, stupid!" Bright Ph.D. graduates could not find academic jobs again. Students were worried. It was getting tougher to obtain permits to work and live in the United States. Once again there was gloom in the corridors where the graduate students roamed. Now it was my turn to tell my anxious students "This is just one of those ups-and-downs".

Sure enough, despite growing unrest and transformational cracking sounds around the world, the Iron Curtain had been torn, the Berlin wall was down, and China started sending lots of inexpensive goods to the world. More openings in academia, more hiring in industry, more new research programs. The golden jubilee of Information Theory was celebrated amidst a new euphoria. Many people were getting rich. You could borrow money with zero interest. The Wall Street wizards were now sporting pink ties.

And then the bust of the telecom industry bubble occurred, followed soon by the infamous attack of 9/11. The budget surpluses started nose-diving. The cracks around the world continued in the manner of pine trees bending under hot dry winds under a blazing sun. The second Iraq war was launched after the incursion into Afghanistan. Almost all countries became members of either NATO or the European Union. Technology continued to boom. Gasoline prices started going up, followed by shortages, global warming, and unrest.

Thus, we arrived to where we are today. The students again are in agony. Enrollments are down. The stock market is catapulting downward. I did not have the nerve to tell my students: "This is just one of those ups-and-downs". But I wanted to believe it. We have to go on. A new era of exuberance should dawn soon. Perhaps!

Source Coding and Simulation *continued from page 1*

- Stationary mappings of iid processes form the arguably most important class of random processes in ergodic theory, as will be discussed. As described in Ornstein's classic exposition [3] of the revolution in ergodic theory that he led in the late 1960s and early 1970s, a fundamental question in the study of random processes is the generality of the class he named *B*-processes (*B* for Bernoulli)—the class of processes which can be modeled as stationary codings of roulette wheels, that is, stationary mappings of iid processes.[1]

Although the focus of ergodic theory was originally on processes with discrete alphabets, the notion of *B*-processes and their principal properties (to be considered shortly) were extended to continuous alphabets [4], so for example linear models formed by stable time-invariant linear filtering of white Gaussian noise are *B*-processes [5]. Hence allowing simulators to include the case of a continuous alphabet iid process driving a stable time-invariant filter, effectively allowing an infinite number of bits per symbol, permits the class of simulator output processes to include the Gaussian autoregressive models popular in speech processing and more general linear models. Our primary interest here, however, is generating simulations with a finite bit rate constraint.

The two systems have an obvious and suggestive similarity, the source coding system from the encoder output through the reproduction looks like a simulation system, random bits are coded to form a process that "resembles" the original source. The notion of "resembles" differs in that in the source coding case the final process should match the original source sequence in some sense, while in the simulation system there is no waveform matching, we need only produce something with similar statistical characteristics. This difference parallels the difference in speech coding between waveform coding, such as ADPCM, which tries to match the input sampled speech waveform, and LPC, which effectively synthesizes or simulates a process at the decoder which has second order statistics (correlation or spectra) which match those of the original speech. This simulation problem was first introduced in 1977 [2] where it was observed that a simulation system yields a source coding system by using the simulator to "color" a trellis which could be searched by a minimum distortion search such as a Viterbi algorithm in order to produce a trellis encoded source coding system. Quantifying the optimal performance of the two systems demonstrated that the optimal simulation performance provided an upper bound to the optimal source coding performance. It was further shown that if the original source was a *B*-process, this bound actually yields the optimal source coding performance as given by the Shannon distortion rate function.

[1]Such processes are also referred to as simply Bernoulli in the ergodic theory literature, but we avoid this usage because of the potential confusion of the term with the engineering literature, where it means coin flips.

[2]We assume that input blocks are mapped into output blocks of the same dimension for simplicity, which means that the discrete time symbols correspond to the same quantity of continuous time. Different block sizes can be handled with more cluttered notation.

An obvious difference between the two systems is that the simulation system assumes an iid input, but the bit stream resulting from encoding a stationary and ergodic source is in general not iid. There is, however, a "folk theorem" to the effect that if the overall source coding system has performance near the Shannon optimum, then the bit stream should be "nearly iid." Results along this line have been proved for block almost lossless coding of iid sources (e.g., [6]), but it turns out that the bits produced by nearly optimally encoding a source are close to coin flips in a very precise sense: Ornstein's $\bar{d}$-distance between the two binary processes is small. The remainder of this paper is devoted to detailing these results and some of the issues involved and providing some thoughts on the implications and possible future paths for investigation.

Preliminaries

Random vectors produced by a source X will be denoted by $X^N = (X_0, X_1, \cdots, X_{N-1})$, with distribution μ^N. The Shannon entropy of a random vector is given as usual by

$$H(X^N) = H(\mu^N)$$

$$= \begin{cases} -\sum_{x^N} \mu^N(x^N) \log \mu^N(x^N) & A_X \text{ discrete} \\ \infty & \text{otherwise} \end{cases}$$

Both notations, one emphasizing the random vector and the other its distribution, will be useful. The Shannon entropy (rate) of the stationary random process X is given by

$$H(X) = H(\mu) = \inf_N H(X^N)/N = \lim_{N\to\infty} H(X^N)/N.$$

Other information measures and information rates, such as average mutual information, will also be considered, but we leave their definitions to the literature.

Block and Sliding-Block Coding

The formulation of both source coding and simulation is in terms of codes, and the nature of the results and their interpretation differs according to what kind of codes are considered. Two basic coding structures for coding a process X with alphabet A_X into a process Y with alphabet A_Y are of interest here. Block coding, the standard approach in information theory, maps each nonoverlapping block of source symbols into an index or block of encoded symbols (e.g., bits). Sliding-block coding, the standard coding method in ergodic theory, maps overlapping blocks of source symbols into single encoded symbol (e.g., a single bit). A block code is described by a vector mapping $\mathcal{E} : A_X^N \to A_Y^N$ (or other index set), where N is the block length.[2] A sequence is encoded by applying the vector mapping to a sequence of *nonoverlapping* input vectors. A sliding-block code is described by a vector-to-scalar mapping $f : A_X^N \to A_Y$, and a sequence is encoded by "sliding" the coding window a single symbol at a time in order to produce a single output symbol at each time; that is, overlapping input blocks are mapped into individual output symbols. A sliding-block code has a total length of $N = N_1 + N_2 + 1$, where the coding

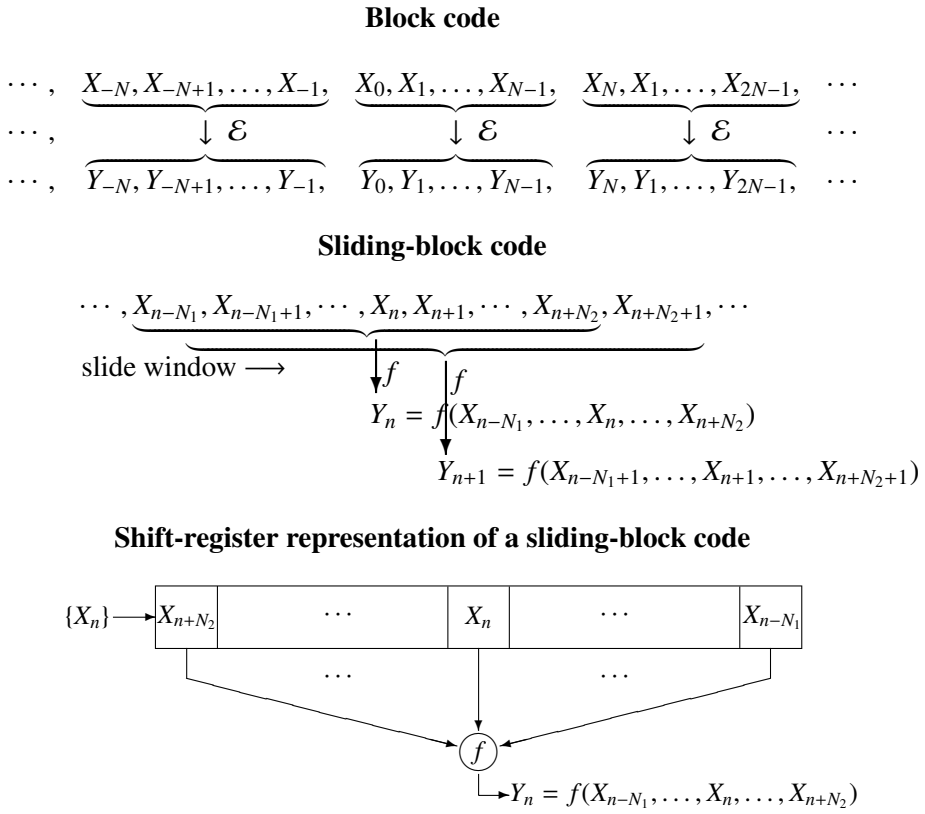


Figure 2: Block and sliding-block coding.

window views N_1 past symbols and N_2 future symbols along with the current symbol. These operations are depicted in Figure 2 along with a representation of a sliding-block code as an operation of on the contents of a shift register filled with input symbols.

There are constructions in information theory and ergodic theory for constructing block codes from sliding-block codes and vice-versa (see, e.g., [7], [8]).

Block codes dominate theory and practice and far more is known about their design, but unfortunately they lack several properties useful for theory. Block coding a stationary and ergodic process produces a coded process that is in general neither stationary nor ergodic. The encoded process will be block stationary (N-stationary if the blocklength is N), but it will not be N-ergodic. This loss of stationary and ergodicity complicates proofs and in both source coding and simulation it means the final reproduction process will lack the stationarity and ergodicity of the original source. Block coding can introduce periodicities at the block length. The resulting processes can be "stationarized" by introducing a uniform random startup time, but this simply results in a mixture of block stationary processes and hence the periodicities will remain in the sample paths.

Sliding-block codes are the standard approach in ergodic theory partially because they retain the stationary properties of the original process. Another difference is that the block coding theorems of information theory imply a sequence of codes that yield asymptotically optimal performance, but there is no "limiting" infinite code to which the finite block length codes converge—a block code does not make sense for an infinite blocklength. On the other hand, infinite-length sliding-block codes are well defined and one can prove theorems by first demonstrating the existence of an idealized infinite-length code, and then finding arbitrarily good approximations by finite-window codes.

Sliding-block codes provide better models for some real-world coding structures that do not have a block structure. For example, time-invariant convolutional codes, predictive quantization, nonlinear and linear time-invariant filtering, and wavelet coefficient evaluation are all naturally described as sliding-block operations.

The original importance of sliding-block codes derives from their use in proving the fundamental result of modern ergodic theory, the Kolmogorov-Sinai-Ornstein isomorphism theorem. Two processes are *isomorphic* if there exists an *invertible* sliding-block coding from either process to the other. Mathematically, isomorphic processes are the "same" in that one can code from one to the other and back again. Although the focus of ergodic theory is on processes with discrete alphabets, the definition and results also hold for the case of continuous alphabets where stationary "coding" would be better described in engineering terms as time-invariant filtering. In a sense, isomorphism can be viewed as a limiting form of lossless coding for discrete processes and inverse filtering for a continuous process.

In a series of papers in the early 1970s, Ornstein and his colleagues showed that for a class of random processes called *B-processes*, a

necessary and sufficient condition for two processes to be isomorphic is that they have equal entropy rate (see, e.g., [3], [4] for details and references). One of many equivalent characterizations of the class of *B*-processes is that they are processes which can be represented as a stationary coding (or filtering) of an iid process, exactly the class of processes considered here for simulation systems. Kolmogorov and Sinai had early demonstrated that equal entropy rate was necessary for isomorphism, Ornstein showed sufficiency for *B*-processes. It was also shown that entropy rate was not sufficient for isomorphism in the next most general known class of stationary and ergodic processes, the so-called "purely nondeterministic" processes which satisfy the 0-1 law of probability theory, called *K-automorphisms* (K for Kolmogorov) in ergodic theory.

Source Coding

The classic (Shannon) formulation for source coding assumes block codes. The setup is described by the following components:

Distortion measure $d_N(x^N, y^N) = \frac{1}{N}\sum_{i=0}^{N-1} d_1(x_i, y_i)$

Codebook/Decoder $\mathcal{C}_N = \{\mathcal{D}_N(i); i \in \mathcal{I}\}, |\mathcal{I}| = M$

Encoder $\mathcal{E}_N : A_X^N \to \mathcal{I}$

Distortion $D(\mathcal{E}_N, \mathcal{D}_N) = E(d_N(X^N, \mathcal{D}_N(\mathcal{E}_N(X^N))))$

Rate $R(\mathcal{E}_N) = \begin{cases} \frac{1}{N}\log M & \text{fixed-rate} \\ N^{-1}H(\mathcal{E}_N(X^N)) & \text{variable-rate} \end{cases}$

The fixed-rate case is emphasized here for simplicity.

The optimal distortion-rate performance is described by the operational distortion-rate function (DRF) defined by[3]

$$\delta_{\mathrm{BC}}^{(N)}(R) = \inf_{\mathcal{E}_N, \mathcal{D}_N : R(\mathcal{E}_N) \le R} D(\mathcal{E}_N, \mathcal{D}_N)$$

$$\delta_{\mathrm{BC}}(R) = \inf_N \delta_{\mathrm{BC}}^{(N)}(R) = \lim_{N\to\infty} \delta_{\mathrm{BC}}^{(N)}(R).$$

The operational DRF is not in general directly computable from its definition. Shannon introduced an information theoretic DRF (the Shannon DRF) which can often be computed analytically or numerically:

$$D_X(R) = \inf_N D_N(R) = \lim_{N\to\infty} D_N(R)$$

$$D_N(R) = \inf_{p^N : p^N \Rightarrow \mu^N, N^{-1}I(X^N, Y^N) \le R} E d_N(X^N, Y^N)$$

where the infimum over all joint distributions for X^N, Y^N with marginal distribution μ^N and $N^{-1}I(X^N, Y^N) \le R$. Shannon and his successors proved that subject to suitable technical conditions, the two DRFs are equal.

[3]Actually, Shannon introduced the dual function, the rate-distortion function.

> *Block Source Coding Theorem:* For a stationary and ergodic source, $\delta_{\mathrm{BC}}(R) = D_X(R)$.

Similar results hold for sliding-codes:

Encoder $f_{N_1+N_2+1} : A_X^{N_1+N_2+1} \to A_U$

$U_n = f_{N_1+N_2+1}(X_{n-N_1}, \ldots, X_{n+N_2})$

Decoder $g_{K_1+K_2+1} : A_U^{K_1+K_2+1} \to \hat{A}_X$

$\hat{X}_n = g_{K_1+K_2+1}(U_{n-K_1}, \ldots, U_{n+K_2})$

Distortion $D(f, g) = E(d_1(X_0, \hat{X}_0))$

Rate $R(f) = \log |A_U|$

The sliding-block code operational DRF is defined by

$$\delta_{\mathrm{SBC}}(R) = \inf_{f,g : R(f) \le R} D(f, g)$$

where the infimum is over all sliding-block codes, of finite or infinite window length. Note here that unlike the block coding case, the optimal quantity can be defined as an optimum over all codes rather than only as a limit of finite-dimensional optimizations.

The sliding-block source coding theorem (see, e.g., [7]) shows that (subject to the usual technical conditions) for stationary and ergodic sources the block and sliding-block operational DRFs equal each other and hence also the Shannon DRF.

> *Sliding-block Source Coding Theorem:* For a stationary and ergodic source, $\delta_{\mathrm{BC}}(R) = \delta_{\mathrm{SBC}}(R) = D_X(R)$.

Figure 3 depicts the complete source coding system for each coding style and provides context for one of the previous questions. Suppose that the code operates near to the Shannon limit, is it then true that the encoded process U is "nearly iid"? One way of quantifying the notion of "nearly iid" is to introduce a distortion or distance measure on random processes and to say that a process is "nearly iid' if it is close to an iid process in the sense of having a small distance to an iid process.

Process Distance Measures

In his development of the proof of the isomorphism, Ornstein introduced a distance on processes he called the $\bar{d}$ (d-bar) distance, a distance closely linked to the Kantorovich or transportation distance intimately connected with the development of linear programming. Consider two stationary random processes: X with distribution μ and Y with distribution ν, and a vector distortion measure $d_N(X^N, Y^N)$ defined for all

dimensions N. Define

$$\bar{d}_N(\mu^N, \nu^N) = \inf_{p^N \Rightarrow \mu^N, \nu^N} E_{p^N} d_N(X^N, Y^N)$$

$$\bar{d}(\mu, \nu) = \sup_N \bar{d}_N(\mu^N, \nu^N) = \inf_{p \Rightarrow \mu, \nu} E_p d_1(X_0, Y_0)$$

Thus $\bar{d}(\mu, \nu)$ quantifies the smallest achievable distortion between two processes with given distributions ("marginal" distributions for the separate processes) over all joint distributions consistent with marginals. There are many equivalent definitions; e.g., if both processes are ergodic, how much does a typical sequence of one have to be changed to be confused for a typical sequence of the other?

Such distances have a long and interesting history of repeated rediscovery and application to a variety of problems. Much of the history can be found in [10], [11]. Briefly, Kantorovich in the early 1940s developed the vector distance on compact metric spaces as an integral part of his contribution to the development of linear programming (for which he shared the 1975 Nobel Prize in Economics). Much of the early development focused on scalar spaces and noted contributors included Dall'Aglio, Frechet, Dobrushin, and Vasershtein (or Wasserstein) (see [10] for reference details). These authors focused mostly on scalar spaces and ℓ_r norms. Ornstein focused on the Hamming distance and discrete processes and extended the definition from vectors to processes, proving many of the key theoretical properties of such distances (see [3]). In [12] Ornstein's $\bar{d}$-distance was extended to continuous alphabet processes with additive distortion measures, including the ubiquitous squared error distortion. The name "$\bar{\rho}$"-distance was used to distinguish the continuous distortion from the discrete Hamming case used by Ornstein, but the basic ideas held even if the underlying vector distortion measures d_N were not metrics and hence $\bar{\rho}$ would be better called a "distortion" rather than a "distance" as it might lack the triangle inequality. The vector case was subsequently considered as the "L_r-minimal metric" in the literature, where the Kantorovich distance was applied to vectors using the ℓ_r norm. Here the definitions become

$$\bar{\rho}^{1/r}(\mu^N, \nu^N) = \ell_r(\mu^N, \nu^N) \triangleq [N\bar{d}_N(\mu^N, \nu^N)]^{\frac{1}{r}}$$

$$= \inf_{p^N \Rightarrow \mu^N, \nu^N} [E(\|X^N - Y^N\|_r^r)]^{\frac{1}{r}}$$

The notation $\bar{d}$ is usually reserved for the Ornstein (Hamming) case, $\bar{\rho}$ is used here for ℓ_r^r. The $\bar{\rho}/L_r$-minimal metric was rediscovered in the computer science literature as the "earth mover's distance" (where it was defined for possibly nonnormalized distributions) and used in clustering algorithms for pattern recognition, taking advantage of its evaluation as a linear programming problem. It was subsequently renamed the "Mallows distance" after a 1972 rediscovery of the Kantorovich distance.

These process distances have several useful properties [3], [12]:

- Ornstein's $\bar{d}$ distance and the L_r-minimal distance / $\bar{\rho}^{1/r}$ are metrics.
- The infimum is actually a minimum.
- The class of all B-processes of a given alphabet is the closure under Ornstein's $\bar{d}$ of all kth order mixing Markov processes of that alphabet.
- Entropy rate is continuous in $\bar{d}$, Shannon's DRF and the operational DRFs are continuous in $\bar{\rho}$.
- The $\bar{\rho}$ distortion for the squared error case is known for iid processes and for purely nondeterministic Gaussian processes and for the output of a stable time-invariant linear filter driven by a uniform iid process.

The process distance measure permits careful statements of the source coding and simulation problems and highlights their connections. We will use the word "distance" even when the measure does not satisfy the triangle inequality (but note that, for example, the square root of $\bar{\rho}$ with respect to the squared error distortion is a true distance or metric).

Application 1: Geometric View of Source Coding

The $\bar{\rho}$ distance gives a geometric interpretation DRF as the closest process in the $\bar{\rho}$ sense to the original source having constrained entropy rate [14]:

$$\delta_{\mathrm{BC}}(R) = \delta_{\mathrm{SBC}}(R) = D_X(R) = \inf_{\nu: H(\nu) \le R} \bar{\rho}(\mu, \nu)$$

This can be viewed as a form of simulation, but it does not meet the constraint imposed here since ν need not be a B-process. The geometric definition bears a strong resemblance to the process definition of the distortion-rate function [13], [14]:

$$D_X(R) = \inf_{p: p \Rightarrow \mu, I(X,Y) \le R} E[d_1(X_0, Y_0)]$$

where the infimum is now subject to an average mutual information rate constraint instead of an output entropy rate constraint.

Application 2: Quantization as Distribution Approximation

Pollard [15] and Graf and Luschgy [16] argued that a (vector) quantizer or block source code was equivalent to a probability distribution, that is, the codebook together with the probabilities inherited by applying it to a source is a random vector described by a distribution. This yields a geometric characterization of the fixed dimension operational DRF as

$$\delta_{\mathrm{BC}}^{(N)}(R) = \inf_{\nu^N} \bar{\rho}_N(\mu^N, \nu^N),$$

where the minimum is over all discrete distributions ν^N with 2^{NR} atoms. This can be viewed as a simulation of a random vector. A random process can be simulated by independent repetitions of the random vector simulation. This does not quite fit the current simulation problem because the resulting process is only block sta-

$$\begin{array}{ccccc}
\underbrace{X_0, X_1, \ldots, X_{N-1}} & \underbrace{X_N, X_1, \ldots, X_{2N-1}} & \underbrace{X_{2N}, X_{2N+1}, \ldots, X_{3N-1}} & \cdots \\
\downarrow \mathcal{E}_N & \downarrow \mathcal{E}_N & \downarrow \mathcal{E}_N & \cdots \\
\overbrace{U_0, U_1, \ldots, U_{N-1}} & \overbrace{U_N, U_{N+1}, \ldots, U_{2N-1}} & \overbrace{U_{2N}, U_{2N+1}, \ldots, U_{3N-1}} & \cdots \\
\downarrow \mathcal{D}_N & \downarrow \mathcal{D}_N & \downarrow \mathcal{D}_N & \cdots \\
\underbrace{\hat{X}_0, \hat{X}_1, \ldots, \hat{X}_{N-1}} & \underbrace{\hat{X}_N, \hat{X}_1, \ldots, \hat{X}_{2N-1}} & \underbrace{\hat{X}_{2N}, \hat{X}_{2N+1}, \ldots, \hat{X}_{3N-1}} & \cdots
\end{array}$$

vs.

$$\begin{array}{c}
\cdots, X_{n-N_1-1}, \underbrace{X_{n-N_1}, \cdots, X_n, \cdots, X_{n+N_2}}, X_{n+N_2+1}, \cdots \\
\downarrow f \\
\cdots, U_{n-K_1-1}, \underbrace{U_{n-K_1}, \cdots, U_n, \cdots, U_{n+K_2}}, U_{n+K_2+1}, \cdots \\
\downarrow g \\
\cdots, \hat{X}_{n-1}, \hat{X}_n, \hat{X}_{n+1}, \cdots
\end{array}$$

Figure 3: Block and sliding-block source coding.

tionary, but it provides a similar mechanism connecting simulation to source coding.

Application 3: Optimal Simulation and Source Coding

The process distortion yields a definition of optimal simulation of process $X \sim \mu$ using a sliding-block coding of an iid process Z [2]:

$$\Delta(X|Z) = \inf_{\tilde{\mu}} \bar{\rho}(\mu, \tilde{\mu})$$

where the infimum is over all processes $\tilde{\mu}$ formed as the output of a sliding-block code driven by Z.[4] Since sliding-block coding reduces entropy rate, $H(Z) \geq H(\tilde{\mu})$ and hence

$$\Delta(X|Z) \geq \inf_{\text{stationary ergodic } \hat{\mu}: H(\hat{\mu}) \leq H(Z)} \bar{\rho}(\mu, \hat{\mu}) = D_X(H(Z)). \quad (1)$$

Thus the best simulation performance provides a bound to the best source coding performance. It is shown in [2] that if X is a B-process, then the converse is true. The proof is easy to sketch in the case where the encoded alphabet is binary and the Z_n are equiprobable coin flips. From the sliding-block source coding theorem we can find an encoder f and decoder g such that the overall distortion from the input X to the reproduction $\hat{X}$ is approximately $D_X(1)$. Since X is a B-process, it is a sliding-block coding of an iid process and hence the reproduction $\hat{X}$ is also a B-process since it is produced by sliding-block decoding the sliding-block encoding of X, so it is itself a sliding-block coding of an iid process. The entropy rate of the reproduction $H(\hat{X})$ must be less than the entropy rate of the encoded process, which in turn is less than 1. If the distribution of $\hat{X}$ is $\hat{\mu}$, this implies that $D_X(1) \leq \bar{\rho}(\mu, \hat{\mu})$ for a B-process $\hat{\mu}$ with $H(\hat{\mu}) \leq 1$, which implies $D_X(1) \geq \Delta(X|Z)$.

[4] Other notions of optimal simulation have since been introduced which also use process distance measures in their formulation. See in particular Steinberg and Verdú [17].

Thus if the source is a B-process, then the source coding and simulation problems are equivalent in the sense that the following theorem holds.

> *Source coding and simulation theorem:* Given an iid process Z, if X is a B-process then $\Delta(X|Z) = D_X(H(Z)) = \delta_{\text{SBC}}(H(Z))$.

Bit Behavior for Nearly Optimal Codes

The results of the previous section are over three decades old. The preparations for the Shannon Lecture and conversations with Tamás Linder led to a reevaluation of the folk theorem regarding encoded bit behavior for nearly optimal codes and to the result described in this section. Again considering Figure 3 and armed with the notion of the $\bar{d}$ distance, we return to the question of whether indeed the U process constructed by encoding the original source is nearly iid when the source code is nearly optimal, where now the source is stationary and ergodic, but not necessarily a B-process.

First consider the block coding case with a block code $\mathcal{C}_N$ and the induced probability mass function (pmf) on the codewords or indexes π. What can be said about π if the code performance is near Shannon optimal? In particular, is it approximately uniform, like 2^N coin flips? The answer is "yes" in a highly qualified way. A more complete answer will result shortly in the sliding-block case. The block source coding theorem implies that there is an asymptotically optimal sequence of block codes $\mathcal{C}^{(N)}$ for which $D_N = Ed_N(X^N, \hat{X}^N) \downarrow D_X(1)$. Since $R_X(D)$ is a continuous function, standard inequalities imply that

$$1 = N^{-1} \log_2 2^N \geq N^{-1} H(\mathcal{E}(X^N)) \geq N^{-1} H(\hat{X}^N)$$

$$\geq N^{-1} I(X^N; \hat{X}^N) \geq R_N(D_N) \geq R_X(D_N) \underset{N \to \infty}{\to} 1$$

As the block length grows, the indexes have *maximal per symbol entropy* and hence can be thought of as *approximately uniformly distributed*. Unfortunately, however, the U processes so constructed are not in general stationary or ergodic and we have only a theorem for the behavior of vectors, albeit large dimensional vectors. The processes U cannot be said to converge in $\bar{d}$ to equiprobable coin flips. Stationarizing the code by a uniform random start does not help.

Using sliding-block codes instead of block codes, however, easily yields a rigorous process version of the theorem. Choose a sequence of sliding-block source codes $f^{(N)}, g^{(N)}$ so that $D_N = D(f^{(N)}, g^{(N)}) \downarrow D_X(1)$. Let $U^{(N)}$, $\hat{X}^{(N)}$ denote the resulting encoded and reproduction processes (necessarily stationary and ergodic). Then a similar string of inequalities to that used in the block coding case for finite-order entropies now yields results for entropy rates:

$$1 \geq H(U^{(N)}) \geq H(\hat{X}^{(N)}) \geq I(X, \hat{X}^{(N)})$$
$$\geq R(D_N) \underset{N\to\infty}{\to} 1$$

Linder observed that Marton's inequality for relative entropy and $\bar{d}$ [19] implies that if the entropy rate of a sequence of binary processes converges to 1, then the sequence converges to fair coin flips in a $\bar{d}$ sense,

$$\lim_{N\to\infty} \bar{d}(U^{(N)}, Z) = 0.$$

This provides a rigorous formulation and proof of the folk theorem that *as the average distortion nears the Shannon limit for a stationary ergodic source, the induced binary channel processes approach a process of equiprobable coin flips in $\bar{d}$.*

Recap

An old problem regarding the equivalence of source coding and simulation has been described, including some old results and one new one. Long known is the fact that if a source is a stationary filtering of an iid process (a *B*-process, discrete or continuous alphabet), then the source coding problem and the simulation problem have the same solution (and the optimal simulator and decoder are equivalent). The new result is a precise formulation of the fact that if source coding a stationary ergodic source approaches the Shannon optimum, then the encoded process is close to iid in $\bar{d}$. This result reinforces the intuitive connection of source coding and simulation—if the source coding is nearly optimal, then the system from bits to reproduction is nearly an optimal simulation (only "nearly" because the bits are only $\bar{d}$ close to iid, not actually iid). This suggests that source code design may profit from the study of simulating stationary and ergodic sources, and that good source codes yield good simulators (by using an actual iid driver instead of the approximately iid driver, as is done in LPC speech coding, for example). This tour has also provided a thinly disguised excuse to present ideas of modeling, coding, and process distance measures common to ergodic theory and information theory.

Final Thoughts and Questions

- The $\bar{d}$ close to iid property is nice for intuition, but is it actually useful? For example, *B*-processes have many special properties. Are there weak versions of those properties for processes that can be represented as a stationary coding of a process $\bar{d}$-close to iid? Which of the properties are "continuous" with respect to $\bar{d}$ distance and which are not?

- Does the equivalence of source coding and simulation hold for the more general case of stationary and ergodic sources? The results of Steinberg and Verdú [17] hold more generally, but in ergodic theory it is known that there are stationary, ergodic, mixing, purely nondeterministic processes which are *not* $\bar{d}$ close to a *B*-process.

- Source coding can be interpreted as an "almost isomorphism" between the source and the encoded binary process; it avoids the hard part (invertibility) in ergodic theory and simply settles for an approximate reconstruction of the source. Is this idea useful pedagogically? A suitably stated "almost isomorphism" theorem might show that if two processes have the same rate distortion function, then one can be coded approximately into the other, which in turn can be decoded to reconstruct approxately the original source. This weak version of isomorphism should hold more generally and be much simpler to prove.

- How does fitting a model using $\bar{\rho}$ compare to the Itakura-Saito distortion used in speech processing to fit autoregressive speech models to real speech? Variations of Marton's $\bar{d}$ inequalities in terms of relative entropy rate have been developed by Talagrand [20] for $\bar{\rho}$-type distortions, but these results hold only when one of the sources is either iid or a very structured Markov source. Can these inequalities be extended to cover the Gaussian autoregressive models popular in speech?[5]

- *B*-processes have a significant shortcoming in the real-world signal coding and processing example of speech. They well model unvoiced sounds, but voiced sounds are better modeled by a periodic input to some filter type, that is, by a 0-entropy process rather than a *B*-process. Is there a useful theory for composite (switched or mixture) models which produce a process by concatenating long segments of *B*-processes and 0-entropy processes? For example, to simulate proper sounding speech one would need to change both the filters and the drivers, sometimes producing *B*-processes and sometimes producing 0 entropy processes.

- Finally, are there good design algorithms for simulators? For example, how do you design a 1 bit per sample fake Gaussian process that is stationary and ergodic? One idea is the following: Suppose that the input process is an iid sequence of equally probable ±1s. Consider a very long shift register with a sparse collection of taps chosen at irregular times feeding

[5] Steinberg and Verdú [17] considered relative entropy rates in their simulation problem formulation.

into a summation whose output is scaled. The central limit theorem implies that the output will be approximately Gaussian, and choosing the taps properly should yield a small output correlation for nonzero lags. Intuitively this should be close in $\bar{\rho}$ to an iid Gaussian process. Now put this process into a filter with transfer function determined by the power spectral density of the target Gaussian process. This will produce a process of approximately the correct spectrum. With suitable choices is this near optimal for a one bit per sample process?

Acknowledgment

The Shannon Lecture slides devoted many slides to acknowledgments and thanks (see http://ee.stanford.edu/gray/Shannonintro.pdf). Briefly, I thank the IEEE Information Theory Society, the National Science Foundation, my graduate advisers, and my coauthors and former students. I specifically thank Tamás Linder for his help in preparing the lecture and these notes and Benjamin Weiss for many helpful discussions of the the history and concepts of ergodic theory.

References

[1] C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE National Convention Record, Part 4*, pp. 142–163, 1959.

[2] R.M. Gray, "Time-invariant trellis encoding of ergodic discrete-time sources with a fidelity criterion," *IEEE Trans. on Info. Theory*, vol. IT-23, pp. 71–83, Jan. 1977.

[3] D. Ornstein, "An application of ergodic theory to probability theory," *Ann. Probab.*, vol. 1, pp. 43–58, 1973.

[4] D. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems*. Yale University Press, New Haven, 1975.

[5] M. Smorodinsky, "A partition on Bernoulli shift which is not 'weak Bernoulli'," *Math. Systems Theory*, vol. 5, no. 3, pp. 201–203, 1971.

[6] *Information Theory and Network Coding*, R.W. Yeung, Springer, 2008 (Section 5.3).

[7] R.M. Gray, *Entropy and Information Theory*. Springer-Verlag, New York, 1990. Available at http://ee.stanford.edu/~gray/it.html.

[8] P.C. Shields, *The Ergodic Theory of Discrete Sample Paths*, AMS Graduate Studies in Mathematics, Volume 13, American Mathematical Society, 1996.

[9] L.V. Kantorovich, "On one effective method of solving certain classes of extremal problems," *Dokl. Akad. Nauk*, vol. 28, pp. 212–215, 1940.

[10] Rüschendorf, "Wasserstein-metric," in *Encyclopaedia of Mathematics. Supplement, I, II, III*, Hazewinkel, Michiel (ed.): Kluwer Academic Publishers (1997–2001).

[11] A. Vershik, "Leonid Vital'evich Kantorovich: A man and a scientist," vol. 1, pp. 130–152. Izd. SO RAN, Novosibirsk, 2002. English translation at http://www.pdmi.ras.ru/%7Evershik/papers.html#math.

[12] R.M. Gray, D.L. Neuhoff, and P.C. Shields, "A generalization of Ornstein's d-bar distance with applications to information theory," *Annals of Probability*, vol. 3, no. 2, pp. 315–328, Apr. 1975.

[13] K. Marton, "On the rate distortion function of stationary sources," *Probl. Contr. Inform. Theory*, vol. 4, pp. 289–297, 1975.

[14] R.M. Gray, D.L. Neuhoff and J.K. Omura, "Process definitions of distortion rate functions and source coding theorems," *IEEE Trans. on Info. Theory*, vol. IT-21, no. 5, pp. 524–532, Sept. 1975.

[15] D. Pollard, "Quantization and the method of k-means," *IEEE Trans. on Info. Theory*, vol. IT-28, no. 2, pp. 199–205, Mar. 1982.

[16] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Springer, Lecture Notes in Mathematics, 1730, Berlin, 2000.

[17] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distortion theory," *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 63–86, Jan. 1996.

[18] P.C. Shields, *The Theory of Bernoulli Shifts*. University of Chicago Press, Chicago, 1973.

[19] K. Marton, "Bounding $\bar{d}$ distance by informational divergence: A method to prove measure concentration," *Annals of Probability*, vol. 24, (1997) pp. 857–866, 1997.

[20] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Publ. Institut des hautes études scientifiques*, vol. 81 pp. 73–205, 1995.

A Citations Tornado

Conference Report: The 2008 IEEE International Symposium on Information Theory, Toronto Canada

Frank R. Kschischang and En-hui Yang

The 2008 IEEE International Symposium on Information Theory took place at the Sheraton Centre Toronto hotel in Toronto, Canada, from July 6–11, 2008. There were 856 registrants from 39 countries, distributed as follows:

Country	Attendees	Country	Attendees	Country	Attendees
Argentina	1	Greece	3	Norway	7
Australia	4	Hong Kong	12	Portugal	3
Austria	2	Hungary	2	Qatar	1
Belgium	1	India	17	Russia	2
Brazil	2	Iran	1	Saudi Arabia	1
Canada	130	Ireland	1	Singapore	3
China	7	Israel	34	South Africa	1
Cyprus	1	Italy	9	Spain	9
Denmark	3	Japan	44	Sweden	11
Finland	4	Korea	57	Switzerland	40
France	27	Lebanon	1	Taiwan	15
Germany	25	Netherlands	10	Turkey	2
Great Britain	11	New Zealand	1	United States	341

Among the 856 registrants, 333—or 39%—were students.

The highlight of ISIT was the 2008 Claude E. Shannon Award Lecture given on Wednesday by Professor Robert M. Gray of Stanford University on "Source Coding and Simulation."

The four Sunday tutorials drew 306 attendees. The tutorial presenters and titles were:

- J. Barros and S. W. McLaughlin, Information Theoretic Security: Theory and Practice;
- K. Ramchandran and S. Pradhan, Distributed Source Coding: Foundations, Constructions and Applications;
- N. Jojic, Information Exchange in Viral Infections: Applications to Vaccine Design;
- A. Ozdaglar, Networks' Challenge: Where Game Theory Meets Network Optimization.

The Symposium opened on Sunday evening with a Welcome Reception that featured live music performed by the Skule Jazz Combo, a group consisting entirely of University of Toronto engineering students.

Four excellent plenary lectures were presented:

- A. R. Calderbank: "Golay, Heisenberg and Weyl";
- A. Orlitsky: "From Two to Infinity: Information Theory and Statistics for Large Alphabets";
- E. H. Knill: "Building Quantum Computers"; and
- A. Wigderson: "Randomness—A Computational Complexity View".

A special session in memory of Sergio Servetto was organized by Toby Berger. A recent results poster session with 21 poster presentations took place immediately following the Shannon Lecture. As described in the previous Newsletter, the annual Awards Luncheon took place on Tuesday, with IEEE Information Theory Society president Dave Forney presiding.

Wednesday afternoon/evening excursions included a Niagara Falls tour (with 144 attendees) and a Toronto Blue Jays Baseball game (with 24 attendees). Many attendees took the afternoon off to explore Toronto.

The panel session on "Balancing Career and Personal Life" held at lunchtime on Wednesday (with panelists Andrea Goldsmith, Robert M. Gray, Ubli Mitra and Lalitha Sankar, and organized by Muriel Médard) was very well attended (by women *and* men). The student committee, under the leadership of Aylin Yener, organized two lunchtime events: a

A. R. Calderbank.

R. M. Gray.

"Roundtable Research Discussion" and a "Panel Discussion and Meeting."

The conference banquet took place on Thursday evening, the highlight of which was the announcement of Jorma Rissanen as the recipient of the 2009 Claude E. Shannon Award.

The dedication and hard work of many people resulted in a Symposium that ran very smoothly. Besides the Organizing Committee, Technical Program Committee, speakers and attendees, we wish to thank the staff at the Sheraton Centre Toronto for providing helpful and friendly service. Like at ISIT 2004 and ISIT 2006, Conference Management Services (CMS) provided superb logistical support (conference web site, paper-handling system, registration, publications, etc.). CMS staffer Lance Cotton, in particular, went beyond the call of duty in his efforts on behalf of the Symposium. In addition, the Symposium also received generous support from Research In Motion, the Ontario Centers of Excellence, IBM Research, Microsoft Research, and the U.S. National Science Foundation, which in turn helped many students attend the Symposium.

Some readers may be wondering about the strange choice of title for this article. It's an anagram: of "ISIT, Toronto, Canada!"

R. Urbanke (right) introduces A. Orlitsky.

E. H. Knill.

A. Wigderson.

David Neuhoff, Robert Gray and Anthony Ephremides.

Robert Gallager and Edmund Yeh.

The 2008 Information Theory society paper award has been jointly won by two 2006 IT Transactions papers:

"*Near-optimal signal recovery from random projections: universal encoding strategies?*", by Emmanuel J. Candès and Terence Tao and "*Compressed Sensing*" by David Donoho.

The Information Theory society awards committee viewed these two ground-breaking papers as independently introducing the new area of compressed sensing, which holds great promise for processing massive amounts of data, and has already had a broad impact on a diverse set of fields, including signal processing, information theory, function approximation, MRI, radar design, and sigma-delta conversion. Given the parallel nature of the work and the inspiration each research group had on the other, the committee recognized both papers jointly for pioneering this important research area. The committee also acknowledged the role of the paper "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information" by Emmanuel Candès, Justin Romberg, and Terence Tao, IEEE Trans. Inform. Theory, Feb. 2006, in developing some of the preliminary ideas that sparked the field of compressed sensing.

Information Theory Society Paper Award: Reflections on Compressed Sensing

Emmanuel J. Candès and Terence Tao

What is Compressed Sensing?

Current data acquisition methods are often extremely wasteful. Indeed, many protocols acquire massive amounts of data which are then—in large part—discarded without much information or perceptual loss, if any, by a subsequent compression stage, which is usually necessary for storage and transmission purposes. The two papers selected to receive the Information Theory Society Paper Award this year challenged this state of affairs and proposed an alternative acquisition paradigm perhaps best known as *compressed sensing* or *compressive sampling*. This paradigm asserts that it is possible to resolve signals from far fewer measurements or data bits than was thought necessary. In practice, this means for example that high-resolution imaging is possible with fewer sensors, or that one can speed up signal acquisition time in biomedical applications by orders of magnitude simply by taking far fewer specially coded samples.

Mathematically speaking, in a discrete-time setting, the setup for compressed sensing may be formulated as follows. We wish to acquire an object $x \in \mathbb{R}^n$, an n-pixel image for instance, and gather information about this object by collecting m linear measurements of the form

$$y_i = \langle a_i, x \rangle, \quad i = 1, \ldots, m.$$

Using matrix notations, one can write this as $y = Ax$ where A is an $m \times n$ matrix with the vectors a_i's as rows. The situation of interest corresponds to a heavy undersampling scenario in which the number m of measurements or samples is far lower than the dimension n of the object we wish to reconstruct. In general, everybody would agree that reconstructing x accurately from y is impossible. But suppose now that x is compressible in the sense that x is sparse or nearly sparse (the paper argues that most signals of interest are in this category). Then one can show that if the sensing vectors are somehow spread out, one can reconstruct x by solving the convenient optimization program

$$\min \|\tilde{x}\|_{\ell_1} \quad \text{subject to} \quad A\tilde{x} = y,$$

with optimization variable $\tilde{x} \in \mathbb{R}^n$. In plain English, among all the objects consistent with the data, we simply pick that with minimum ℓ_1 norm. The surprise is that this strategy recovers x accurately if x is approximately sparse, and even exactly when x is sparse. This is surprising because the information matrix or sensing protocol A is fixed in advance, and does not depend on the signal. That is, the sensing mechanism does not make any attempt to understand what the signal x is made of. In short, compressed sensing says that one can design nonadaptive measurement techniques that condense the information in a compressible signal into a small amount of data. The number of measurements necessary is comparable with the signal's sparsity level or the information rate if you will, rather than its nominal length. After the acquisition process, one solves an optimization program to "decompress" y and recover (an approximate version of) the original object x.

There are two principles at work here. The first is that randomness or pseudo-randomness can be used as an effective sensing mechanism; correlating a compressible signal against incoherent waveforms—even random waveforms—is efficient as it captures most of the information about x in just a few numbers. The second is that ℓ_1 minimization is able to disentangle the components of x from the compressed data. *Taken separately*, these facts were known—at least to some degree.

Leading Up to Compressed Sensing: High-Dimensional Geometry

At the theoretical level, compressed sensing is deeply connected with the field of geometric functional analysis, a field concerned with geometric and linear properties of convex bodies, and which has developed an astonishing array of deep geometric, probabilistic and combinatorial methods. In the seventies, for instance, Kashin studied the following problem [22]: suppose we are interested in recovering elements from the ℓ_1 ball $\{x \in \mathbb{R}^n : \|x\|_{\ell_1} \le 1\}$. How many and which linear measurements of the form $y = Ax$ do we need to take so that it is in principle possible to recover x from y to within precision ϵ? This is a question in high dimensional geometry. After seeing y, we would know that x lies in an affine space parallel to the nullspace of A. How then should we select this nullspace so that its intersection with the ℓ_1 ball has minimal radius? Kashin, and later Garnaev and Gluskin [17], answered this question very precisely. First, selecting the nullspace at

random provides, with very large probability, an intersection with nearly minimal radius. And second, the optimal number of measurements needed to achieve a precision ϵ is within a multiplicative constant from what the compressed sensing theory actually supplies. The fact that randomly selecting the nullspace of A (or equivalently the row space of A) nearly gives the most precise information about the object of interest shows, at least abstractly, that randomness leads to efficient sensing mechanisms. In a related but different direction, the theory of compressed sensing also connects with the deep probabilistic techniques of Bourgain [3], Bourgain-Tzafiri [4]–[6], and Rudelson [29]; in particular, the technique of generic chaining as developed by Bourgain [3] and Talagrand [33] to control suprema of random processes is a key technical tool to verify that certain measurement ensembles (such as the random Fourier ensemble) are indeed well-behaved for compressed sensing purposes.

ℓ_1 Minimization

Another feature of compressed sensing is of course the role played by the ℓ_1 norm as a sparsity-promoting functional, a fact that was made clear as early as the 60's. Suppose we observe

$$y(t) = f(t) + n(t), \qquad t \in \mathbb{R},$$

where $f(t)$ is bandlimited, $f \in B(\Omega) := \{f : \hat{f}(\omega) = 0 \text{ for } |\omega| > \Omega\}$, and $n(t)$ is an impulsive noise term supported on a sparse set T. In his Ph.D. thesis, Logan [23] observed the following phenomenon: if we recover f by

$$\min_{\tilde{f} \in B(\Omega)} \|y - \tilde{f}\|_{L_1(\mathbb{R})},$$

then the recovery is exact provided that $|T||\Omega| \leq \pi/2$! This holds whatever $f \in B(\Omega)$, whatever the size of the noise. The point here is that f is sparse in the frequency domain and that under this assumption, ℓ_1 minimization perhaps yields a rather unexpected result. Another leading early application of ℓ_1 minimization was reflection seismology, in which a sparse reflection function (indicating meaningful changes between subsurface layers) was sought from bandlimited data. On the practical side, Taylor, Banks and McCoy [34] and others proposed the use of ℓ_1 to deconvolve seismic traces. This idea was later refined to better handle observation noise [31]. On the theoretical side, rigorous results began to appear in the late-1980's, with Donoho and Stark [14] and Donoho and Logan [13] who extended Logan's 1965 result and quantified the ability to recover sparse reflectivity functions from bandlimited data. The application areas for ℓ_1 minimization began to broaden in the mid-1990's, as the LASSO algorithm [35] was proposed as a method in statistics for sparse model selection, Basis Pursuit [10] was proposed in computational harmonic analysis for extracting a sparse signal representation from highly overcomplete dictionaries, and a related technique known as total variation minimization was proposed in image processing [18].

Other Works

There are other precedents in the literature. Compressed sensing is related to coded aperture imaging, which also acquires information about an image by using a random array of pinholes instead of a single camera pinhole. It is also connected to a massive amount of work concerned with super-resolution, i.e., the problem of reconstructing high-frequency information from low-frequency content only. In coding theory, it is connected to Reed-Solomon codes and the related Finite Rate of Innovation theory [37]. In the literature of theoretical computer science algorithms, compressed sensing is also related to the theory of heavy hitters [18].

What Compressed Sensing has to Offer

A legitimate question is then what compressed sensing (a term only introduced recently in [12]) has to offer. With respect to the work on high-dimensional geometry, we see essentially three elements of novelty. (1) The first is that compressed sensing comes with an algorithm—ℓ_1 minimization—so that the results from Banach space theory are not just existential but also practical since one can recover those objects by linear programming. What is remarkable here is that the same algorithm is nearly optimal simultaneously over many classes of signals of interest to the signal processing community, namely, such the widely used ℓ_p balls with $p < 1$. (2) The second is that it brings a renewed understanding about the geometry of Banach spaces. Kashin's theorem was once considered as extremely deep and the argument nearly impenetrable (his proof is between 100 and 200 pages long). With the tools from compressed sensing, Kashin's theorem can be reduced to a maximum of two pages which are accessible to an undergraduate student with a basic knowledge of linear algebra and probability theory. (3) Finally, these new tools and techniques allow to tackle other algorithmic or information-theoretic questions such as the robustness of the reconstruction vis a vis noise or quantization errors.

With respect to the prior art on ℓ_1 minimization and from a more philosophical perspective, compressed sensing offers a unified organizational framework with which to view the previous scattered and specialized results on signal recovery, and on how randomness and the geometry of high-dimensional spaces can be exploited to reconstruct signals efficiently and robustly. One particularly valuable insight emphasised by the compressed sensing theory is that ℓ_1-based methods are well adapted to sparse or compressible signals (in contrast to the more classical ℓ_2-based methods, which are better suited for uniformly distributed signals).

A New Wave of Results

As just reviewed, this early work on compressed sensing may have crystallized a series of ideas about ℓ_1 minimization which were already in the air. But this work also went one step further. By combining the power of ℓ_1 minimization and randomness (leveraging the lessons from high-dimensional geometry), this work enabled novel practical acquisition approaches which more effectively apply system resources to find the useful information content embedded in a complex environment and directly measure it in a far more concentrated form. Perhaps the greatest surprise to us is seeing how much this has captured people's imagination, and how quickly these ideas have taken off. The number of people who contribute to the field of compressed sensing nowadays is staggering. More interestingly, this community brings together people with very different research interests: from pure mathematicians to circuit designers, from statisticians to radiologists, from information and communications theorists to optical systems designers, from theoretical computer scientists to numerical optimization experts. All have exploited, refined, extended, and pushed the theory and practice of compressed sensing in exciting and totally new directions. It is impossible in this note to report on all the amazing progress that have been made in the last

few years. To get a sense of both the vitality and the current breadth of the field, the reader may want to visit a few blogs on the world wide web, or browse through the pages of various international conference programs in the mathematical and statistical sciences, in information theory and signal processing, in medical imaging, in computational optics and so on. For instance, the blog *Nuit Blanche* [1], edited by Igor Carron, has at least one entry per day, each introducing two or three—sometimes even more—papers in the field. This blog is widely read and Igor reports a few thousand daily hits. In the academic community, the level of excitement is also clearly perceptible in several special issues that have been dedicated to this topic, see [2] for example.

Even though we cannot understandably survey all the many accomplishments of the compressed sensing community, we would nevertheless like to provide an idea of the range of the body of knowledge that is available today and, of course, was not when we wrote our paper.

• On the theoretical side, the theory has been considerably simplified thanks to the role played by the restricted isometry property or RIP (termed the uniform uncertainty principle in our paper), and also perhaps by Rudelson's selection theorem. We have learned that compressed sensing is robust to noise, which is a must for any real-world application [8], [9]. We have learned to identify sharp transitions between exact recovery and failure regions [15], [16]. We have learned that one could also obtain strong results with other algorithms, e. g., with greedy algorithms such as Orthogonal Matching Pursuit [36], Robust Orthogonal Matching Pursuit [26], CoSamp [25]. We have learned that one could get guaranteed instance optimality[1] [11].

• On the algorithmic side, there has been considerable effort in developing efficient ℓ_1 solvers and related algorithms in order to solve large scale problems. We now know far more about methods for solving ℓ_1 minimization problems than we did just four years ago. (As an aside, we have often heard from the optimization community its fondness for the flurry of activity around compressed sensing for it gives a sense of being even more relevant, as if such a thing were possible.) Other remarkable works in theoretical computer science have focused on the design of clever sensing matrices by borrowing ideas from discrete mathematics and more specifically from the literature on expander graphs, which come with fast recovery procedures [20], [38].

• On the application side, progress have been made in speeding up signal acquisition time in biomedical applications such as Magnetic Resonance Imaging by taking fewer samples [24]. Compressed sensing is already changing the way engineers think about signal acquisition in the area of analog-to-digital conversion. There are ongoing efforts to design subNyquist analog-to-digital receivers capable of sampling a wide radio-frequency band at a rate much below the Nyquist without much information loss, see [19] and the grant program DARPA BAA 08-03 for more information about these efforts. Compressive sensing is also changing the way some people think about digital optics; we now have single-pixel cameras [32], and it seems that new architectures for implementing compressive imagers in CMOS come out a rapid pace these days [21], [28]. Related works include those of David Brady, of William Freeman, and of their many colleagues.

We are sorry for the many omissions in the nonexhaustive list above. The reality is that we are humbled and happy at the same time to see such a large and broad community of enthusiastic and energetic researchers dramatically advancing this area of research, and systematically exploiting the results of our paper.

The Future

Will this flourishing activity be short lived? It is of course difficult to tell. Things seem to come and go at a much faster pace these days than they used to. We are already seeing a shift as people, inspired by compressed sensing, have realized that objects other than vectors (e.g., matrices and information tables) could also be recovered from what appear to be highly incomplete sets of sampled entries [7], [27]. What is clear, however, is this: we and others have learned about the amazing power of convex relaxations; we and others have learned about the fruitful interactions between analysis and probability theory; we and others have and learned that it was easier than ever to integrate sensing and processing, "to bring mathematics into the lens" to paraphrase Dennis Healy. It is probably not too speculative to assume that these will bear fruits for decades to come.

References

[1] Nuit blanche. http://nuit-blanche.blogspot.com/.

[2] *IEEE Signal Processing Magazine*, vol. 25, special issue on sensing, sampling, and compression, March 2008.

[3] J. Bourgain, "Bounded orthogonal systems and the $\Lambda(p)$-set problem," *Acta Math.*, vol. 162, no. 3–4, pp. 227–245, 1989.

[4] J. Bourgain and L. Tzafriri, "Complements of subspaces of l_p^n, $p \geq 1$, which are uniquely determined," in *Geometrical aspects of functional analysis (1985/86)*, number 1267 in Lecture Notes in Math., pages 39–52. publisher, Berlin, 1987.

[5] J. Bourgain and L. Tzafriri, "Invertibility of "large" submatrices with applications to the geometry of banach spaces and harmonic analysis," *Israel J. Math.*, vol. 57, no. 2, pp. 137–224, 1987.

[6] J. Bourgain and L. Tzafriri, "On a problem of Kadison and Singer," *J. Reine Angew. Math.*, vol. 420, pp. 1–43, 1991.

[7] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," 2008. Submitted to *Foundations of Computational Mathematics.*

[8] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[1] Roughly speaking, instance optimality asserts that the error (in a suitable norm, such as ℓ_1) between the original signal and the recovered signal in the given data class—in this case, sparse signals—differs by at most a constant factor from the distance between the original signal and the closest signal in the data class.

[9] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, pp. 2313–2351, 2007.

[10] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[11] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation, 2006, Preprint.

[12] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[13] D.L. Donoho and B.F. Logan, "Signal recovery and the large sieve," *SIAM J. Appl. Math.*, vol. 52, no. 2, pp. 577–591, Apr. 1992.

[14] D.L. Donoho and P.B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, June 1989.

[15] D.L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *Journal of the American Mathematical Society*, vol. 2006. To appear.

[16] C. Dwork, F. McSherry, and K. Talwar, "The price of privacy and the limits of LP decoding," in *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 85–94, New York, NY, USA, 2007.

[17] A. Garnaev and E. Gluskin, "The widths of a Euclidean ball," *Dokl. A. N. USSR*, vol. 277, pp. 1048–1052, 1984. In Russian.

[18] A.C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M.J. Strauss, "Near-optimal sparse Fourier representations via sampling," in *34th ACM Symposium on Theory of Computing*, Montreal, May 2002.

[19] D. Healy and D.J. Brady, "Compression at the physical interface," *IEEE Signal Processing Magazine*, 25, March 2008.

[20] P. Indyk, "Explicit constructions for compressed sensing of sparse signals," in *Symposium on Discrete Algorithms*, 2008.

[21] L. Jacques, P. Vandergheynst, A. Bibet, V. Majidzadeh, A. Schmid, and Y. Leblebici, "CMOS Compressed Imaging by Random Convolution," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[22] B. Kashin, "The widths of certain finite dimensional sets and classes of smooth functions," *Izvestia*, vol. 41, pp. 334–351, 1977.

[23] B.F. Logan, *Properties of High-Pass Signals*. PhD thesis, Columbia University, 1965.

[24] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–195, Dec. 2007.

[25] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, 2008. To appear.

[26] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. To appear, 2008.

[27] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization. 2007. Submitted to *SIAM Review*.

[28] R. Robucci, L.K. Chiu, J. Gray, J. Romberg, P. Hasler, and D. Anderson, "Compressive sensing on a CMOS separable transform image sensor," in to appear in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, 2008.

[29] M. Rudelson, "Random vectors in the isotropic position," *J. Funct. Anal.*, vol. 164, no. 1, pp. 60–72, 1999.

[30] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[31] F. Santosa and W.W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 4, pp. 1307–1330, 1986.

[32] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K.F. Kelly, and R.G. Baraniuk, "A compressed sensing camera: New theory and an implementation using digital micromirrors," in *Proc. Comp. Imaging IV at SPIE Electronic Imaging*, San Jose, California, January 2006.

[33] M. Talagrand, "Majorizing measures: The generic chaining," *Ann. Probab.*, vol. 24, no. 3, pp. 1049–1103, 1996.

[34] H.L. Taylor, S.C. Banks, and J.F. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, Jan. 1979.

[35] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[36] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[37] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[38] W. Xu and B. Hassibi, "Efficient compressive sensing with determinstic guarantees using expander graphs," in *IEEE Information Theory Workshop, Lake Tahoe*, September 2007.

Information Theory Society Paper Award: Reflections on Compressed Sensing

David L. Donoho

1. Introduction

Compressed Sensing (CS) is one label for a very active area of research currently attracting talented and energetic researchers from applied mathematics, information theory, mathematical statistics and optimization theory.

CS is sometimes reduced to a slogan: "Many signals (e.g., images) are naturally compressible; this can be exploited by developing novel methods of data acquisition which, in comparison to traditional approaches, involve far less data collection time, but on the other hand, require relatively far more processing effort in order to obtain a reconstructed signal."

The slogan is provocative, seeming to contradict both mountains of serious modern engineering work *and* the Fundamental Theorem of Linear Algebra (see below). Fortunately, some beautiful theory provides a well-founded and rigorously valid foundation. It's a pleasure to have this chance to point out a few of the very pretty ideas that had crystallized by the time my paper 'Compressed Sensing' [6] was submitted in 2004, and then briefly talk about the avalanche of work taking place over the last few years.

2. Compressibility is Key

Consider the following idealization: we are interested in an image or signal, conventionally represented by a vector f containing N real values—think of these as N pixel values. How many measurements are needed to obtain full knowledge of f? Group the measurements in a vector y with n entries, and suppose that the measurements are linear combinations of pixel values f obtained by an $n \times N$ matrix M according to $y = Mf$. How big should n be so that the measurements y determine f?

The fundamental theorem of linear algebra—"as many equations as unknowns"—gives the answer $n = N$: there should be as many measurements as pixels. N is of course a sufficient number—simply choose $M = I_N$, the $N \times N$ identity matrix—this corresponds to making one measurement for each pixel. And indeed, in many fields, current practice follows closely the $n = N$ prescription, magnetic resonance imaging and magnetic resonance spectroscopy being standard examples.

In some interesting and apparently surprising situations, $n < N$ measurements suffice. Indeed, suppose that f *is compressible* by some known transform coding scheme. Examples of such schemes include JPEG and JPEG-2000. Such codecs, in use heavily by internet browsers, cell phones and other devices, represent the signal in a basis different than the pixel basis, and rely on the empirical fact that the typical signals have relatively sparse representations in the transform basis. Thus, JPEG relies on the fact that images have relatively sparse representations in the discrete cosine basis, while JPEG-2000 relies on the fact that images have relatively sparse representations in the wavelet basis.

At the core of CS research is the idea that such transform sparsity allows fewer measurements. The next two sections give increasingly sophisticated ways to make rigorous this basic slogan.

3. Sparsity Measured by Number of Nonzeros

To see why transform sparsity might be helpful, let's push things to an extreme, and suppose a particularly strong form of transform sparsity. Suppose there is a fixed transform for the space of N-pixel images with the property that every signal f of interest has at most k nonzeros in that transform representation. Let Φ denote the $N \times N$ matrix representing the corresponding basis, so that $f = \Phi x$ where the columns of Φ are the basis elements and the N-vector x contains the coefficients of f in that basis. Define the $n \times N$ abstract measurement matrix $A = M\Phi$; the problem of recovering f from measurements $y = Mf$ is equivalent to recovering x from measurements $y = Ax$.

3.1. Unique Recovery

Our question becomes: x is known to be a vector with at most k nonzeros, $k < N$. How many measurements do we need in order to uniquely determine x?

Definition 3.1. *The matrix A has property $(Unique_0(k))$ if, for every $y \in \mathbf{R}^n$ representable as $y = Ax$ for some k-sparse vector $x \in R^N$, this representation is unique.*

Lemma 3.2. *[9], [10], [21]. If the $n \times N$ matrix A has its columns in general position then A has property $Unique_0(k)$ for every $k < n/2$.*

Here is a formal way to put things. Let $\|x\|_0$ denote the number of nonzero entries in x and consider the optimization problem

$$(P_0) \qquad \min \|x\|_0 \text{ subject to } y = Ax.$$

(P_0) expresses in a formal way the problem of finding the sparsest solution to the system $y = Ax$. The results just mentioned imply that, if A has property $(Unique_0(k))$, then the n abstract measurements provided by A followed by solution of (P_0) suffice to reconstruct x exactly. Here is an (impractical) method to solve (P_0): simply enumerate the k-subsets of columns of A; for each such k-subset I, say, form the matrix A_I consisting of just the columns from A with indices in I; check if it is possible to solve $y = A_I x_I$; if so, stop, announcing success. Of course, that strategy is totally unrealistic for interesting values of k, n, and N: there are $\binom{N}{k}$ subsets to examine, each one requiring $O(Nk^2)$ operations to decide if the corresponding system has a solution.

Surprisingly, there are practical methods that can efficiently find the solution to (P_0) when A is sufficiently special.

Suppose first that we know $x \geq 0$, and consider the inequality-constrained optimization problem

$$(LP) \qquad \min 1'x : \text{ subject to } y = Ax, x \geq 0.$$

This is a linear programming problem in standard form and can be solved using standard linear programming software. Such

linear programs are now considered efficiently solveable. In many cases, solving such problems will yield the unique k-sparse solution, if one exists.

Definition 3.3. *The matrix A has property $(Unique_+(k))$ if, for every $y \in \mathbf{R}^n$ representable as $y = Ax_0$ for some nonnegative k-sparse vector $x_0 \in R^N$, that vector x_0 is the unique solution of (LP).*

Lemma 3.4. *[11], [14], [20]. The n by N matrix A, $n < N$, consisting of the first n rows of the N by N real Fourier matrix, has property $(Unique_+(k))$ for every $k < n/2$.*

(The same conclusion holds for a large family of matrices A, not only the first n rows of the Fourier matrix).

The case $x \geq 0$ is restrictive, but suggests we are on to something. The situation where x may have positive and negative entries is far more applicable, and also more challenging to understand. Consider the optimization problem

$$(P_1) \qquad \min \|x\|_1 \ : \ \text{subject to } y = Ax.$$

note that here, x may be of either sign. Such problems can be solved by standard linear programming software and had previously been shown in [2] in 1998 to be efficiently solvable in some very large problems, for example with $n = 8,192$ and $N = 256,000$. Scott Chen's thesis had shown empirically that in some synthetic examples with very heavily underdetermined problems and very sparse solutions, the numerical solution of (P_1) was exactly the same (up to computer arithmetic accuracy) as the generating sparse solution.

Definition 3.5. *The matrix A has property $(Unique_\pm(k))$ if, for every $y \in \mathbf{R}^n$ representable as $y = Ax_0$ for some k-sparse vector $x_0 \in R^N$, that vector x_0 is the unique solution of (P_1)*

Incremental steps in the understanding of $(Unique_\pm(k))$ were made in the period 1999–2004, many of them appearing in *IEEE Trans. Inf. Thry.* Reference [10] in 2001 showed that the $n \times 2n$ matrix $A = [IF]$ where F is the n by n Fourier matrix, has property $(Unique_\pm(k))$ for $k < \sqrt{n}/2$, and various improvements soon followed; [18], [19], [22], [29] are examples.

The thrust of all this work is that apparently, for special kinds of matrices A, property $(Unique_\pm(k))$ holds for every $k = 1, 2, \ldots$ up to some threshold $k^*(A)$. Since $k^* < n/2$ and the matrices A in question are in general position, the property $Unique_0(k)$ also holds for those same k. Hence there is equivalence between solutions of (P_1) and (P_0) for k below a certain *breakdown point* for the equivalence [7]. It thus becomes natural to want to understand more deeply what determines this breakdown point, and how large it might be made.

3.2. Geometry of Unique Recovery

The properties $(Unique_+)$ and $(Unique_\pm)$ are equivalent to a beautiful geometric phenomenon concerning high-dimensional convex sets.

Suppose we have points $a_1, \ldots, a_N$ in the affine space $\mathbf{R}^n$; taking their convex hull gives a polytope, $P = conv(a_1, \ldots, a_N)$. The polytope is called k-**neighborly** if every subset of $k+1$-points generates a k-face of the convex hull, for $k = 0, \ldots, n/2 - 1$ [23].

1-Neighborliness can seem a very paradoxical and surprising property. It implies that every pair of points (a_i, a_j) spans an edge of $P = conv(a_1, \ldots, a_N)$: each line segment connecting some a_i with an a_j does not intersect the interior of P! Our low-dimensional intuition suggests, to the contrary, that such a line segment will generally cross the interior. Neighborliness is a genuinely high-dimensional phenomenon.

Lemma 3.6. *[14] Let the matrix A have columns $a_1, \ldots, a_N$ in general position in $\mathbf{R}^N$. If the polytope $P = conv(a_1, \ldots, a_N)$ is k-neighborly, then A has property $(Unique_+(k))$.*

The famous examples of neighborly polytopes [23] involve cyclic polytopes; these correspond to partial Vandermonde matrices and to partial Fourier matrices.

What about sparse vectors with negative and positive entries? The phenomenon $(Unique_\pm(k))$ is likewise intimately related to a standard notion in polytope theory.

Suppose we have points $a_1, \ldots, a_N$ in the affine space R^n; consider the polytope generated by these points and the reflections through the origin $P = conv(+a_1, \ldots, +a_N, -a_1, \ldots, -a_N)$. The polytope is symmetric about 0 and is called centrally k-neighborly if every subset of $\ell + 1$-points not containing an antipodal pair generates a face of the convex hull, for $l = 0, \ldots, k$.

Central neighborliness is again a paradoxical property, one genuinely not suggested by low-dimensional intuition.

Lemma 3.7. [5]. *Let the matrix A have columns $a_1, \ldots, a_N$ in general position in $\mathbf{R}^N$. The matrix has property $(Unique_\pm(k))$ if and only if the polytope $P = conv(a_1, \ldots, a_N, -a_1, \ldots, -a_N)$ is k-centrally neighborly.*

In short, phenomena concerning sparse solutions of underdetermined systems of linear equations which can be observed in signal processing correspond one-to-one with pre-existing questions in pure mathematics. Translating back into CS language:

1. Let $n \geq 2k + 1$. There are $n \times N$ matrices A with the property that, for every nonnegative k-sparse vector x, the solution of (LP) is the unique sparsest solution of $y = Ax$. It follows that every signal $f = \Phi x$ where x has at most k nonzeros can be uniquely recovered by linear programming from the n linear measurements generated by $M = A\Phi^{-1}$.

2. Suppose there exist k-centrally neighborly polytopes with $2N$ vertices in n dimensions. Then there are $n \times N$ matrices A with the property that, for every k-sparse vector x, the solution of (P_1) is the unique sparsest solution of $y = Ax$. It follows that every signal $f = \Phi x$ where x has at most k nonzeros can be uniquely recovered by linear programming from the n linear measurements generated by $M = A\Phi^{-1}$.

Since we know how to make the matrices A referred to in (1), the case with nonnegative coefficients, the main open issue concerns (2), the case with coefficients of either sign.

3.3. Random Matrices

Vershik and Sporyshev in the early 1990's had proposed to make polytopes that were so-called 'weakly neighborly' using random matrices [30]. Jared Tanner and I showed that the approach actually yielded true highly neighborly polytopes [13] and developed precise theory for the degree of neighborliness in case of large N, n. Adapting this approach to the centrosymmetric case, we can construct k-centrally neighborly polytopes [8]; this seems to be the best known way to make centrally-neighborly polytopes with n substantially less than N.

Let A be an $n \times N$ matrix with entries chosen i.i.d. from the standard normal $N(0,1)$ distribution. Then the centrosymmetric polytope $P = conv(a_1, \ldots, a_N, -a_1, \ldots, -a_N)$ is likely to be highly centrally neighborly. The exact degree of neighborliness is random, but for large N approximates a certain threshold function $\rho : [0,1] \mapsto [0,1]$ derived and studied in [8], [16]. If $k < n \cdot \rho(n/N)(1-\epsilon)$, where ϵ is small and N and n are large, then P is overwhelmingly likely to be k-centrally neighborly, while if $k > n \cdot \rho(n/N)(1+\epsilon)$, P is overwhelmingly likely to not be k-centrally neighborly. Here $\rho(1/2) \approx .09$, while $\rho(1/10) \approx .05$ [8, 16]. Informally, $\rho(1/10) \approx .05$ means that n needs to be about 20 times the underlying 'information content' k, while taking a tenth the conventional number of samples N.

Translating back into the compressed sensing framework, we could say that there is a *sampling theorem* for k-sparse signals: for $k \le k^*(A) \approx \rho(n/N) \cdot n$ every k-sparse vector x can be recovered from n random measurements $y = Ax$ by ℓ_1 minimization, while for substantially larger k, ℓ_1 minimization fails to recover some k-sparse vectors from those measurements. Here $\rho(1/2) \approx .09$, while $\rho(.1) \approx .05$ [8], [16].

4. Sparsity Measured by ℓ_1 Norm

Sparsity can allow for accurate and efficient solution of underdetermined systems—when sparsity is defined in a particularly strong way, by the number of nonzeros. Of course, "real" signals have transform coefficients which are not strictly speaking k sparse for any $k < N$.

In this section, we consider vectors x which have ℓ_1 norms bounded by 1. Such a vector can be approximated in ℓ_2 norm by a k-sparse vector with error at most $1/\sqrt{k}$. How many measurements do we need to reconstruct it approximately?

4.1. Information-Based Complexity

A general framework to approach related questions is called by the names 'Optimal Recovery' or 'Information-Based Complexity' IBC has been actively developed since the late 1970's. OR has its roots even earlier, but appears to have been subsumed by IBC.

A central question in that literature: we have an unknown function f, known only to belong to some specific function class $\mathcal{F}$ of interest to us. How many measurements n are really necessary so that the error of reconstructing f from those measurements is less than a desired threshold ϵ? More formally, suppose we let $I_n(f) = (\langle a_i, f \rangle, i = 1, \ldots, n)$ where a_i are some vectors and let R_n be a reconstruction operator, taking n numbers and returning a function f. The *optimal reconstruction error* is defined by:

$$e_n(\mathcal{F}) = \inf_{I_n, R_n} \sup_{f \in \mathcal{F}} \| f - R_n(I_n(f)) \|_2;$$

here we are optimizing over all ways of collecting information I_n and over all ways of reconstructing based on that information.

This type of question has been intensively studied over the last 25 years, principally in the Journal of Complexity, and in other papers and books associated with the IBC community. Each interesting research project in that field involves choosing a class of functions $\mathcal{F}$, a class of information operators I_n and some class of reconstruction operators R_n, and studying the resulting e_n.

The IBC/OR framework immediately makes it clear that ℓ_1 minimization is close to optimal for recovering a signal which is sparse in ℓ_1-norm. Let $\mathcal{F}_N$ denote the class of N-pixel images whose wavelet coefficients have unit ℓ_1 norm. Define a reconstruction rule $R^1_{n,N}$ based on ℓ_1 minimization as follows: For a given abstract measurement matrix A, form the matrix $M = A\Phi^{-1}$. Take measurements $y = Mf$; solving (P_1) with said y and A, yields the solution x_1. Put $f_1 = \Phi x_1$. (Thus $f_1 = R^1_{n,N}(Mf)$ actually depends on A and f.) Some general facts in IBC imply that this reconstruction operator $R_{n,N}$ obeys

$$\inf_M \sup_{\mathcal{F}_N} \left\| f - R^1_{n,N}(Mf) \right\|_2 \le 2 \cdot e_n(\mathcal{F}_N). \qquad (1)$$

Thus ℓ_1 minimization is within a factor 2 of optimal if we use the most clever choice of measurement matrix.

It becomes of interest to know $e_n(\mathcal{F}_N)$ and the type of measurement matrix which achieves it; this leads to deep and surprising high-dimensional geometry.

4.2. Random Sections of the ℓ_1 ball

The field of "Geometry of finite-dimensional Banach Spaces" has been intensively developed since the mid-1970's. Some of the major developers include Vitali Milman, Alain Pajor, Stan Szarek and Nicole Tomczak-Jagerman. In that field, researchers study the properties of norms on finite-dimensional Banach spaces (such as the ℓ_1-norm) and properties of the unit ball in such norms when we take sections or projections of them.

Here is a question which is completely natural in that theory. Take the ℓ_1-ball B_1^N in N-dimensional space (N is large). Take an $N-n$ dimensional linear subspace V and intersect it with B_1^N—i.e., consider the *section* $B_1^N \cap V$. What is its radius?

To measure the radius, let $B_2^N(0,r)$ denote the ℓ_2-ball of radius r in $\mathbf{R}^N$ and let $r_{n,N} = r_{n,N}(V)$ be the smallest $r > 0$ such that

$$B_1^N \cap V \subset B_2^N(0, r_{n,N}).$$

Our low-dimensional Euclidean intuition suggests that since B_1^N is a kind of 'ball' of radius 1, $B_1^N \cap V$ will probably have a radius about 1. Instead, Banach space ideas yield a surprise. There is some subspace V of codimension n obeying

$$r_{n,N}(V) \le c \cdot \log(1 + N/n)^{1/2} n^{-1/2}. \qquad (2)$$

For large enough n and N, this radius $r_{n,N}$ can be quite small—something that could only be true in high dimensions and goes against any low-dimensional pictures we can conjure up. In fact, such exotic behavior is hard to exhibit explicitly; the strategy of proof developed by the Banach spacers is to use *random* subspaces of codimension n: most of them will work.

Another basic insight at the heart of IBC gives

$$e_n(\mathcal{F}_N) = \inf_V r_{n,N}(V), \qquad 1 \le n \le N, \qquad (3)$$

where the infimum is over subspaces of codimension n.

Combining (1)–(3) gives:

Theorem 4.1. *Let $\mathcal{F}_N$ denote the class of N-pixel images whose wavelet coefficients have an ℓ_1 norm bounded by 1. There is an $n \times N$ matrix M so that the reconstruction error from ℓ_1-minimization obeys:*

$$\left\| f - R^1_{n,N}(Mf) \right\|_2 \le c \log(1 + N/n)^{1/2} n^{-1/2} \qquad \forall f \in \mathcal{F}_N. \qquad (4)$$

In this relationship, the dominant behavior involves n, and N plays only a very weak role (appearing in the logarithm). If N is large, there is a dramatic effect on the number of measurements needed. Suppose we want to achieve a reconstruction error $e_n(\mathcal{F}) < \epsilon$. then we only need to sample n measurements, where

$$(\epsilon/c)^2 = (\log(1 + N/n)/n).$$

If $\epsilon/c = \sqrt{1/1000}$, and N is 10^6, we need only about $n = 2000$ measurements to guarantee a reconstruction error $\le \epsilon$. So we have reduced from 10^6 measurements to $2 \cdot 10^3$.

The fact that the radius $r_{n,N}$ can be so small turns out to be equivalent, by duality, to a phenomenon first discovered in the mid-1970s by Boris Kashin [31,32]. Kashin was trying to complete the program—inspired by Kolmogorov and pursued almost to completion by many other researchers—of computing the linear n-widths of Sobolev classes. In order to calculate the Kolmogorov n-widths of the recalcitrant Sobolev classes, he was forced into a novel kind of calculation involving random subspaces. In retrospect, we can infer by duality that his calculation, with later refinements by Garnaev and Gluskin [32], imply (2).

Erich Novak in the mid 1990's pointed out that Kashin's results ought to be of significance for people working in IBC.

The relations (1)–(3) were the starting point for the paper [6] I am supposed to place in context here. Theorem 4.1 is a special case of results in that paper; another set of results in that paper made connections between results on the k-sparse notion of sparsity and the ℓ_1 notion.

5. Recent Developments

So far, I have mentioned theoretical issues that were well-understood by the end of 2004. Since then, many talented, energetic researchers have expanded the picture enormously. Space requirements make it impossible to even begin to describe it all; for example, the paper "Compressed Sensing" has over 300 citations in Google Scholar. I'll just make a few observations, apologizing in advance to the very large number of authors whose work I am unable to correctly cite here.

- *Quantifying Breakdown:* the breakdown point for k-sparse solution is now very well-understood; Jared Tanner and I derived the appealing asymptotic form $k^* \sim n/(2e\log(N/n))$ for $n \ll N$; [16].

- *Typical Sparsity:* statements can be made about recovering almost all, rather than all k-sparse vectors; asymptotically, the precise threshold $\sim n/(2\log(N/n))$ for $n \ll N$; [16].

- *Finite-N Results:* It is also possible to give results bounding probability of breakdown at specific, finite N; see work by Mark Rudelson, and Roman Vershynin, and also [15].

- *Impacts of Noise and Stability:* The ℓ_1 minimization approach continues to work even with moderate noise, perhaps after modifying the minimization objective to include the possibility that $y \approx Ax$; [1], [17], [27].

- *Other kinds of sparsity:* in addition to $\|x\|_0 \le k$ and $\|x\|_1 \le C$, one can measure sparsity in other ways, for example by the ℓ_p 'norm', $0 < p < 1$. [1], [6].

- *Algorithms:* A tribe of researchers have rushed to speed up ℓ_1 minimization software, often with great success [4]. Joel Tropp and others have shown that heuristic algorithms seem to work well in theory and often in practice. [28], [29].

- *Alternate Randomizations:* The elegant theory concerns uniformly-distributed random subspaces V; or, equivalently, Gaussian iid random matrices A. Empirical and theoretical evidence suggests that many different ensembles of random matrices work fairly well; see recent work by Anna Gilbert, Jared Tanner, Alain Pajor and respective co-authors.

- *Derandomization:* There are also efforts to develop completely deterministic matrices A. In the case where the coefficients are nonnegative, this was solved by Lemma 3.4 above; but much remains to be done in the case where coefficients can be plus or minus.

Many teams are developing applications, in areas as diverse as MRI imaging, NMR spectroscopy, and radar signal processing. I

leave it to others to comment, mentioning only the fact that from my first hand experience with serious MR imaging researchers I believe the MR Imaging efforts are likely to have quite serious impacts [24].

6. Personal Comments

There's a very long personal story I could tell, spanning decades, beginning with my first job after college. I'll boil what I say down to a few observations.

- *Contributions of Industry:* I became aware of the rather remarkable sparsity-seeking properties of ℓ_1 minimization while working in the seismic oil exploration business many years ago. Industrial research groups understood by the early 1980's that one could use ℓ_1 minimization to seemingly solve underdetermined systems of equations representing geophysical inverse problems where the low-frequency behavior of an object was completely unobserved. Seemingly, industry was almost thirty years ahead of today's academic trends! To be fair, serious mathematical contributions inspired by these industry insights came in the next decade [3], [11], [12], [26], but today's academic surge only came decades later.

- *Contributions of Theory:* I wish I could say that this is a success story of theory 'showing the way to applications,' like Moses bringing down the law from the mountain. The theory has spurred today's surge of interest, but there has been applied research in related directions for decades.

  My co-authors, NMR spectroscopists Jeff Hoch and Alan Stern published the idea of undersampling in the Fourier domain followed by maximum entropy reconstruction in the early 1990's [25]. By 2004 there were dozens of applied publications on undersampling and nonlinear reconstruction of various kinds in NMR spectroscopy and MR imaging—too many to cite here.

  In array signal processing, Bhaskar Rao and Irina Gorodnitsky were getting good results in the mid 1990's on the sparsity-seeking algorithm FOCUSS [21]. In the case of nonnegative x and A the Partial Fourier Matrix, we have the famous spectral extension problem. The utility of ℓ_1 estimation under sparsity was established already more than 15 years ago in [11]; Jean-Jacques Fuchs was publishing about ℓ_1 methods in spectral estimation in the mid 1990's.

  The pioneers in the applied community who advocated undersampling long ago frequently met resistance and disbelief among their peers; frequently they were told that Nyquist, or Shannon, or the Fundamental Theorem of Linear Algebra proved that undersampling was for fools. Solid theory offers a scholarly way to respond to such criticism. It also provides a *morale boost*; practitioners can now be more confident in their research programme than previously.

  Theory *has* made decisive contributions; practitioners were not very clear about the role that *transform* sparsity played in their work, in how to effectively exploit it, or *how much* sparsity was needed. Theory has shown the importance of sparsifying transforms such as wavelet transforms, the value of ℓ_1 minimization and related ideas and has proved rigorously that failure will occur beyond sharp, specific limits knowing the point at which failure must occur seems important for credibility.

- *Contributions of Frameworks:* Many mathematical scientists are suspicious of framework-building; I can recall as a young researcher hearing skepticism about IBC, saying in effect that such frameworks only repackage what was already known. CS gives a counterexample to such skepticism. In fact attempts to explain CS without carefully using this framework often lead simply to confusion and seeming contradiction. Frameworks do matter, and we should honor the framework builders who give us ways to correctly frame and communicate new ideas. I'd like to toast Joe Traub, Greg Wasilkowski, and Henryk Wozniakowski for their leadership in developing IBC. I'd like to also toast the major developers of OR: Charles Micchelli, and even earlier, I.J. Schoenberg and Hans Weinberger.

- *Contributions of Russians:* Two major figures in the above story, Kashin and Vershik, are products of the Soviet/Russian mathematical tradition. The profound intellectual contributions of this mathematical community are still not fully appreciated.

- *Contributions of Students and Colleagues:* I have worked over the years with a series of collaborators who have authored with me a series of papers involving sparsity, ℓ_1, and underdetermined systems of equations; these include Phil Stark, Jeff Hoch, Alan Stern, Iain Johnstone, Ben Logan, Scott Chen, Xiaoming Huo, Jean-Luc Starck, Michael Elad, Jared Tanner, and Michael Lustig. A former student who eventually became interested in this area, to notable effect, was Emmanuel Candès.

  One of the pleasures of getting old(er) is the humbling realization of the immense contributions of those who came before and the overwhelming energy and talent of the generation to come.

  Of these, I toast my mentor Ben Logan, to me the greatest harmonic analyst of the twentieth century, after Polya, Szego, Beurling and Wiener. Already in his 1965 Ph.D. thesis (in Electrical Engineering!) he proved the first ℓ_1-uncertainty principle, the first reference I know of hinting at the remarkable interaction between sparsity and ℓ_1 minimization. I also toast my 'mentees' Jean-Luc Starck and Michael Lustig for taking serious career risks to introduce sparsity-based methods into their fields (astronomy and medical imaging) with impressive energy, imagination, and persistence.

References

[1] E.J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies," *IEEE. Trans. Info. Thry.*, Vol. 52, No. 12, pp 5406-5425, December 2006.

[2] S.S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001. Reprinted from SIAM J. Sci. Comput., vol. 20, (1998), no. 1, 33–61.

[3] P.B. Stark and D.L. Donoho, "Uncertainty principles and sig-

nal recovery, "*SIAM Journal on applied mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[4] D.L. Donoho and Y. Tsaig, "Fast solution of ℓ_1 minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, no. 11, 2008.

[5] D.L. Donoho, "Neighborly polytopes and sparse solutions of underdetermined linear equations," Technical report, Department of Statistics, Stanford University, 2005.

[6] D.L. Donoho, "Compressed sensing," *IEEE. Trans. Info. Thry.*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] David L. Donoho, "For most large systems of underdetermined equations, the minimum ℓ^1-norm solution is the sparsest solution," *Comm. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2006.

[8] D.L. Donoho, "High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension," *Disc. Comput. Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[9] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proc. Natl. Acad. Sci.*, USA, vol. 100, no. 5, 2003, pp. 2197–2202 (electronic).

[10] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[11] D.L. Donoho, I.M. Johnstone, J.C. Hoch, and A.S. Stern, "Maximum entropy and the nearly black object," *J. Roy. Statist. Soc. Ser. B*, vol. 54, no. 1, pp. 41–81, 1992. With discussion and a reply by the authors.

[12] D.L. Donoho and B.F. Logan, "Signal recovery and the large sieve," *SIAM J. Appl. Math.*, vol. 52, no. 2, pp. 577–591, 1992.

[13] D.L. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, 2005, pp. 9452–9457.

[14] D.L. Donoho and J. Tanner, "Sparse nonnegative solutions of underdetermined linear equations by linear programming," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 27, 2005, pp. 9446–9451.

[15] D.L. Donoho and J. Tanner, "Exponential bounds implying construction of compressed sensing matrices, error-correcting codes and neighborly polytopes by random sampling," *preprint*, 2007.

[16] D.L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *Journal of the AMS*, 2008.

[17] M. Donoho, D.L. Elad, and V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[18] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[19] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[20] J.-J. Fuchs, "Sparsity and uniqueness for some specific underdetermined systems," In *IEEE International Conference on Acoustics, Speech, and Signal Processing, March 2005*, Philadelphia, PA., 2005.

[21] I.F. Gorodnitsky and B.D. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted norm minimization algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600 – 616, 1997.

[22] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[23] B. Grünbaum, *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.

[24] D. Lustig, D. Donoho, and J.M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[25] G. Wagner, P. Schmieder, A.S. Stern, and J.C. Hoch, "Application of nonlinear sampling schemes to cosy-type spectra," *Journal of Biomolecular NMR*, vol. 3, no. 5, p. 569, 1993.

[26] F. Santosa and W.W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM Journal on Scientific and Statistical Computing*, 1986.

[27] J.A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[28] J.A. Tropp and A. Gilbert, Signal recovery from partial information by orthogonal matching pursuit. Technical report, Mathematics Department, University of Michigan, 2005.

[29] J.A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[30] A.M. Vershik and P.V. Sporyshev, "Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem," *Selecta Math. Soviet.*, vol. 11, no. 2, pp. 181–201, 1992.

[31] A.Y. Garnaev and E.D. Gluskin. On the widths of the euclidean ball. Doklady AN USSR, 277:10481052, 1984. In Russian; English Translation in Soviet Mathematics – Doklady 30 1984.

[32] B. S. Kashin. Widths of certain finite-dimensional sets and of classes of smooth functions. Izv. Akad. Nauk SSSR Ser. Mat., 41:234251, 1977.

Joint CommSoc/IT Paper Award: Reflections on Accumulate Repeat Accumulate Codes

Aliazam Abbasfar, Dariush Divsalar, and Kung Yao



1. Introduction and Motivation of the Work

Low Density Parity Check (LDPC) codes were proposed by Gallager [1] in 1962. After introduction of turbo codes by Berrou et al. [2] in 1993, researchers revisited the LDPC codes, and extended the work of Gallager. During 1962 to 1993, only few people, notably Tanner in 1981 [3], paid attention to the work of Gallager and made some contributions. After 1993 huge number of contributions have been made to LDPC codes; see for example [6], [12]–[17] and references there.

Recently RA and IRA codes, as simple subclasses of LDPC codes with fast encoder structure, were proposed [7], [8]. RA and IRA codes can also be considered as serial concatenated codes [18]. RA codes use fixed repetition for input bits. On the contrary, IRA codes inspired by RA and irregular LDPC [5] codes have irregular repetition for input bits. In fact, node degree distribution can be optimized to achieve low threshold performance. To achieve very low threshold for IRA, as for LDPC codes, maximum repetition for some portion of input bits can be very high.

ARA codes are concatenation of an accumulator with RA or IRA codes. Before elaborating on the role of this accumulator as a precoder for RA or IRA codes and graph representation of ARA codes, we use the definition of protograph introduced by Thorpe in [9]. A similar definition called projected graph was introduced by Richardson et al. in [10] for implementation of the decoder for LDPC codes. They show that if an LDPC code can be represented by the smallest base-graph or projected graph, then high speed implementation of the decoder will be more feasible. Protograph definition also facilitates the minimal graph representation for the overall graph of an LDPC code. We will show that ARA codes have such a protograph or projected graph representation.

A protograph [9] is a Tanner graph with a relatively small number of nodes. A protograph $G = (V, C, E)$ consists of a set of variable nodes V, a set of check nodes C, and a set of edges E. Each edge $e \in E$ connects a variable node $v_e \in V$ to a check node $c_e \in C$. Parallel edges are permitted, so the mapping $e \rightarrow (v_e, c_e) \in V \times C$ is not necessarily 1:1. As a simple example, we consider the protograph shown in Figure 1. This graph consists of $|V| = 4$ variable nodes and $|C| = 3$ check nodes, connected by $|E| = 9$ edges. The four variable nodes in the protograph are denoted by "0,1,2,3" and the three check nodes by "0,1,2". By itself, this graph may be recognized as the Tanner graph of an LDPC code with ($n = 3$, $k = 1$). In Figure 1, the variable nodes connected to the channel are shown with dark circles i.e., transmitted nodes. Blank circles are those variable nodes not transmitted through the channel, i.e., punctured nodes. Check nodes are circles with plus sign inside. Under certain conditions on the corresponding parity check matrix, i.e., full rank, the code rate for the protograph code can be defined as $R_c = (|V| - |C|)/|V_t|$, where V_t represent a set of transmitted nodes in the protograph. There are $|E| = 9$ types of edges in the protograph representation of RA code. In fact the protograph LDPC codes are subclass of multi edge type LDPC codes introduced in [11]. In multi edge type LDPC code, a group of edges may represent a type, whereas in the protograph LDPC codes each edge is a type.

For a given protograph, we can obtain a larger graph by a copy-and-permute operation. For more details on protographs see [9]. The resulting larger graph is called the derived graph and the corresponding LDPC code is a protograph code. In general, we can apply the copy-and-permute operation to any protograph to obtain derived graphs of different sizes. This operation consists of first making N copies of the protograph, and then permuting the endpoints of each edge among the N variable and N check nodes connected to the set of N edges copied from the same edge in the protograph. Equivalently the derived graph can be viewed as a multi edge type LDPC code in which each edge in the protograph is a distinct type. Thus the derived graph can be obtained by replacing each edge of the protograph with a permutation of size N. In other words, LDPC codes with protograph are multi edge codes with equal number (N) of edges for each type. In our examples we will consider both multi edge type LDPC codes and protograph LDPC codes. The difference is mainly on the use of permutation on multiple edges or on single edges. Multi edge type LDPC codes with rational degree distribution can also have a projected graph description. In Figure 2, ensembles of three types of LDPC codes are presented, namely unstructured LDPC, Multi edge LDPC and protograph LDPC codes.

In Figure 1(a), the encoder with single permutation may represent a multi edge type RA code. In Figure 1(b) protograph based RA code is shown (as long as the permutation π is chosen to be decomposable into permutations along each edge of the protograph). In the figure, the minimum threshold of protograph RA code is also shown. As follows "threshold" means "E_b/N_o iterative decoding threshold".

Irregular repetition in RA reduces the iterative decoding threshold. We asked ourselves the following question. Is it possible to reduce iterative decoding threshold by other means? While performing experiments on hybrid concatenated codes (parallel concatenation of a serial concatenated code with another convolutional code in parallel) [21], we noticed that if the parallel rate-1 con-

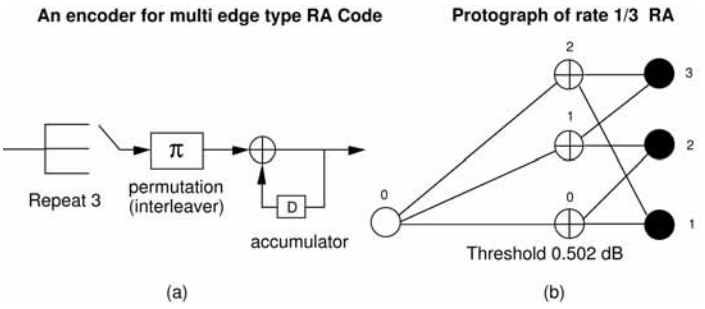


Fig. 1. Rate 1/3 RA code with repetition 3, multi edge and protograph versions.

volutional code is non-recursive, in particular, if it is a differentiator, then the iterative decoding threshold can be reduced. We applied this observation to RA code which is a serial concatenated code. In hybrid concatenated code, the information data usually is applied to the input of the outer code. If instead we apply the information data to the output of differentiator, then differentiator can be viewed as an accumulator. Thus we learned that an accumulator as a rate-1 precoder (rate-1 LDGM code) applied before the repetition code can improve the iterative decoding performance. Other types of rate-1 LDGM codes were tried but none have shown a better improvement in threshold when concatenated with RA code. These observations led us to discover the Accumulate Repeat Accumulate protograph LDPC code. The protograph representation of such construction is shown in Figure 3(a).

II. Accumulate-Repeat-Accumulate Codes

Let us consider a rate 1/3 systematic RA code with repetition 3 as in Figure 3(b). Alternatively consider the same repetition 3 node in RA that is precoded by an accumulator. Let us compare the extrinsic SNR behavior of these two cases using Gaussian density evolution as shown in Figure 3(c). As the Gaussian density evolution analysis shows, the use of a rate-1 accumulator dramatically improves the extrinsic SNR behavior of repetition 3 at high extrinsic SNR region. However, it slightly deteriorates the behavior of repetition 3 code at very low extrinsic SNR region.

Now let us use the inner accumulator code. Since the serial concatenation consists of outer accumulator, middle repetition, and inner accumulator, we call it Accumulate-Repeat-Accumulate (ARA) code. The rate 1/3 ARA code, the corresponding protograph, and the extrinsic input-output SNR curves using Gaussian density evolution are shown in Figure 3. Except for this example where we used Gaussian density evolution to show the advantage of precoding, in this paper actual density evolution on protographs will be used. Using density evolution, we obtain the exact iterative decoding threshold of -0.048 dB for the protograph shown in Figure 3(a). If we remove the precoder, and transmit the node 1 in the Figure 3(b), the threshold will be 0.73 dB. Thus precoder improves the iterative decoding threshold by 0.77 dB. We call such performance improvement due to the use of precoder as "**Precoding gain**."

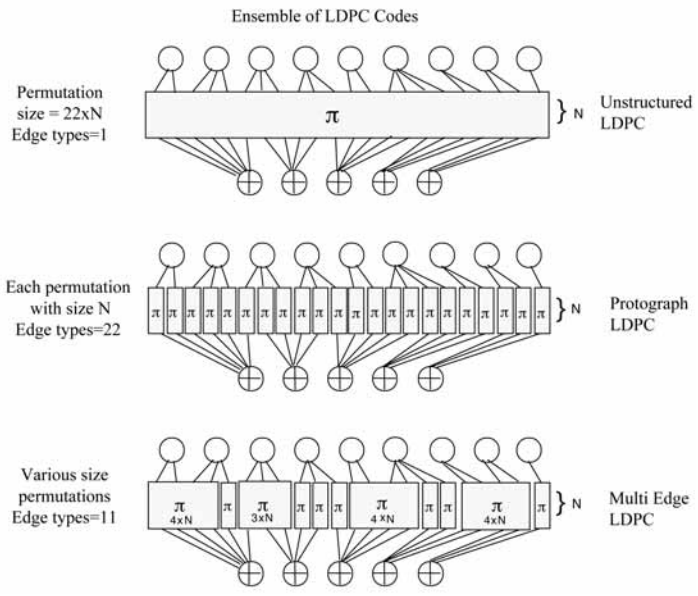
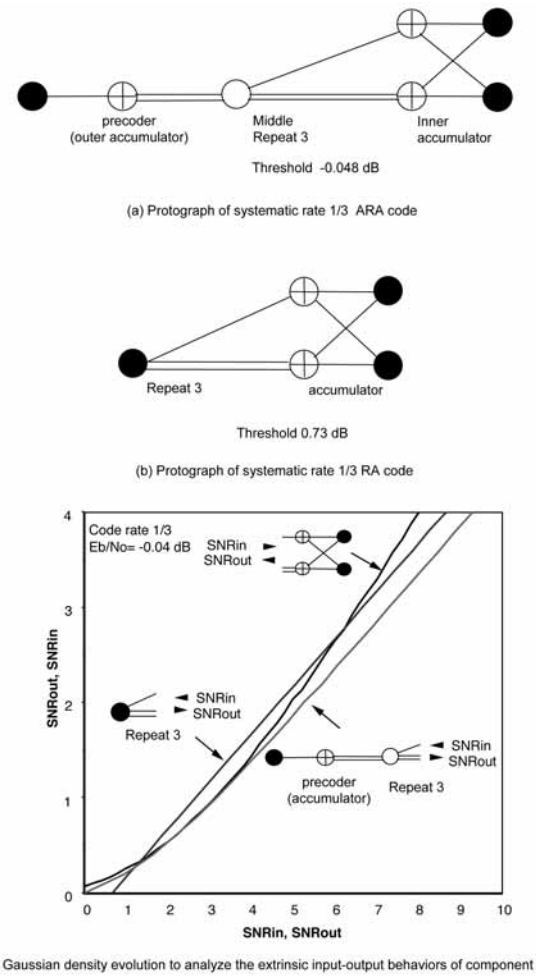


Fig. 2. Ensemble of LDPC codes; unstructured, multi edge and protograph versions.

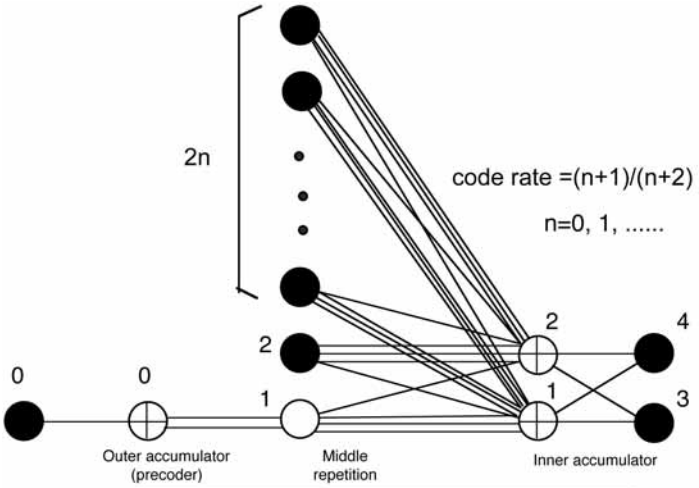
These comparisons are fair if we fix the maximum variable node degree. Shortly we will demonstrate such comparisons with unstructured irregular LDPC codes. In a similar way we can construct rate 1/2 and higher ARA codes. However, as we increase the code rate, the number of check nodes are decreased, and degree of inner checks (corresponding to the accumulator part) may increase beyond three. In such case we require some portion of the input bits not to be precoded in order to allow the iterative decoding to start, i.e., "**Doping**" [24] is required for iterative decoder to start. An example of a simple rate 1/2 ARA code, its protograph, and the corresponding threshold are shown in Figure 4 when $n = 0$. We also propose a constructive method to design codes with higher rates from a rate 1/2 ARA code and its protograph by adding $2n$ additional variable nodes each with degree 4 to the protograph. This is like adding $2n$ repeat 4 RA codes to a single rate 1/2 ARA code. In this case the addition is done at the inner check nodes of rate 1/2 ARA, and one common inner accumulator is used. One example of such a family of ARA codes for various code rates is shown in Figure 4.



(c) Gaussian density evolution to analyze the extrinsic input-output behaviors of component codes

Fig. 3. (a) Protograph of systematic rate 1/3 ARA, (b) systematic RA code, and (c) extrinsic message analysis using Gaussian density evolution showing the improvement due to the use of precoder.

We can also use irregular repetition instead of regular repetition in ARA codes, which results in Irregular ARA codes or simply IARA codes. A simple example of a rate 1/2 IARA code was discussed in [4]. It was shown that the threshold for that IRA example was 0.99 dB. The precoded version (IARA) has a threshold of 0.36 dB. Thus 0.63 dB improvement. In [4], an example of low threshold (0.264 dB) rate 1/2 ARA code was provided. The protograph for this example has a maximum degree of 5. The best rate 1/2 unstructured irregular LDPC code with a maximum degree of 5 in [5] has a threshold of 0.72 dB. There are few reasons for such a difference. In [5], the degree of variable nodes are greater or equal to 2, and punctured variable nodes were not allowed. If we look at the protographs of ARA codes, they contain degree 1 variable nodes and punctured variable nodes. In fact, later Richardson and Urbanke [11], [13] mentioned that the degree 1 variable nodes and punctured variable nodes also can be used in the multi edge type LDPC code design. As we mentioned earlier, protograph based ARA codes use permutations per each edge of protograph. When we have accumulators a protograph based ARA can be constructed if the structure of permutation in ARA codes is based on edge connections in the ARA protograph between middle variable nodes and inner check nodes (right most check nodes), i.e., between repetition and accumulator. However a "multi edge" type ARA code can be defined when permutations are used per group of edges, e.g., a single permutation is used for repetition.



Code Rate	Protograph Threshold	Capacity	Difference
1/2	0.560	0.187	0.373
2/3	1.414	1.059	0.355
3/4	1.980	1.626	0.354
4/5	2.396	2.040	0.356
5/6	2.717	2.362	0.355
6/7	2.980	2.625	0.355
7/8	3.197	2.845	0.352
8/9	3.385	3.033	0.352

Fig. 4. Protograph of ARA code family with repetition 4, for rates 1/2 to 8/9, and the corresponding iterative decoding thresholds in dB.

Simulation results for protograph ARA code family in Figure 4 with circulant permutations per each edge of protograph are shown in Figure 5 for code rates 1/2 through 7/8 . Use of circulant permutations instead of random permutations is preferred for implementation of encoder and decoder. For decoding of ARA codes the message passing algorithm was used. Low rate ARA codes can be constructed by concatenating an LDGM code with an RA code [20]. For example a rate 1/10 ARA code is constructed with iterative decoding threshold of −1.028 dB. For performance simulation results of protograph ARA codes for various code rates see [19], and [20]. For encoders for protograph based ARA codes with circulant permutations see [22].

III. Follow up Contributions to ARA Codes

The most significant contribution to ARA codes was made by Henry Pfister and Igal Sason [25]. They proved that ensembles of accumulate-repeat-accumulate codes are capacity-achieving codes for the erasure channel with bounded complexity. The next contributing factor emerged from a JPL coding research team including; Kenneth Andrews, Dariush Divsalar, Sam Dolinar, Jon Hamkins, Chris Jones, and Jeremy Thorpe (now with Google Inc.). This group noticed that if at least one of degree 2 variable nodes in the inner accumulator of ARA code is replaced with a degree 3 node, the resulting protograph code will have a linear minimum distance property. They called this new codes Accumulate Repeat Jagged Accumulate (ARJA) protograph LDPC codes [23]. A family of ARJA codes for various code rates proposed to Consultative Committee for Space Data Systems (CCSDS) for Standard. Concatenation of an LDGM (accumulator) code with a regular (3,6) LDPC code was also considered in [23] but with smaller improvement.

Conclusion

In this paper we proposed a new channel coding scheme called Accumulate Repeat Accumulate codes (ARA). This class of codes can be viewed as a subclass of Low Density Parity Check (LDPC) codes with fast encoder structure, and they have a projected graph or protograph representation, which allows for high speed iterative decoding implementation using belief propagation. Based on density evolution for protograph based ARA codes, we have shown that for maximum

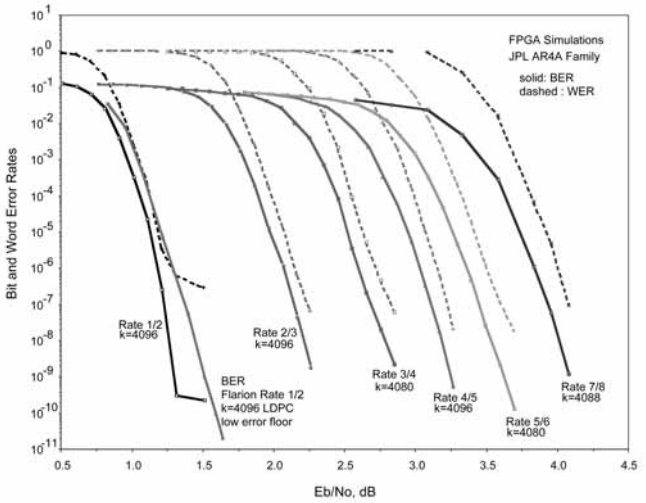


Fig. 5. Simulation results for examples of rate 1/2 and higher ARA codes.

variable node degree 5 a minimum bit SNR as low as 0.08 dB from channel capacity for rate 1/2 can be achieved as the block size goes to infinity . Such a low iterative decoding threshold cannot be achieved by RA, IRA, or unstructured irregular LDPC codes with the same constraint on the maximum variable node degree. We constructed families of higher rate protograph based ARA codes with iterative decoding thresholds that stay close to their respective channel capacity thresholds uniformly. The weight distribution of some simple multi edge type ARA codes is obtained in [4]. Through existing tightest bounds it was shown in [4] that the ML performance of ARA codes approaches very closely the performance of random codes.

References

[1] R.G. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: MIT Press, 1963.

[2] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes,'' *IEEE Trans. Commun.*, vol. 44, pp. 1261–1271, Oct. 1996.

[3] M.R. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*,'' vol. 27, pp. 533–547, 1981.

[4] A. Abbasfar, D. Divsalar, and K. Yao, "Accumulate-repeat-accumulate codes," *IEEE Trans. on Communications*, vol. 55, no. 4, pp. 692–702, Apr. 2007.

[5] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, 2001.

[6] S.-Y. Chung, D. Forney, T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit,'' *IEEE Communication Letters*, vol. 5, pp. 58–60, 2001.

[7] D. Divsalar, H. Jin, and R. McEliece, "Coding theorems for Turbo-like codes," in *Proceedings of the 1998 Allerton Conference*, 1998, pp. 201–210.

[8] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd International Symposium on Turbo Codes*, 2000, pp. 1–8.

[9] J. Thorpe, "Low Density Parity Check (LDPC) Codes Constructed from Protographs," JPL INP Progress Report 42–154, Aug. 15, 2003.

[10] T. Richardson et al., "Methods and apparatus for decoding LDPC codes," United States Patent 6,633,856, Oct. 14, 2003.

[11] T. Richardson, " Multi-Edge Type LDPC Codes," presented at the Workshop honoring Prof. Bob McEliece on his 60th birthday (but not included in the proceedings), California Institute of Technology, Pasadena, California, May 24–25, 2002.

[12] D.J.C. MacKay and R.M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronics Letters*, vol. 32, Issue 18, 29 Aug. 1996, Page(s) 1645.

[13] T. Richardson and R. Urbanke,"The Renaissance of Gallager's Low-Density Parity-Check Codes," *IEEE Communications Magazine*, pp. 126–131, Aug. 2003

[14] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, "Analysis of low density codes and improved designs using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, 2001.

[15] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message- passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, 2001.

[16] Y. Kou, S. Lin, and M.P.C. Fossorier, "Low-density parity-check codes based on finite geometries: A rediscovery and new results," *IEEE Transactions on Information Theory*, volume: 47 Issue: 7, Nov. 2001, Page(s): 2711–2736.

[17] F.R. Kschischang, "Codes defined on graphs," *IEEE Communications Magazine*, vol. 41, Issue 8, Aug. 2003, Pages 118–125.

[18] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. Info. Theory*, vol. 44, pp. 909–926, May 1998.

[19] D. Divsalar, S. Dolinar, J. Thorpe, and C. Jones, "Constructing LDPC Codes from Simple Loop-Free Encoding Modules," *IEEE ICC 2005*, Seoul Korea, 16–20 May 2005.

[20] D. Divsalar, S. Dolinar, J. Thorpe, and C. Jones, "Low-rate LDPC Codes with Simple Protograph Structure," *IEEE International Symposium on Information Theory 2005*, Adelaide, Australia 4–9 Sept., 2005.

[21] D. Divsalar and F. Pollara, "Hybrid Concatenated Codes and Iterative Decoding," *IEEE International Symposium on Information Theory 97*, Germany July 1997.

[22] K. Andrews, S. Dolinar, and J. Thorpe, "Encoders for Block-Circulant LDPC Codes,'' *IEEE International Symposium on Information Theory 2005*, Adelaide, Australia 4–9 Sept., 2005.

[23] D. Divsalar, C. Jones, S. Dolinar, and J. Thorpe, "Protograph Based LDPC Codes with Minimum Distance Linearly Growing with Block Size," IEEE Globecom 2005.

[24] S. ten Brink, "Code doping for triggering iterative decoding convergence," *Proc. IEEE Int. Symp. Inform. Theory*, July 2001, p. 235.

[25] H. Pfister and I. Sason, "Accumulate-repeat-accumulate codes: Capacity-achieving ensembles of systematic codes for the erasure channel with bounded complexity," *IEEE Trans. on Information Theory*, vol. 53, no. 6, pp. 2088–2115, June 2007.

Golay, Heisenberg and Weyl

Plenary talk presented at the 2007 IEEE International Symposium on Information Theory, Toronto Canada

Robert Calderbank

I would like to begin by thanking the organizers of ISIT 2008 for their invitation to speak, and Daniela Tuninetti for her invitation to share elements of the story with readers of the IT Newsletter. I would like to thank my sponsors, the Division of Mathematical Sciences at NSF, AFOSR, ONR and DARPA for their support and guidance. It is also a real pleasure to acknowledge the contributions of my collaborators, with special thanks to Stephen Howard from the Australian Defence Science and Technology Organisation, Bill Moran from Electrical Engineering at the University of Melbourne, and Ali Pezeshki from Electrical and Computer Engineering at Colorado State University.

The advent of phased array radars and space-time adaptive processing has given radar designers the ability to make radars adaptable on receive. The current state of radar technology allows the transmission of wavefields that vary across space, polarization, time and frequency and which can be changed in rapid succession. The ability to exploit space-time adaptive processing is limited by the computational power available at the receiver and optimal design only makes things worse unless the waveforms are properly designed to simplify processing at the receiver.

Exciting applications are emerging that are separate from defense. For example, the NSF Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere is prototyping a network of radars with the aim of changing the time scale on which the public is warned of severe weather. As our planet warms we expect to see a greater number of extreme climate events, such as tornados, hail storms, flooding, and wildfires, yet the curvature of the earth means that the US network of weather radars only provides universal coverage of phenomena above 10,000 feet and misses extreme weather emerging at low altitude.

My story starts sixty years ago with efforts by Marcel Golay [2], [3] to improve the sensitivity of far infrared spectrometry that led to the discovery of pairs of complementary sequences. Shortly thereafter Welti proposed to use Golay sequences in radar, but they have found very limited application to date. I hope to convince you that suitably transmitted and processed, radar waveforms based on Golay sequences provide new primitives for adaptive transmission that enable better detection and finer resolution, while managing computational complexity at the receiver.

Sixty years ago saw the birth of Information Theory and Coding in 1948, but measurement is a much older subject. Hamming codes were discovered in 1950, but the dual codes, the first order Reed Muller codes were discovered in 1942 by the famous statistician Fisher in the context of design of experiments. Earlier still is the development of weighing designs by Frederick Yates [8], where we are given a measurement device with mean zero and variance σ^2 and we want to minimize the number of measurements we need to make in order to measure each of N objects with variance σ^2/N. This idea of measuring objects in combination is characteristic of some of the most exciting modern work in detection and estimation. One example is the single pixel camera developed at Rice University which uses a micro-mirror array to form random linear combinations of pixels. We follow Golay and take our linear combinations in the frequency domain.

In the middle of the 20th century, infrared spectrometry was a challenge. There was a kind of twilight zone between measurements that could easily be made optically and those that could easily be made with radio-frequency techniques and this gap is described by Golay in his 1952 paper *Bridges across the infrared-radio gap*. Golay published two different designs for far infrared spectrometers designed to operate at room temperature, and we will examine the second in a little more detail.

Golay faced three problems; weak sources, strong background noise and insensitive detectors—temperature sensors that could not by themselves distinguish between different frequencies of infrared radiation and essentially integrated energy across the entire band. Optical spectrometers employ diffraction, so that different frequencies appear at different angles, and slits are used to separate these frequencies. Narrow slits improve discrimination but do not pass enough energy for the detector to work well. Wide slits pass enough energy but do not provide sufficient frequency discrimination.

Golay split the infrared radiation into two paths and made each path pass through an entry and exit mask as shown in Figure 1. The path label specifies the slit pattern in the entry mask. The desired frequency, indicated in black, arrives orthogonal to the masks, and the background radiation, indicated in gray, arrives obliquely. The desired radiation from path 1 proceeds to the detector as shown, and the desired radiation from path 2 is blocked. The background radiation casts a gray shadow on each exit mask that is a translation of the corresponding input mask. The two sequences x and y are such that the amount of radiation (at each background frequency) arriving at the detector is the same for both paths. Subtraction now provides a measurement of radiation at the desired frequency. The mathematical restriction that is placed on the sequences x and y is simply that the sum of the two autocorrelation functions is a delta function (that is $E_S = E_1 - E_2$ if and only if $R_x(k) + R_y(k)$ is a delta function).

Golay Complementary Waveforms in Radar

In active sensing we illuminate the scene with a waveform and analyze the return to detect the presence of scatterers, estimate their range from the round trip delay and estimate velocity from the Doppler Effect. In the absence of Doppler, consider the problem of estimating range from the round trip delay. We are looking for peaks when the return is matched to the illuminating waveform. The ideal would be an autocorrelation function that is impulse-like, so we want a narrow mainlobe with lots of energy and we want to suppress range sidelobes because we worry about a strong scatterer masking a weak scatterer that is close by.

When we incorporate Doppler, the ambiguity function may be viewed as the point spread function of the radar imaging system. It

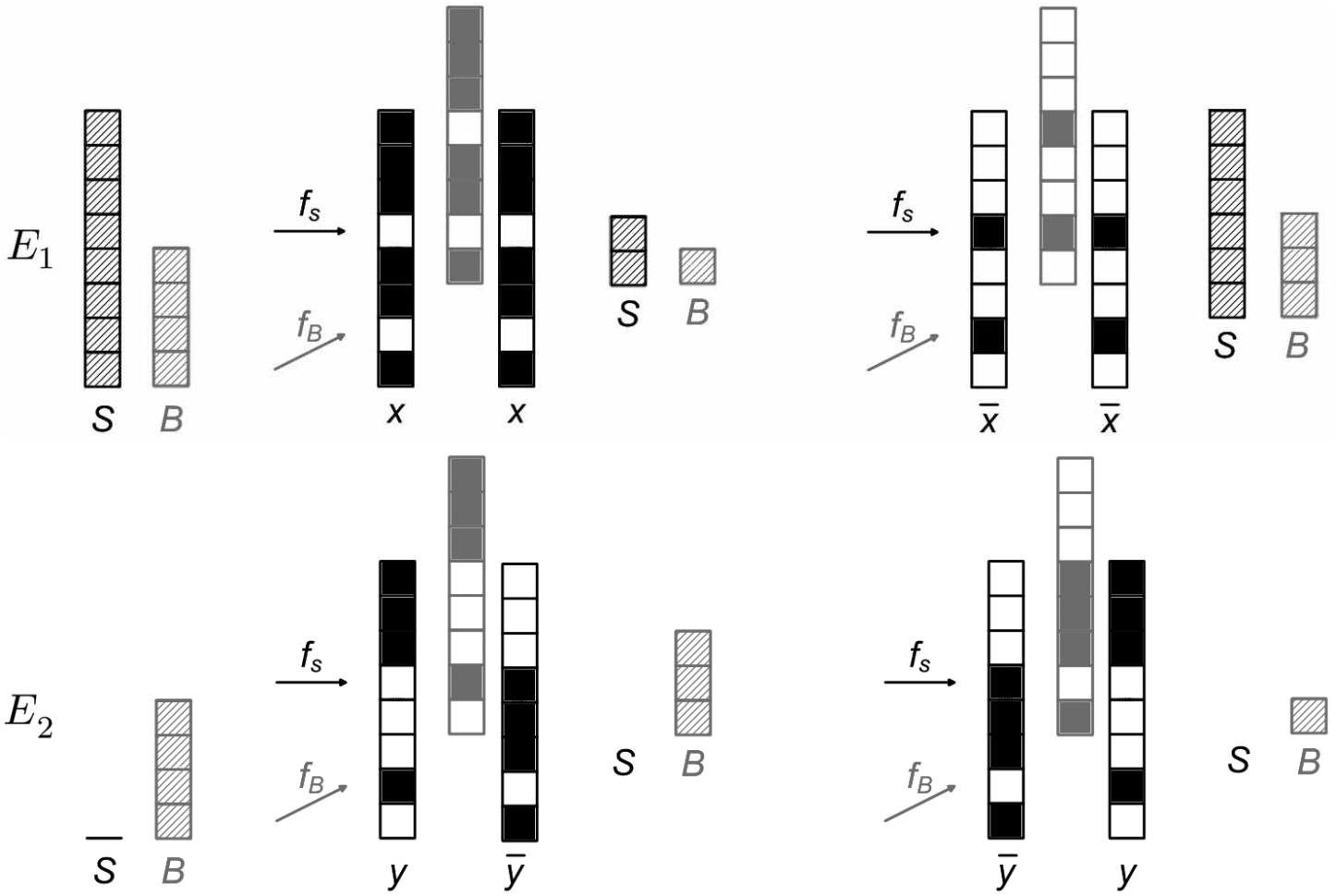


Fig. 1. The origin of Golay Complementary Sequences.

captures the blurriness of the image as a function of the radar waveform, when matched filtering is used at the receiver. Moyal's Identity places a lower bound on the volume under the squared ambiguity surface as a function of the energy in the signal. It encapsulates in a slightly different guise the Heisenberg Uncertainty Principle. The ideal ambiguity function would be a thumbtack, but since that is not possible, the aim of the radar engineer is to move the inevitable volume under the squared ambiguity function to regions where it matters less for the operational task of the radar. Figure 2 shows the ambiguity function for the Barker code of length 13 and the left half of the ambiguity surface viewed from above. Note the range sidelobes at zero Doppler appearing in black.

By separating waveforms in time we reap the usual advantages that attach to modular design in engineering. The waveforms are short (order 1 μs) with respect to the Pulse Repetition Interval or PRI (order 100 μs) and we can ignore Doppler at the chip level.

It was in 1960 that Welti proposed the use of Golay sequences in pulsed radar. Fifty years later they are nowhere to be seen, and every practicing radar engineer *knows* that they *don't* work because of sensitivity to Doppler. The problem is that the ambiguity function of a Golay pair of phase coded waveforms is free of range sidelobes only at zero Doppler, and off the zero Doppler axis it has large sidelobes in range. This means that a weak target located near a strong target can be masked by the range sidelobes of the ambiguity function centered on the strong target.

However all is not lost, and the freedom to sequence different Golay pairs makes possible the design of pulse trains for which the composite ambiguity function maintains ideal shape inside a specific Doppler interval [6]. The trick is to sequence the Golay pairs so that the pulse train ambiguity function has a higher order null at the Doppler frequency where we want to eliminate range sidelobes. We employ a binary sequence p to coordinate transmission of the components of a Golay pair, resulting in the ambiguity function

$$\underbrace{\frac{1}{2}[R_x(k)+R_y(k)]\sum_{n=0}^{2^M-1}e^{jn\theta}}_{\text{Sidelobe free}}+\underbrace{\frac{1}{2}[R_x(k)-R_y(k)]\sum_{n=0}^{2^M-1}(-1)^{p_n}e^{jn\theta}}_{\text{Range sidelobes}}$$

When we separate mainlobe from range sidelobes we see that the magnitude of the range sidelobes is determined by the spectrum

$$S_{\mathbf{p}}(\theta)=\sum_{n=0}^{2^M-1}(-1)^{p_n}e^{jn\theta}$$

This should be very familiar to those who have worked on the design of spectral null codes for magnetic recording, and in fact this is a more forgiving problem because we do not care about

Ambiguity Function:

$$A_s(\tau,\nu) = \int_{-\infty}^{\infty} s(t)\overline{s(t-\tau)}e^{-j2\pi\nu t}dt$$

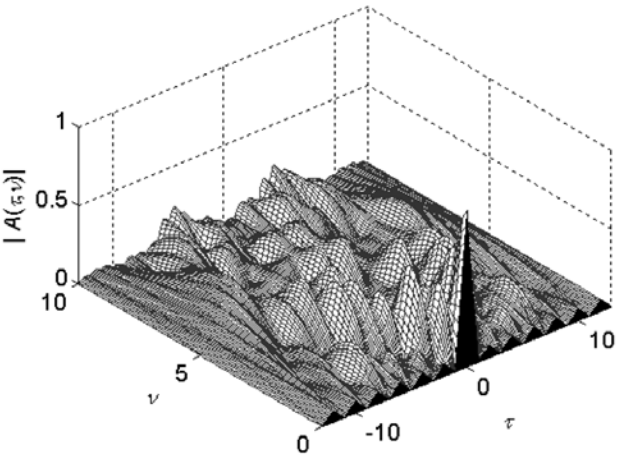


Fig. 2. Radar imaging.

rate. If we focus on creating a higher order null at zero Doppler, then we are led to a famous sequence in number theory.

Prouhet-Thue-Morse Sequence

The nth term in the PTM sequence p_n is the sum of the binary digits of n mod 2:

n	(0) = 0000	(1) = 0001	(2) = 0010	(3) = 0011
p_n	0	1	1	0

x	y	y	x	y	x	x	y
0	1	1	0	1	0	0	1

The PTM sequence of length 2^{M+1} kills M Taylor moments and classical coding theorists will have noticed from the iterative construction that this sequence is a codeword in the first order Reed Muller code. The PTM pulse train can bring out slow-moving weak scatterers such as pedestrians out of the shadow of strong stationary reflectors such as buildings (see Figure 3).

We have only explored time as a degree of freedom, but this approach extends to polarization and space. The basic unit of illumination is a matrix of phase coded waveforms indexed by transmit antenna and by PRI, where the polarization of constituent waveforms may vary. It turns out that choosing this matrix to be a unitary filter bank simplifies signal processing considerably, and in two dimensions the filter bank has the same structure as the Alamouti block space-time code. The generalization to multiple dimensions uses the filter banks introduced by Tseng and Liu [7] to study acoustic surface waves, and we refer the interested reader to [1] for more details.

In 1953 Woodward introduced the narrowband radar ambiguity function in his book *Probability and Information Theory with Applications to RADAR* where he acknowledges the influence of Shannon's work. He ends the book with the following quote: *The reader may feel some disappointment, not unshared by the writer, that the basic question of what to transmit remains substantially unanswered.* It is extraordinarily self-deprecating quote, and I think ultimately incorrect. Looking to the future, it is possible to envisage unitary filter banks as a new illumination paradigm that enables broad waveform adaptability across time, space, frequency and polarization.

Heisenberg-Weyl Groups and Sequence Design

Where do Golay complementary sequences come from? The mathematical framework for sequence design turns out to be the framework for quantum error correction. The error group of an m-qubit quantum system is the m-fold Kronecker product of the dihedral group D_4 extended by multiplication by the square root of -1. The notation $D(a, b)$ keeps track of how the Kronecker product is formed; the binary vector a captures appearances of the bit flip x and the binary vector b captures appearances of the phase flip z. There are 2^{2m+2} matrices, all of them square to I or $-I$, any pair of matrices commute or anticommute and a simple symplectic form governs which possibility occurs.

$$\mathrm{xz} \otimes \mathrm{x} \otimes \mathrm{z} \otimes \mathrm{xz} \otimes I_2 \leftrightarrow D(\overset{a}{11010}, \overset{b}{10110})$$

$$D(a,b)D(a',b') = (-1)^{a'.b+b'.a}D(a',b')D(a,b)$$

$$D(a,b)^2 = (-1)^{a.b}I_{2^m}$$

A different Heisenberg-Weyl group provides the mathematical foundation for the world of Fourier analysis. The operators $\Delta(k, 0)$ are cyclic time shifts, the operators $\Delta(0, j)$ are the corresponding frequency shifts, and these two groups are interchanged by the Fourier transform. In the binary world the operators $D(a, 0)$ are the time shifts, the operators $D(0, b)$ are the frequency shifts, and these two groups are interchanged by the Walsh-Hadamard transform. Think of Walsh functions as the sinusoids of the binary world. Why sinusoids? It is because they are eigenfunctions of the time shift operators.

What are chirps in this binary world? The answer is second order Reed Muller codewords (realized over Z_4 as in [5]). Each coset of the first order Reed Muller code corresponds to an orthonormal basis of the underlying Hilbert space. The cosets are indexed by binary symmetric matrices P and the vector

$$\phi(x) = i^{xPx^T+2bx^T}, x \in \mathbb{Z}_2^m$$

is an eigenvector of the commutative subgroup of operators $D(a, aP)$. This is entirely analogous to the properties of classical chirps (see [4]).

The operators $D(a, b)$ are an orthonormal basis for the space of operators with respect to the trace inner product, and the Weyl transform gives the basis expansion of an arbitrary operator.

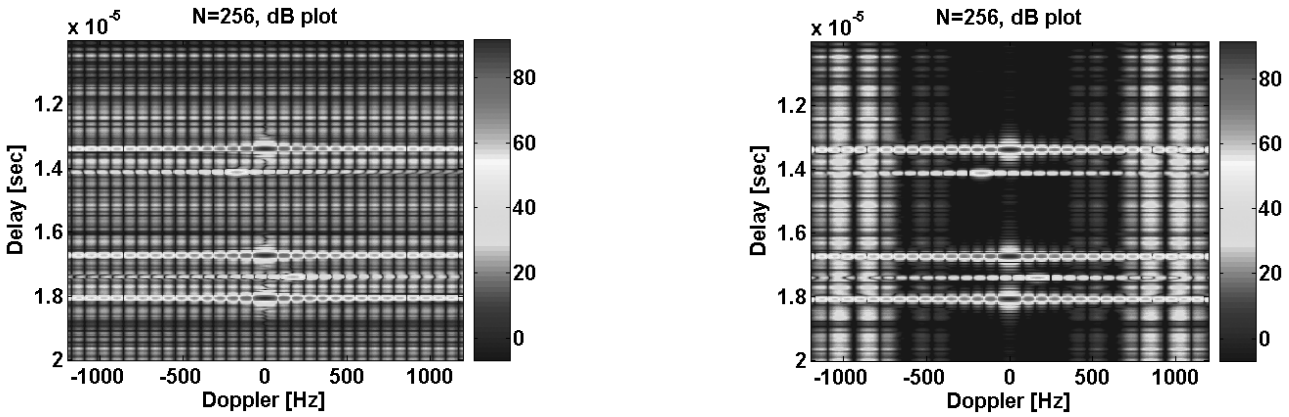


Fig. 3. By transmitting a Golay pair according to the PTM sequence we can clear out the range sidelobes along modest Doppler frequencies.

Rather than look at sequences, we are going to work with the corresponding rank 1 projection operators. We connect periodic and aperiodic correlation by padding with zeros and we express the Golay complementary property as follows:

$$\theta^{\dagger}\Delta(k,0)\theta + \varphi^{\dagger}\Delta(k,0)\varphi = 0 \quad \text{for } k \neq 0$$
$$Tr((P_{\theta} + P_{\varphi})\Delta(k,0)) = 0 \quad \text{for } k \neq 0$$

We want to remove the overlap between the support of the cyclic time shifts and the support of the sum of the rank 1 projection operators. If we start with a sequence that has a maximal commutative symmetry group, then the support of the corresponding projection operator vanishes outside this subgroup. We can also calculate (in the $D(a,b)$ basis) the union of the supports of the Weyl transforms of the cyclic time shift operators. The answer turns out to be a pair of Sirpinski triangles! Now all we have to do is find a maximal commutative subgroup (coset of $RM(1,m)$) that has minimal intersection with the triangles, and choose two eigenvectors so that the support of the sum of the corresponding projection operators has empty intersection. This is how Golay complementary pairs are made.

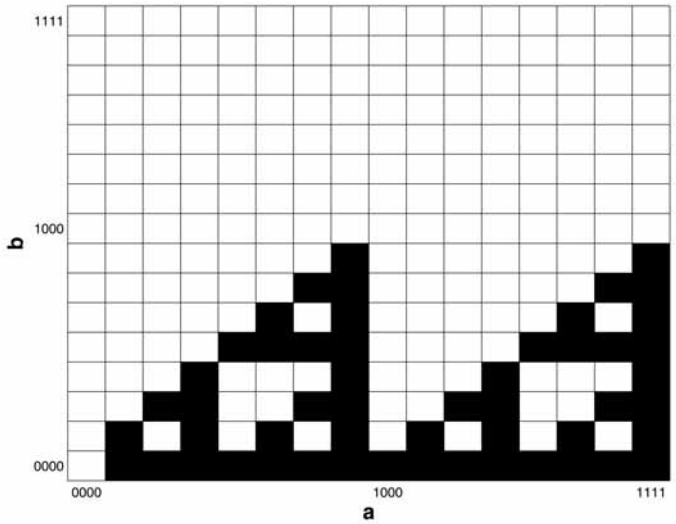


Fig. 4. The support of the time shift operators $\Delta(k,0)$ for $m = 4$.

Postscript I would like to conclude with one final thought, and that is that classical coding theory and Fourier analysis are just different sides of the same coin.

References

[1] A.R. Calderbank, S.D. Howard, and W. Moran, "Waveform diversity in radar signal processing," *IEEE Signal Processing Magazine*, to appear, 2009.

[2] M.J.E. Golay, "Multi-slit spectrometry," *Journal of the Optical Society of America*, vol. 39, no. 6, p. 437, 1949.

[3] M.J.E. Golay, "Static multi-slit spectrometry and its application to the panoramic display of infrared spectra," *Journal of the Optical Society of America*, vol. 41, no. 7, pp. 468–472, 1951.

[4] S.D. Howard, A.R. Calderbank, and W. Moran, "The finite Heisenberg-Weyl groups in radar and communications," *EURASIP Journal on Applied Signal Processing*, Article ID 85685, 2006.

[5] A.R. Hammons, Jr., P.V. Kumar, A.R. Calderbank, N.J.A. Sloane and P. Sole, "The Z_4-linearity of Kerdock, Preparata, Goethals and related codes," *IEEE Transactions on Information Theory*, vol. IT-40, pp. 301–319, 1994.

[6] A. Pezeshki, A.R. Calderbank, W. Moran and S.D. Howard, "Doppler resilient waveforms with perfect autocorrelation," *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4254–4266, Sept. 2008.

[7] C.C. Tseng and C.L. Liu, "Complementary sets of sequences," *IEEE Transactions on Information Theory*, vol. 18, no. 5, pp. 644–652, 1972.

[8] F. Yates, "Complex experiments," Supplement to the *Journal of the Royal Statistical Society*, vol. 2, no. 2, pp. 181–247, 1935.

Building Quantum Computers

Plenary talk presented at the 2007 IEEE International Symposium on Information Theory, Toronto Canada

E. Knill

In theory, quantum computers can be used to efficiently factor numbers, quadratically speed up many search and optimization problems, and enable currently impossible physics simulations. At first, quantum states appeared to be too fragile for implementing large quantum computers. Fortunately, because of theoretical advances in quantum error correction and fault tolerance, there are now no fundamental obstacles to realizing quantum computers. However, building quantum computers is difficult. Current experiments can barely achieve adequate control of two quantum bits. Nevertheless, the gap between theoretical and practical quantum computing is closing. In what follows, I give a brief explanation of what quantum computers are, explain why we believe that in principle, arbitrarily large quantum computations can be accurately implemented, and survey the experimental state of the art and main implementation challenges.

A simple way to think of a quantum computer is as a traditional, *classical* computer with access to a quantum state machine. Thus quantum computing is an extension of classical computing with all the classical programming constructs available for problem solving and control of the quantum state machine. The quantum state machine is specified by its state space, initial state, transition operators and readout operators. It can be thought of as the result of applying the superposition principle to the 2^n *configurations* (bit strings) of n bit systems together with the ability to exploit interference, where n may vary during a computation. In particular, the state space consists of the unit vectors in a 2^n-dimensional Hilbert space with a distinguished orthonormal basis, whose elements are denoted by $|b\rangle$ with b bit strings of length n. The unit vectors can therefore be written as *superpositions* $|\psi\rangle = \sum_b \alpha_b |b\rangle$, where the complex numbers α_b are called the amplitudes of the superposition and $\sum_b |\alpha_b|^2 = 1$. For $n = 1$, the state space is that of a *quantum bit* (*qubit* for short). Just as bit strings of length n are the configurations of n bit systems, the superposition states $|\psi\rangle$ are considered to be states of n qubits. This is done by identifying the 2^n-dimensional Hilbert space with the tensor product of the n 2-dimensional Hilbert spaces associated with the qubits. The distinguished basis is obtained from the tensor products of the distinguished basis elements of the component qubits. Note that it is necessary to clearly distinguish between systems (such as qubits) and their states. This also makes it easier to understand the relationship between the formal definition of our state machines and their physical realizations.

The initial state of a quantum state machine has no qubits. To add qubits, we can make use of a transition operator that maps the state of n qubits $\sum_b \alpha_b |b\rangle$ to the state of $n+1$ qubits $\sum_b \alpha_b |b0\rangle$, where $b0$ is the length $n+1$ bit string obtained by appending 0. The representation of the states of a quantum state machine as the states of n qubits is important for defining unitary transition operators that may be applied to the states. One such operator is the Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

acting on one qubit. H can be applied to the k'th of n qubits by tensoring with identity operators acting on the other $n-1$ qubits. Another is the Toffoli gate, which acts on three qubits by linear extension of the map on three bits that flips the third bit if the first two are 1. To define the linear extension, bit strings are identified with the corresponding distinguished basis elements. The Toffoli gate can be applied to three of n qubits by tensoring with identity operators acting on the remaining qubits. The Hadamard and Toffoli gates are sufficient for quantum computing. Nevertheless it is convenient to be able to apply any one-qubit unitary and the controlled-not gates. The controlled-not gate is the linear extension of the map that flips the second bit if the first one is 1. The Toffoli gate can be decomposed into a product of one-qubit unitary and controlled-not gates.

Information about the state of a quantum state machine is obtained by measurement. Suppose that the state of the machine is $\sum_b \alpha_b |b\rangle$. A full, destructive measurement returns the bit string b with probability $|\alpha_b|^2$ and resets the machine to its initial state. It is convenient, but not necessary, to be able to make nondestructive measurements of any one of the qubits. To learn how such measurements act, and for an introduction to quantum computing, see, for example, Nielsen and Chuang's textbook [10].

A phenomenon that is often mentioned as a source of the power of quantum computing is quantum parallelism, which involves the application of a classical reversible algorithm implemented by Toffoli gates "simultaneously" to all bit patterns in a superposition with exponentially many non-zero amplitudes. This is simply the generalization of the linear extension principle by which we defined the Toffoli gate. Transition operators such as the Hadamard gate must be used to prepare the state. Because the measurement cannot access amplitudes except by an exponentially complex analysis of the statistics of measurement outcomes, any use of such quantum parallelism must be followed by large scale interference of the state's amplitudes to extract the desired information. Interference refers to the effect by which one can reversibly increase amplitudes in some states in a way that is sensitive to relative phases. For example, the Hadamard gate applied to $|0\rangle$ yields the state $\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, which when measured returns 0 or 1 with equal probability. Applying it again restores both the state $|0\rangle$ and determinism of the measurement outcome. If a process that changes the sign of the amplitude of $|1\rangle$ is applied before the second Hadamard gate, the final state is $|1\rangle$, demonstrating the sensitivity of the interference effect in the second Hadamard gate to the relative phases of the amplitudes. It is worth contrasting these effects to what is possible with probabilistic computing, where instead of superpositions involving amplitudes, we have *mixtures* involving probabilities of states. Gates correspond to Markov processes, which are reversible only if they are deterministic.

Building quantum computers requires physical systems with quantum state spaces that are capable of realizing qubit states and are sufficiently controllable. DiVincenzo [6] gives five require-

ments for the physical realization of quantum computers that correspond to the earlier specifications of a quantum state machine. The first is the availability of arbitrarily many independent quantum-information-carrying systems. The second requires that the quantum systems' state can be consistently intialized. Skipping the third requirement for now, the fourth asks for the ability to apply quantum operations sufficient for implementing arbitrary quantum computations efficiently. The fifth involves the ability to measure the systems so as to enable the required readout. These four requirements have been demonstrated individually in a number of physical systems.

The third and so far the most difficult requirement to demonstrate experimentally is that the states and operations are subject to sufficiently low noise. The continuous nature of the amplitudes and the sensitivity of interference effects to seemingly small changes in phases imply that quantum states and gates must be protected not only from bit flip errors, but from a continuous family of unwanted effects including changes in the phases of amplitudes. These unwanted effects are referred to as *decoherence*. Decoherence is associated with incomplete isolation from the environment and imperfect calibration of control fields required to implement gates. The microscopic nature of most suitable quantum systems and the need for strong interactions with the control and measurement apparatus makes it particularly difficult to reduce the effects of decoherence.

Like quantum gates, general quantum errors exhibit interference effects that preclude purely probabilistic models. Nevertheless, we commonly refer to gates having independent probabilities of error. This is justified if unwanted effects act independently in time and space and are unbiased. Although this is not generally the case, actions can be taken to increase independence and decrease bias. Alternatively, it is understood that the probability refers to the square of an amplitude in the operators expressing the effect of an error. Originally it was believed that in order to realize a quantum computation of size N, the probability of error per gate must be sufficiently smaller than $1/N^2$, where the square accounts for the possibility that errors add in amplitude rather than probability. However, as for classical computing with errors, it has been proven that under reasonable assumptions on the errors, if the probability of error per gate is smaller than some constant, then it is possible to efficiently quantum compute arbitrarily accurately. This result is known as the threshold theorem. See Sect. 10.6 of [10] for an overview of quantum fault tolerance and versions of this theorem. Since there are many ways to parameterize quantum error models and many physical constraints (such as spatial layout) to consider, the error threshold claimed by the theorem is best understood as defining a region of the relevant space of parameters and constraints where *scalable* quantum computing is possible in principle. Note that if the parameters are near the boundary of this region, the overhead required for implementing computations fault tolerantly becomes impractical.

Fault tolerant quantum computing involves using quantum error-detecting and -correcting codes to protect quantum information. To maintain and compute with the protected quantum information, we use carefully designed sequences of gates that ensure that any errors in the gates themselves do not disturb the protected information. All schemes for protecting quantum or classical information can be understood in terms of *subsystems*. Consider the trivial problem of protecting one qubit when we are given three physical qubits, where only the first two are subject to errors. The solution is to have the third qubit carry the protected information. The third qubit is a subsystem of the three-qubit physical system. Protected states are associated not with single states but with subspaces of states of the physical system, and the errors preserve these subspaces. Formally, a quantum subsystem of a physical system whose state space consists of unit states in the Hilbert space $\mathcal{H}$ is a tensor factor of a subspace of $\mathcal{H}$. The other factor is called the *cosubsystem*. Equivalently, finite quantum subsystems are characterized by subalgebras of the algebra of bounded operators on $\mathcal{H}$, where the subalgebras are ismorphic to matrix algebras. From a physical point of view, such subalgebras of operators consist of the (complex) *observables* of the subsystem and characterize the measurements that one can make of the states of the subsystem. The general scheme for protecting information is to determine a subsystem of the physical system that has the property that, provided the cosubsystem's state is suitably prepared, errors perturb only the cosubsystem's state with high probability. If the cosubsystem's state does not matter, then no action needs to be taken to maintain protection. Otherwise, it is necessary to periodically restore the cosubsystem to a state that ensures future protection. In the traditional view, this action is accomplished by error correction and re-encoding. From the subsystem view, the protected information never requires "correction"; it is sufficient to reset the cosubsystem after errors occurred. One can think of errors as increasing the entropy of the cosubsystems, and the protection procedure as a way of removing the entropy. Therefore, physical quantum computers generally require an entropy sink to protect information from errors.

The analysis of fault-tolerant quantum computing leads to strategies for eventually building large-scale quantum computers. Most suitable physical systems consist of localized quantum subsystems with at least two distinguishable states that can represent qubit states. These physically explicit qubits are normally subject to a significant amount of decoherence. The first task is to ensure sufficient control of the physical systems, including the ability to *couple* them, and to use whatever experimental techniques are available to reduce the effects of decoherence to the point where general-purpose error-correction techniques can be applied according to the threshold theorem. Eventually, fault-tolerant techniques are used to protect *logical* qubit subsystems that are nontrivially supported by many physical systems. Depending on the errors, it may be necessary to recursively construct qubit subsystems of lower-level logical qubits, a strategy known as *concatenation*. It helps to recognize that there are a number of common information processing tasks that are much easier to perform fault tolerantly than implementing unitary gates on logical qubits. These tasks include state preparation, measurement and quantum communication. In fact, the constraints on errors in physical operations used for these tasks are significantly weaker than on errors in unitary control. Thus, provided there is some means of establishing quantum communication channels between physical systems used to support logical qubits, one can initially focus on building very accurate quantum registers with only a small number (three or four) of qubits. One can rely on communication for computations requiring more

qubits. Unlike classical communication, quantum communication and remote computation can be performed by what is known as quantum *teleportation*, which has the advantage of having no quantum latency. This implies that the speed of remote computation is not limited by slow quantum processes, only by the classical communication required for control. Although focusing on small but accurate quantum registers makes sense now, the ultimate goal is to ensure that good quantum gates are not much slower than classical circuit elements. This will require a tight integration of quantum and classical processing and fault tolerance.

Targeted experimental efforts to build quantum computers started with Shor's discovery of the quantum algorithm to factor large integers around 1994. Since then there have been many proposals to build quantum computers using a variety of physical systems. For a survey, see [1]. The clear current front runner for building small to medium size quantum computers is based on atomic qubits in ion traps. There are currently three other approaches that can claim to have demonstrated coherent two-qubit control: Liquid state nuclear magnetic resonance (NMR) quantum computing, postselected photonic qubits, and superconducting qubits. Of these approaches, the first two have little hope of constructing quantum registers with more than about ten qubits, because of inherent exponential inefficiencies that require a significant change or addition to the underlying technology.

To be able to usefully solve problems currently infeasible on classical computers with known quantum algorithms requires thousands of qubits and billions of gates. Although up to eight qubits have been nontrivially manipulated with atomic qubits in ion traps, at this point no one has clearly demonstrated a computationally useful two-qubit register. It is expected that this will be achieved shortly in ion traps.

In ion-trap quantum computing, the physical qubits are represented by two energy levels of ions that are electromagnetically trapped. The ions can be manipulated by means of laser pulses. The combination of the trapping potential and Coulomb repulsion leads to common vibrational modes that can be exploited for applying nontrivial two-qubit gates. This approach to quantum computing can be scaled by having multiple traps with the ability to move ions between them as proposed by Wineland and coauthors [12]. All but the requirement for sufficiently low noise have been individually demonstrated. There are three main challenges for experimental ion-trap quantum computing. The first is to realize gates with sufficiently low error probabilities. Error probabilities of about 0.5% have been demonstrated for two-qubit gates [2]. The current guidelines for demonstration of the low-noise requirement are to have less than 0.01% probability of error per unitary gate. State preparation and measurement can have probabilities of error of 1%, which has been demonstrated in ion traps. The second challenge is to show that all the requirements can be met in one device. This is a problem of technology integration and is typically much harder than demonstrating each requirement independently. The third challenge is to have an efficient way of quantum communicating between ion-trap quantum registers, preferably by optical interconnects. The first steps in this direction have been taken by Moehring and coauthors [8].

Superconducting qubits are based on the collective dissipation-less behavior of electrons in superconducting circuits. There are a number of different ways to design such circuits to exhibit the desired two-level subsystems needed to represent qubits. For reviews of the relevant physics, see [4], [5]. It was not clear whether the collective effects were experimentally accessible until some coherent control and measurement of qubits in superconducting circuits was demonstrated by Nakamura and coworkers [9]. Unexpectedly, experiental quantum computing with superconducting qubits is progressing rapidly and has overtaken other seemingly more promising approaches. A possible advantage of superconducting qubits is that it is possible to have gates that are much faster than is practical with atomic qubits. Because noise also acts on shorter time scales, this is also a challenge, requiring high-quality control involving very fast electronics. At this time, slow gates are an advantage as the electronics required for control is off-the-shelf. The path toward large scale quantum computing with superconducting qubits is not yet as well defined as for atomic qubits in ion traps, so the requirements have not been demonstrated as clearly. Because current realizations of superconducting qubits require temperatures well below 1 K, the challenge of integrating technology seems more severe at the moment. Communication with devices in separate refrigeration units is also difficult and no means for doing so has been demonstrated so far.

There are many other approaches to building quantum computers that are being investigated experimentally. Promising ones include atomic qubits of trapped atoms in optical lattices [3] and various quantum-dot-based schemes [7], both of which have two-qubit gate demonstrations in progress. There are also esoteric approaches, such as topological quantum computing based on anyonic excitations, which is claimed to be intrinsically robust against noise. Whether and where these excitations can be found in experimentally accessible condensed matter phases is a subject of theoretical controversy and experimental investigation [11].

Since the challenge of building quantum computers has no analogue in the history of computing, this is a great time to be doing research in quantum technologies. There are many theoretical and experimental problems to be solved and challenges to be met, and although difficult, they are likely surmountable. The associated improvements in quantum control have wide applicability beyond quantum computing proper. Assuming no fundamental physics surprises, which would of course be welcome, I expect the use of quantum mechanics in practical technology and computation to become pervasive.

Acknowledgments

This is a contribution of the National Institute of Standards and Technology, an agency of the U.S. government, and is not subject to U.S. copyright.

References

[1] Special issue on implementations of quantum computers. *Fort. Phys.*, vol. 48, no. 9–11, 2000.

[2] J. Benhelm, G. Kirchmair, C.F. Roos, and R. Blatt, "Towards fault-tolerant quantum computing with trapped ions," *Nature Phys.*, vol. 4, pp. 463–466, 2008.

[3] I. Bloch, "Quantum coherence and entanglement with ultracold atoms in optical lattices," *Nature*, vol. 453, pp. 1016–1022, 2008.

[4] J. Clarke and F. Wilhelm, "Superconducting quantum bits," *Nature*, vol. 453, pp. 1031–1042, 2008.

[5] M.H. Devoret, A. Wallraff, and J.M. Martinis, Superconducting qubits: A short review. quant-ph / 0411174, 2004.

[6] D.P. DiVincenzo, "The physical implementation of quantum computation," *Fort. Phys.*, vol. 48, pp. 771–783, 2000.

[7] R. Hanson and D.D. Awschalom, "Coherent manipulation of single spins in semiconductors," *Nature*, vol. 453, pp. 1043–1049, 2008.

[8] D.L. Moehring, P. Maunz, S. Olmschenk, K.C. Younge, D.N. Matsukevich, L.-M. Duan, and C. Monroe, "Entanglement of single-atom quantum bits at a distance," *Nature*, vol. 449, pp. 68–71, 2007.

[9] Y. Nakamura, Y.A. Pashkin, and J.S. Tsai, "Coherent control of macroscopic quantum states in a single-cooper-pair box," *Nature*, vol. 398, pp. 786–788, 1999.

[10] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2001.

[11] S. Das Sarma, M. Freedman, and C. Nayak, "Topological quantum computation," *Phys. Today*, vol. 59, pp. 32–38, 2006.

[12] D.J. Wineland, C. Monroe, W.M. Itano, D. Leibfried, B.E. King, and D.M. Meekhof, "Experimental issues in coherent quantum-state manipulation of trapped atomic ions," *J. Res. Nat. Inst. St. Tech.*, vol. 103, pp. 259–328, 1998.

Randomness—A Computational Complexity Perspective

Plenary talk presented at the 2007 IEEE International Symposium on Information Theory, Toronto Canada

Avi Wigderson

Abstract

Man has grappled with the meaning and utility of randomness for millennia. Randomness is paramount to many fields of science, and in particular plays a key role in information theory and coding theory. Here we explore some key aspects of the meaning and utility of randomness in computation.

1. Introduction

The marriage of randomness and computation has been one of the most fertile ideas in computer science, with powerful consequences in a wide variety of areas ranging from cryptography to computational learning theory to distributed computing. It enabled new understanding (and uses) of fundamental concepts such as knowledge, secret, learning, proof, and indeed, randomness itself.

In this short note we discuss the role of randomness in algorithms, and its impact on their efficiency. It is taken, with some modifications, from my survey "P, NP and Mathematics" [10][1]. This survey contains motivation, intuition and precise definitions of computational complexity theory for the interested reader. In particular, it discusses the rich field (which we do not discuss here) that evolved from studying the role of randomness in *proofs*, leading to such paradoxical (and highly useful) notions of *zero-knowledge* proofs and *holographic* (or *probabilistically checkable*) proofs. Other good sources on randomness in computation are the books [6], [16] and the relevant chapters in [20].

Finally we remark on the connections to Information Theory and Coding Theory. Randomness is of course an integral part of both, in modeling and technique. Besides the well known presence of randomness and probabilistic algorithms in Coding Theory, e.g., in the probabilistic construction of codes and in probabilistic decoding algorithms, there are many new connections between these fields and Computational Complexity Theory, benefitting both sides. Many of these were born out of the computational study of randomness in computation. Specific areas of interaction include list-decodable codes, locally-decodable codes and expander codes, for which computer science interest arose partly from the study of probabilistic proofs, randomness extractors and derandomization. Some of these connections and consequences to coding are discussed in the monographs [9], [24], [25].

2. Preliminaries

We briefly and informally describe some of the complexity theoretic notions used in this paper, including Turing machines and Boolean circuits, and the classes $\mathcal{P}$ and $\mathcal{NP}$. More precise definitions can be found in [10], as well as standard texts on Computational Complexity such as [7], [18].

Algorithms are discussed informally, but assuming some formal underlying computational model such as a Turing machine. The specific choice of model is not important, since we will consider the time efficiency of algorithms up to polynomial factors. Computational problems are abstracted as functions f, where given an input x, the task is computing $f(x)$. Inputs are encoded as binary strings; this representation induces a length measure on inputs to f. The time complexity of an algorithm $t(n)$ is measured asymptotically as the number of steps performed by the algorithm on (the worst) input of length n. Algorithms (or programs) were traditionally viewed as completely deterministic, and here we will explore enhancing them with access to random coin tosses.

The class $\mathcal{P}$, is the class of functions f for which some algorithm computes f in time polynomial in n. This class is commonly used to capture problems which have efficient[2] solutions.

The class $\mathcal{NP}$ captures in some intuitive sense many of the problems we would like to solve. These are all problems for which correct solutions can be *verified* efficiently (namely in polynomial time). A canonical example of a problem in $\mathcal{NP}$ is SAT, the satisfiability problem. Here the input is a Boolean formula over, say, n Boolean variables. The task is to determine if some assignment to these variables makes the formula evaluate to "True". Obviously, given a guess of such assignment, one can easily verify its correctness by evaluating the formula. But brute for search over all assignments will take exponential time in n.

The famous $\mathcal{P}$ versus $\mathcal{NP}$ question, namely "is $\mathcal{P} = \mathcal{NP}$?", asks if brute-force search can be avoided in all these problems. For example, can some clever algorithm always find satisfying assignment in polynomial time (if one exists). Indeed, this is no mere example; the problem SAT is an $\mathcal{NP}$-complete problem, which means it is "hardest" in the class $\mathcal{NP}$. One important manifestation of this fact is that $\mathcal{P} = \mathcal{NP}$ *if and only if* SAT is in $\mathcal{P}$. $\mathcal{NP}$-complete problems pervade mathematics and all sciences, making this question so central. While having (on the face of it) nothing to do with randomization, we'll later see that it does.

Finally, we'll need another computational model, that of Boolean circuits. Informally, if Turing machines capture software, circuits capture hardware. Boolean circuits compute a *finite* function on

[1] This and other surveys mentioned below are available for personal use from the authors' home pages.

[2] The "loaded" word *efficient* and its numerous interpretations across computational models, resources and applications is the focus of computer science, and the class $\mathcal{P}$ just happens to be one extremely important manifestation of efficiency.

(say) n input bits via a sequence of Boolean gates (a standard set is $\{AND, OR, NOT\}$). To compute functions on arbitrary input lengths one must specify a *family* of circuits, one for every input length n. As before, the size complexity of such a circuit family $s(n)$ is measured asymptotically as the number of gates in the circuit for size n inputs. Circuits are often called "non-uniform" algorithms, since in circuit families no specific relation need exist between the circuits for different input lengths (while in "uniform" Turing machines, the same program handles at once all input lengths).

The relation between the two models is interesting. Circuit families can simulate Turing machines efficiently. But the converse is false, due to this "nonuniformity" in circuit. To see this, note that every function on n bits has a 2^n size circuits (e.g., by expressing the function in disjunctive normal form). Thus *every* function (including undecidable ones) can be computed by circuit families. However, for functions in $\mathcal{NP}$ there is a general belief (substantiated in a weak technical sense which is beyond the scope of this article) that circuit size and Turing machine time behave similarly. No subexponential size (namely $2^{o(n)}$) upper bound is known for e.g., SAT. Indeed, one major direction to proving that $\mathcal{P} \neq \mathcal{NP}$, is attempting to prove the stronger result, that $\mathcal{NP}$ does not even have polynomial size circuits. Like all other attempts on this major problem, not much progress has been made, although very interesting lower bounds were obtained for restricted families of circuits, e.g., monotone and constant-depth circuits. Circuit lower bounds will play a role in the results below.

3. Randomness in Algorithms

The following two sections tell the contradicting stories on the power and weakness of algorithmic randomness.

3.1. The Power of Randomness in Algorithms

Let us start with an example, which illustrates the potential algorithmic power of randomness. It concerns a problem that comes up naturally in many mathematical areas, namely the discovery, verification and proof of algebraic identities. Assume we work here over the field of rationals $\mathbb{Q}$. The $n \times n$ Vandermonde matrix $V(x_1, \ldots, x_n)$ in n variables has $(x_i)^{j-1}$ in the (i, j) position. The Vandermonde Identity is:

Proposition 3.1. $\det V(x_1, \ldots, x_n) \equiv \prod_{i<j}(x_i - x_j)$.

While this particular identity is simple to prove, many others like it are far harder. Suppose you conjectured an identity $f(x_1, \ldots, x_n) \equiv 0$, concisely expressed (as above) by a short arithmetic formula say, and wanted to know if it is true before investing much effort in proving it. Of course, if the number of variables n and the degree d of the polynomial f are large (as in the example), expanding the formula to check that all coefficients vanish will take exponential time and is thus infeasible. Indeed, no subexponential time algorithm for this problem is known! Is there a quick and dirty way to find out?

A natural idea suggests itself: assuming f is *not* identically zero, then the variety of zeros it defines has measure zero, and so if we pick *at random* values to the variables, chances are we shall miss it. If f is identically zero, every assignment will evaluate to zero. So a "random" point in $\mathbb{Q}^n$ will distinguish the two cases with high probability. It turns out that this idea can be made precise, by restricting random choices to a finite domain, and the following can be simply proved:

Proposition 3.2 ([21], [28]). *Let f be a nonzero polynomial of degree at most d in n variables. Let r_i be uniformly and independently chosen from $\{1, 2, \ldots, 3d\}$. Then* $\Pr[f(r_1, \ldots, r_n) = 0] \leq 1/3$.

Note that since evaluating the polynomial at any given point is easy given a formula for f, the above constitutes an efficient *probabilistic* algorithm for verifying polynomial identities. Probabilistic algorithms differ from the algorithms we have seen so far in two ways. First, they postulate the ability to toss coins and generate random bits. Second, they make errors. The beauty is, that if we are willing to accept both (and we should!), we seem to be getting far more efficient algorithms for seemingly hard problems.

The deep issue of whether randomness exists in nature has never stopped humans from assuming it anyway, for gambling, tie breaking, polls and more. A fascinating subject of how to harness seemingly unpredictable *weak sources of randomness* (such as sun spots, radioactive decay, weather, stock-market fluctuations or internet traffic) and converting them into a uniform stream of independent, unbiased coin flips, is the mathematical study of *randomness extractors* which we shall not describe here (see the excellent survey [22]). We shall postulate access of our algorithms to such perfect coin flips, and develop the theory from this assumption. We note that whatever replaces these perfect random bits in the numerous practical implementations of probabilistic algorithms seems empirically to work pretty well.

The error inherent in probabilistic algorithms seems a more serious issue—we compute to discover a *fact*, not a "maybe". However, we do tolerate uncertainty in real life (not to mention computer hardware and software errors). Furthermore, observe that the error of probabilistic algorithm is much more controllable—it can be decreased arbitrarily, with small penalty in efficiency. To see this, assume an algorithm makes error at most $1/3$ on any input (as the one above). Then running it k times, with independent random choices each time, and taking a majority vote would reduce the error to $\exp(-k)$ on every input!

Thus it is natural to revise the notion of efficient computation to allow probabilistic algorithms with small error, and define the probabilistic analog $\mathcal{BPP}$ (for Bounded error, Probabilistic, Polynomial time) of the class $\mathcal{P}$.

Definition 3.3 (The class $\mathcal{BPP}$, [5]). The function f is in $\mathcal{BPP}$ if there exists a probabilistic polynomial time algorithm A, such that for every input x, $\Pr[A(x) \neq f(x)] \leq 1/3$.

Again, we stress that this probability bound is over the internal coin-tosses of the algorithm, and holds for *every* input. Moreover, replacing the error probability $1/3$ by $\exp(-|x|)$, namely expo-

nentially small in the input length, leaves the definition of $\mathcal{BPP}$ unchanged (by the amplification idea above)[3].

Probabilistic algorithms were used in statistics (for sampling) and physics (Monte Carlo methods) before computer science existed. However, their introduction into computer science in the 1970s, starting with the probabilistic primality tests of Solovay–Strassen [23] and Rabin [19], was followed by an avalanche that increased the variety and sophistication of problems amenable to such attacks tremendously—a glimpse to this scope can be obtained e.g., from the textbook [16]. We remark again that here we restrict ourselves only to those probabilistic algorithms which save *time*, and note that randomness seems to help save other computational resources as well!

We list here a few sample problems which have probabilistic polynomial time algorithms, but for which the best known deterministic algorithms require exponential time. These are amongst the greatest achievements of this field.

- **Generating primes ([19], [23]).** Given an integer x (in binary), produce a prime in the interval $[x, 2x]$ (note that the prime number theorem guarantees that a random number in this interval is a prime with probability about $1/|x|$). Verification that the generated number is prime can be done efficiently either probabilistically with the algorithms mentioned above, or deterministically, with the recent breakthrough of [2] which we will mention later again.

- **Polynomial factoring ([15]).** Given an arithmetic formula describing a multivariate polynomial (over a large finite field), find its irreducible factors[4].

- **Permanent approximation ([13]).** Given a nonnegative real matrix, approximate its permanent[5] to within (say) a factor of 2.

- **Volume approximation ([4]).** Given a convex body in high dimension (e.g., a polytope given by its bounding hyperplanes), approximate its volume[6] to within (say) a factor of 2.

The most basic question about this new computational paradigm of probabilistic computation, is whether it really adds any power over deterministic computation.

Open Problem 3.4. Is $\mathcal{BPP} = \mathcal{P}$?

The empirical answer is an emphatic *NO*: we have no idea in sight as to how to solve the problems above, and many others, even in subexponential time deterministically, let alone in polynomial time. However, the next subsection should change this viewpoint.

3.2. The Weakness of Randomness in Algorithms

Let us start from the bottom line: if any of the numerous problems in $\mathcal{NP}$ is *hard*, then randomness is *weak*. There is a tradeoff between what the words *hard* and *weak* formally mean. To be concrete, we give perhaps the most dramatic such result of Impagliazzo and Wigderson [11].

Theorem 3.5 ([11]). *If SAT cannot be solved by circuits of size $2^{o(n)}$ then $\mathcal{BPP} = \mathcal{P}$. Moreover, SAT can be replaced in this statement by any problem which has $2^{O(n)}$-time algorithms*[7].

Rephrasing, exponential circuit lower bounds on essentially any problem of interest imply that randomness can be *always* eliminated from algorithms without sacrificing efficiency (up to polynomial). Many variants of this result exist. Weakening the assumed lower bound does weaken the deterministic simulation of randomness, but leaves it highly nontrivial. For example, if $\mathcal{NP}$ does not have polynomial-time circuits, then $\mathcal{BPP}$ has deterministic algorithms with subexponential runtime $\exp(n^{\varepsilon})$ for every $\varepsilon > 0$. Moreover, analogs are known where the hardness assumption is uniform (of the type $\mathcal{P} \neq \mathcal{NP}$), e.g., [12].

Note the remarkable nature of these results: they show that if one computational task is hard, than another is easy!

We are now faced with deciding which of two extremely appealing beliefs to drop (as we discover that they are contradictory!). Either that natural problems (e.g., $\mathcal{NP}$-complete ones) cannot be solved efficiently, or that randomness is extremely powerful. Given that our intuition about the former seems far more established, we are compelled to conclude that randomness cannot significantly speed up algorithms, and indeed $\mathcal{BPP} = \mathcal{P}$.

Conjecture 3.6. $\mathcal{BPP} = \mathcal{P}$.

We now turn to give a high level description of the ideas leading to this surprising set of results, which are generally known under the heading *Hardness vs. Randomness*[8]. We refer the reader to the surveys in [6], [20] for more.

We are clearly after a general way of eliminating the randomness used by any (efficient!) probabilistic algorithm. Let A be such an algorithm, working on input x, and using as randomness the uniform distribution U_n on binary sequences of length n. Assume A computes a function f, and its error on any input is at most $1/3$. The idea is to "fool" A, replacing the distribution U_n by another distribution D, without A noticing it!

This leads to the key definition of *pseudorandomness* of Yao [26].

[3] This small error easily implies that *some* fixing of the randomness will not err on *any* input of a fixed length n. This is the core of a result of Adleman [1], that every problem in $\mathcal{BPP}$ has polynomial size circuits. It is another demonstration of the power of nonuniformity.

[4] Note that it is not even clear that the output has a representation of polynomial length—but it does!

[5] Unlike its relative, the determinant, which can be easily computed efficiently by Gauss elimination, the permanent is known to be $\#\mathcal{P}$-complete (which implies $\mathcal{NP}$-hardness) to compute exactly.

[6] Again, computing the volume exactly is $\#\mathcal{P}$-complete.

[7] This class includes most $\mathcal{NP}$-complete problems, but far more complex ones, e.g., determining optimal strategies of games, not believed to be in $\mathcal{NP}$.

[8] The title of Silvio Micali's PhD thesis, who with his advisor Manuel Blum constructed the first hardness based pseudorandom bit generator.

Definition 3.7 (Pseudorandomness, [26]). Call a distribution D *pseudorandom* if no efficient process[9] can "tell it apart"[10] from the uniform distribution U_n.

By the definition, any such distribution is as good as U_n, as A's computation on x *is* an efficient process.

Remark 3.8. This definition specializes a more general one of *computational indistinguishability* between probability distributions, which originates in the landmark paper of Goldwasser and Micali [8]. This key *behavioristic* definition of randomness underlies the mathematical foundations of modern cryptography which are laid out in that paper. We also note that computational indistinguishability suggests a metric on probability distributions which is a coarsening of the usual statistical distance (L_1 norm) or informational divergence. Computationally indistinguishable distributions may have drastically different entropies, and we make full use of it below.

Back to our derandomization task. Can we efficiently generate a pseudorandom distribution D from only very few random bits? Specifically, we would like to compute $D = G(U_m)$ where G is a deterministic polynomial time algorithm and $m \ll n$. Such functions G which produce pseudorandom distributions from short random *seeds* are called *pseudorandom generators*. With them, a deterministic simulation will only need to enumerate all possible 2^m seed values (rather than the trivial 2^n). For each such seed it will use the output of G as "randomness" for the computation of A on x, and take a majority vote. As the error of A was at most $1/3$ under U_n, and A's output probability changes by at most $1/9$ between D and U_n, the new error is at most $4/9$, so the majority vote will correctly compute $f(x)$, for *every* x. If m gets down to $O(\log n)$, then $2^m = n^{O(1)}$, and this becomes a deterministic polynomial time algorithm.

But how can we construct such a pseudorandom generator G? Since the definition of pseudorandomness depends on the computational limitations of the algorithm, one might hope to embed some hard function g into the workings of the generator G, and argue as follows. If an efficient process can distinguish the output of G from random, we shall turn it into an efficient algorithm for solving the (assumed hard) function g. This yields a contradiction.

Thus the heart is this conversion of hardness into pseudorandomness. The two main different methods for implementing this idea are the original generator of Blum–Micali and Yao [3], [26] (which must use hard functions g with very special structure, like factoring or discrete logarithm), and the one by Nisan–Wigderson [17] (which can use any hard function g that has an exponential time algorithm). We note that here the hardness required of g is of the *average-case* variety, which is either assumed in the former, or has to be obtained from *worst-case* hardness in the latter. Thus this field invents and uses new types of efficient *reductions*, translating nonstandard computational tasks (from distinguishing a random and pseudorandom distributions, to computing a function well on average, to computing it in the worst case.

We note that this very general line of attack may benefit from specialization. We saw that to derandomize a probabilistic algorithm all we need is a way to efficiently generate a low entropy distribution which *fools* it. But fooling a specific, given algorithm may be easier than fooling them all. Indeed, careful analysis of some important probabilistic algorithms, (specifically, the way they use their randomness), has enabled their derandomization via tailor-made generators. These success stories (of which the most dramatic is the recent deterministic primality test of [2]) actually suggest the route of probabilistic algorithms and then derandomization as a paradigm for *deterministic* algorithm design. More in the textbook [16].

Finally, let us remark on recent progress regarding this mysterious connection between hardness and randomness. In the theorems above, it works in one way. Given a hard function, we can derandomize. A recent remarkable result proves a partial converse. Kabanets and Impagliazzo [14] showed that derandomizing (the extremely specific and simple) probabilistic algorithm embodied in Proposition 3.2 above is *equivalent* to proving certain circuit lower bounds. This news may be taken negatively, saying we are unlikely to prove unconditional derandomization results, or positively, indicating another route to proving lower bounds, namely via derandomization.

[9] This can mean an algorithm or a circuit.

[10] E.g., produce a given output with noticeably different probability, say $1/9$.

References

[1] L. Adleman, "Two Theorems about Random Polynomial Time," *Proceedings of 19th IEEE Symposium on Foundations of Computer Science*, 1978, pp. 75–83.

[2] M. Agrawal, N. Kayal, and N., Saxena, "Primes is in $\mathcal{P}$," *Ann. of Math.*, vol. 160, no. 2, pp. 781–793, 2004.

[3] M. Blum and S. Micali, "How to generate cryptographically secure sequences of pseudorandom bits," *SIAM J. Comput.*, vol. 13, pp. 850–864, 1984.

[4] M. Dyer, A. Frieze, and R., Kannan, "A random polynomial time algorithm for approximating the volume of a convex body," *J. ACM*, vol. 38, no. 1, pp. 1–17, 1991.

[5] J. Gill, "Computational complexity of probabilistic Turing machines," *SIAM J. Comput.*, vol. 6, pp. 675–695, 1977.

[6] O. Goldreich, *Modern Cryptography, Probabilistic Proofs and Pseudorandomness*. Algorithms Combin. 17, Springer-Verlag, Berlin 1999.

[7] O. Goldreich, *Computational Complexity, a Conceptual Perspective*. Cambridge University Press, 2008.

[8] S. Goldwasser and S., Micali, "Probabilistic encryption," J. *Comput. System Sci.*, vol. 28, pp. 270–299, 1984.

[9] V. Guruswami, *Algorithmic Results in List Decoding*, Foundations and Trends in Theoretical Computer Science, vol. 2, no. 2, NOW Publishers, 2007.

[10] A. Wigderson, "$\mathcal{P}$, $\mathcal{NP}$ and Mathematics—a computational complexity perspective," *Proceedings of the International Congress of Mathematicians*, vol. I, EMS publishing, 2007, 665–712.

[11] R. Impagliazzo and A. Wigderson, "P = BPP unless E has Subexponential Circuits: Derandomizing the XOR Lemma," *Proceedings of the 29th annual ACM Symposium on Theory of Computing*, ACM Press, New York 1997, 220–229.

[12] R. Impagliazzo and A. Wigderson, "Randomness vs. Time: De-randomization under a uniform assumption," *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, IEEE Comput. Soc. Press, Los Alamitos, CA, 1998, 734–743.

[13] M. Jerrum, A. Sinclair, and E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries," *J. ACM*, vol. 51, no. 4, pp. 671–697, 2004.

[14] V. Kabanets and R. Impagliazzo, "Derandomizing Polynomial Identity Tests Means Proving Circuit Lower Bounds," *Comput. Complexity*, vol. 13, no. 1–2, pp. 1–46, 2004.

[15] E. Kaltofen, "Polynomial Factorization," in *Computer Algebra: Symbolic and Algebraic Computation*, 2nd ed., Springer-Verlag, Wien, New York 1983, 95–113.

[16] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, Cambridge 1995.

[17] N. Nisan and A., Wigderson, "Hardness vs. Randomness," *J. Comput. System Sci.*, vol. 49, no. 2, pp. 149–167, 1994.

[18] C.H. Papadimitriou, *Computational Complexity*. Addison Wesley, Reading, MA, 1994.

[19] M.O. Rabin, "Probabilistic algorithm for testing primality," *J. Number Theory*, vol. 12, pp. 128–138, 1980.

[20] S. Rudich and A. Wigderson, (eds.), *Computational Complexity Theory*. IAS/Park-City Math. Ser. 10, Institute for Advanced Studies/Amer. Math. Soc., 2000.

[21] J.T. Schwartz, "Fast probabilistic algorithms for verification of polynomial identities," *J. ACM*, vol. 27, no. 4, pp. 701–717, 1980.

[22] R. Shaltiel, "Recent Developments in Explicit Constructions of Extractors," *Bull. EATCS*, vol. 77, pp. 67–95, 2002.

[23] R.M. Solovay and V. Strassen, "A fast Monte-Carlo test for primality," *SIAM J. Comput.*, vol. 6, no. 1, pp. 84–85, 1977.

[24] M. Sudan, *Efficient Checking of Polynomials and Proofs and the Hardness of Approximation Problems*. ACM Distinguished Theses, Lecture Notes in Comput. Sci. 1001, Springer-Verlag, Berlin 1996.

[25] S. Vadhan, *A Unified Theory of Pseudorandomness*, SIGACT News, vol. 38, no. 3, 2007.

[26] A.C. Yao, "Theory and application of trapdoor functions," *Proceedings of the 23th annual IEEE Symposium on Foundations of Computer Science*, IEEE Comput. Soc. Press, Los Alamitos, CA, 1982, pp. 80–91.

[27] A.C. Yao, "How to generate and exchange secrets," in *Proceedings of the 27th annual IEEE Symposium on Foundations of Computer Science*, IEEE Comput. Soc. Press, Los Alamitos, CA, 1986, pp. 162–167.

[28] R.E. Zippel, "Probabilistic algorithms for sparse polynomials," in *Symbolic and algebraic computation* (EUROSCAM '79), Lecture Notes in Comput. Sci. 72, Springer-Verlag, Berlin 1979, pp. 216–226.

Recent Activities of the IT Student Committee

Brooke Shrader, Ivana Maric, Lalitha Sankar and Aylin Yener

First, we would like to note that The First Annual School of Information Theory was held the first week of June at the Penn State Campus, in University Park, PA. A detailed report of this exciting event was presented in the September issue of the newsletter. The preparations for the Second Annual School of Information Theory are underway and will be communicated in the subsequent issues.

Now onto the events at ISIT 2008: For the fourth consecutive year, the Student Committee hosted two well-attended student lunch events at ISIT:

The Roundtable Research Discussion and Lunch was held on Monday, July 7 during ISIT. There were over 130 students and post-docs in attendance - enough to fill the Grand Ballroom East. In fact, the turnout was so overwhelming that the lunch line extended far outside the room! We would like to thank the volunteer discussion leaders, who are listed next to the discussion topics below, for their enthusiastic participation.

- Graphical models - Alex Dimakis (UC Berkeley)
- Source coding - Krish Eswaran (UC Berkeley)
- Feedback - Ramji Venkataramanan (University of Michigan)
- Interference and Secrecy - Xiang He (Penn State)
- Relaying - Bobak Nazer (UC Berkeley)
- Distributed detection and estimation - Anima Anandkumar (Cornell)
- Scaling laws - Awlok Josan (University of Michigan)
- Zero-error Information Theory - Salim El Rouayheb (Texas A&M)
- Optimization - Chee Wei Tan (Princeton)

As in the previous years, a panel event was organized during the Thursday lunch hour. This year, the panel was dedicated to the memory of the late Professor Sergio Servetto, and was titled "What Makes a Great Researcher?" We were lucky to have the following great panelists: Thomas Cover and Andrea Goldsmith of Stanford, Sergio Verdu of Princeton University, Alon Orlitsky of UCSD and Alexander Barg of University of Maryland. The panel moderator was Aylin Yener of Penn State University. Once again the panel event attracted over 130 attendees. As has become a custom, free IT student committee T-shirts were distributed to the participants. We extend our gratitude to ISIT local arrangement chairs Raviraj Adve, and TJ Lim, both of University of Toronto, for their help in the organization of these events. We especially thank Raviraj Adve for his help with ordering T-shirts and lunches.

We are happy to announce that the student paper award - first proposed by the student committee, and presented for the second time at this conference following ISIT'07 in Nice - was awarded to three students: they were Paul Cuff (Stanford) for his paper titled "Communication Requirements for Generating Correlated Random Variables", Satish Babu Korada (EPFL) for his paper coauthored with Ruediger Urbanke (EPFL) and titled "Exchange of Limits: Why Iterative Decoding Works"; and to Yury Polyanskiy (Princeton) for his paper coauthored with H. Vincent Poor (Princeton) and Sergio Verdu (Princeton), titled "New Channel Coding Achievability Bounds". Congratulations to the winners!

Finally, we would like to express our deepest thanks and gratitude to two of the members of our committee, Brooke Shrader and Lalitha Sankar, who have been tireless volunteers on the student committee from the beginning and now moved on to the next phase in their lives. We congratulate Brooke, who recently received her PhD from University of Maryland and now is a scientist at MIT Lincoln Labs. We wish Brooke and Lalitha the best for their future.

At the same time, we welcome our new members, Krish Eswaran of UC Berkeley as our new student co-chair; Nan Liu of Stanford, Deniz Gunduz of Princeton / Stanford, and Xiang He of Penn State as our new event coordinators and members at large; and Matthieu Bloch of Notre Dame for web-related topics. We look forward to working with them towards new and exciting activities.

As always, we look for volunteers to join the committee and help with organization of various future events. If you would like to be a volunteer, please e-mail Aylin Yener at yener@ee.psu.edu, Ivana Maric at ivanam@wsl.stanford.edu or Krish Eswaran at keswaran@eecs.berkeley.edu. We welcome your input, comments and look forward to hearing from you.

Students at the round table event.

Panelists discussing attributes of a great researcher.

Update on the Society's Outreach Efforts

Muriel Medard

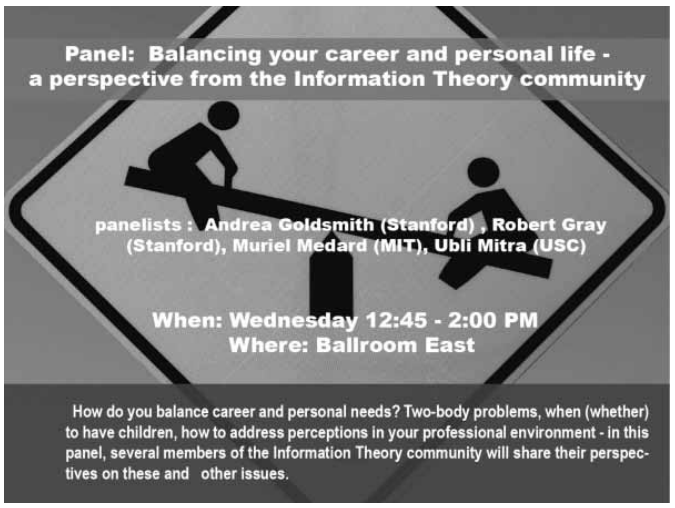
This year the Board of Governors decided at our first meeting of the year in Porto to back an initiative to increase our outreach to encourage diversification of our membership, particularly for women members, in a way that is inclusive and beneficial for all our members, to be led by me, with the advice and assistance of Bob Gray. In doing so, we are following the lead of successful programs in related fields. Many of our members may already be familiar with the very well attended ICASSP women's breakfast or with the networking community's N^2 Women (networking networking women) initiative (http://www.comsoc.org/N2Women/). The N^2 Women have helpfully documented what seem to be the main components of a successful initiative: events at main conferences, co-opting both senior and junior members of the community, a mailing list with very few but useful e-mails, and a visible web site hosted by the relevant society.

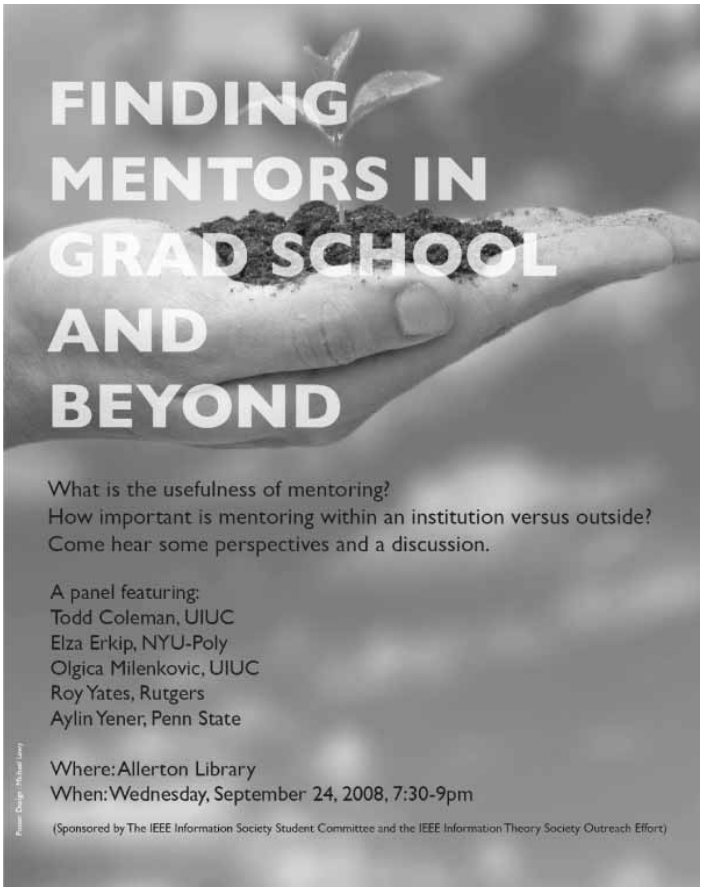
The level of participation of women in the Society provides a mixed picture. The participation of women in our society at different member levels is given in the table below. For the purpose of comparison, two of our sister societies, Signal Processing (SP) and Communications (COM) are also shown. The grades of membership are shown as follows GSM - graduate student member, Mem - member, SM - senior member, Fel - Fellow, LF - Life Fellow.

	GSM-F	Mem-F	SM-F	Fel-F	LF-F	GSM-M	Mem-M	SM-M	Fel-M	LF-M
IT	25	101	22	9	1	178	1613	435	204	137
SP	79	473	134	30	0	613	8176	1828	440	160
COM	330	1385	211	19	2	2331	22039	3458	606	327

The numbers bear out the pattern, often observed in science and engineering, of a shrinking pipeline. Overall, our numbers are comparable to those of our sister societies. The male/female multiplier at the different levels are, over all Fellows (Fel+LF), 34 for IT, 20 for SP, 44 for COM; for GSM they are around 7 for all three societies and for members they are all between 16 and 17. The only worrisome disparity with respect to other societies seems to occur at the senior member level, which is the pipeline for Fellows. The ratios at that level of membership are 19 for IT, 13 for SP and 16 for COM. While the number of Fellows in the last years has risen considerably (an inspection of the list of Fellows shows at least 7 in the last 8 years), it would appear the majority of those (possibly all - IEEE did not have the data available) rose to the level of Fellow through Societies other than IT. The historical data on participation of our women members in ISIT TPCs has been very variable. While no numbers are easily available, an inspection of the list of TPC members (which may suffer from some inaccuracy in my visual inspection, for which I apologize) shows that the numbers were 3/58 for 2005, 1/55 for 2006, 6/56 (and 2/4 TPC co-chairs) for 2007 and 3/63 for 2008. The trend of participation in the future may be increasing. Participation of women in leadership of the Society seems to be encouraging. The participation of women in the Board of Governors seems relatively robust recently, with about 3/20, including our President elect.

The topics of our meetings have been selected in keeping with our mission to provide events that address needs and encourage participation of our underrepresented demographics, while being of interest and use to the community at large. The topic of balancing work and life is often important in career decisions of women and, increasingly, in those of men as well. Our first event took place at ISIT in Toronto, on the topic of balancing career and personal life, with planned panelists of Andrea Goldsmith, Bob Gray, Ubli Mitra and myself. Fittingly, given the topic of the panel, I had





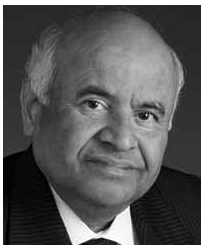
to rush home because of a very sick child, but Lalitha Sankar and Elza Erkip kindly agreed to participate on the panel. The attendance at the event was very encouraging. The attendance was of about 150-200, and the benefit of the event to the entire community was evidenced by the fact that men constituted one half or more of the audience.

The issue of finding mentoring is of particular importance for demographic groups with low representation. Our second event, spearheaded by Todd Coleman, took place at Allerton, on the topic of finding mentors in grad school and beyond, was done in collaboration with the student initiative, led by Aylin Yener. The roughly 40 participants included junior and senior faculty as well as graduate students and postdocs. The panelists (Todd Coleman, Elza Erkip, Olgica Milenkovic, Roy Yates, and Aylin Yener) and the participants discussed topics concerning how to find mentors within and outside one's institution, both during graduate school and one's professional career, and how to be a more effective mentor. Our sincere thanks go to the organizers of ISIT and Allerton 2008 for their help and support in making these events a success.

In order to help foster an online community, we have created a new mailing list withits@mit.edu (Women in the Information Theory Society), which has been collecting new names and which is open to all. This list will be used with great restraint, mainly to advertise future events. We are also collecting web pages from interested members of our Society to create a WithITS webpage, which we shall link to our Society web site. If you wish to be added to this mailing list or to have your web page added to the WithITS, please e-mail me at medard@mit.edu (please send the url of your webpage if you wish for it to be included). Feel free to contact me with suggestions for future events, in particular if you would be interested in helping to lead such an event.

Vijayfest - International Workshop on Advances in Communications

Aaron Gulliver, Ian F. Blake, and Vahid Tarokh



Vijay Bhargava

The International Symposium on Advances in Communications in honour of Vijay K. Bhargava on the occasion of his 60th birthday was held in Victoria, British Columbia on September 21-23, 2008. A prolific scholar, Vijay has made fundamental contributions to communication theory, coding theory, and wireless communications. He is also a great educator, which was recognized with the 2002 IEEE Graduate Teaching Award. During the period that he served as the Chairperson of the Department of Electrical and Computer Engineering at the University of British Columbia, the department underwent significant expansion including a new building, 20 new faculty and a doubling of student numbers.

Vijay is very active in the IEEE and in particular the IT Society. He has served as President of the Information Theory Society, Vice President of the Regional Activities Board, Director of Region 7, Montreal Section Chair and Victoria Section Chair. He is a past member of the Board of Governors of the IEEE Communications Society and the IEEE Information Theory Society. He organized ISIT'83 (St. Jovite) with Ian Blake and L. Lorne Campbell, and ISIT'95 (Whistler) with Ian Blake and Michael Pursley. In 1987 he founded the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, and in 1988 the Canadian Conference on Electrical and Computer Engineering. He was awarded the IEEE Haraden Pratt Award in 1999 and the IEEE Larry K. Wilson Transnational Award in 1996.

Moe Win, Norman Beaulieu, Vijay Bhargava, Sherman Shen, Ross Murch, Khaled Ben Letaief, and Vahid Tarokh.

The Workshop featured twenty-two presentations, plus a graduate student session. The speakers and titles of the workshop papers are listed below:

Speaker	Title
Vincent Poor	Tight Finite-blocklength Bounds in Channel Coding
Vahid Tarokh	Capacity Bounds and Signaling Schemes for Bi-directional Coded Cooperation Protocols
Charlie Yang	An Information Theoretic Study of Stock Market Behavior
David Haccoun	Simple Iterative Decoding Using Convolutional Doubly Orthogonal Codes
Han Vinck	Capacity and Coding for Impulse Noise Channels
Dong In Kim	Near-optimal Receiver for Multiuser UWB Systems
Chengshan Xiao	Channel Equalization and Symbol Detection for Single Carrier Broadband MIMO Systems with Multiple Carrier Frequency Offsets
Poramate Tarasak	Cyclic / Phase Shifting for Active Interference Cancellation on Multiband OFDM UWB Transmission
Angel Bravo	Multireception Systems in Mobile Environments
Andreas Antoniou	On the Roots of Wireless Communications
Hlaing Minn	An Improved Ranging Method in OFDMA Systems
Ross Murch	MIMO Systems and Practical Antenna Considerations
Hamid Jafarkhani	Network Beamforming
Francois Gagnon	Multi-user Detection for Ad-hoc Networks in Emergency and Tactical Scenarios
Stephen Wicker	Advanced Sensing Systems and the Right to Privacy
Vikram Krishnamurthy	Global Games Approach for Decentralized Spectrum Access
Norman Beaulieu	Are Amplify-and-Forward Relaying and Decode-and-Forward Relaying the Same?
Ekram Hossain	Cognitive Radio MAC: Practical Issues, Potential Approaches and Open Problems
Fortunato Santucci	Reliability and Efficiency Analysis of Distributed Source Coding in Wireless Sensor Networks
Qiang Wang	A Generalized Decoding Problem
Anwarul Hasan	Detecting Errors in Elliptic Curve Scalar Multiplication Units
Hugues Mercier	Synchronization-Correcting Codes: Challenges and Applications

The session chairs were all former Ph.D. students of Vijay who are now professors. The symposium was held in the beautiful and historic Fairmont Empress Hotel in the center of Victoria. The highlight of the social program was a magnificant banquet on September 22 (Vijay's birthday), with Alexandre Henri-Bhargava serving as the Master of Ceremonies. Over 100 attendees were entertained with numerous stories about Vijay and his travels around the world.

Aaron Gulliver, Vijay Bhargava, Han Vinck, Vahid Tarokh, Ian Blake and Vincent Poor.

Marvin K. Simon, 1939-2008

Robert McEliece

Voulez-vous savoir le grande drame de ma vie? C'est que j'ai mis mon génie dans ma vie; je n'ai mis que mon talent dans mes œuvres. --- Oscar Wilde.

Marv Simon was a friend of, and an important contributor to, the Information Theory Society, but his primary intellectual home was the Communications Society. So when he died untimely at the age of 68 on September 23 last year, it was certainly appropriate that a full-page obituary appeared in the *Communications Magazine*,[1] rather than the *IT Newsletter*. But as ISIT 2008 recedes into history and Marv's *Yahrzeit* approaches, I'd like to share some of my memories of this remarkable man.

I'll skip his many professional achievements, which were in any event covered in the earlier obituary. Suffice it to say that Marv was one of the most prolific and influential communications researchers of his generation. Without ever having held a full-time university post, he mentored dozens of students and younger colleagues, including IT luminaries Dariush Divsalar and Slim Alouini. He was a true intellectual, and nothing delighted him more than the discovery of a new inequality or a simplified proof, which he would be more than happy to share.

But Marvin was much more than a successful scientist. His nimble mind led him into many unexpected places. For example, in the 1980's he became one of the world's experts on computer adventure games. During this time he wrote several best-selling books with the solutions to many popular games, among which "Keys to Solving Computer Adventure Games" and "Hints, Maps, and Solutions to Computer Adventure Games," are still available on amazon.com. He acquired this expertise during hundreds of hours spent playing computer games with his son Jeffrey. (He was an extraordinarily loyal father to Jeffrey, recently losing 50 pounds by becoming a guinea pig for his son's nutritional theories.) His daughter Brette was, and is, a successful corporate lawyer, which was a great source of naches for Marv. Marv would also have wanted everyone to know that near the end of his life, when his doctors had despaired, Brette refused to accept their pessimistic prognoses, and through her extraordinary intervention skills added at least an extra year to Marv's life.

The most memorable interactions I personally had with Marvin were musical. In January 2004 Marvin agreed to be the piano player for my performance of the song "Thank You Very Much" in the 2004 Shannon Lecture, and we began rehearsing weekly at his house. During these rehearsals, I learned that Marv led two lives: research engineer, which I of course knew about, and accomplished musician, which I didn't. It turned out that Marv had been studying music, mostly piano, since age 9, and he was good enough to make a living at it. And although he had had to make a binary career choice (choosing engineering over music) as a teenager, he had never stopped studying, playing, and performing.

As far as my own needs went, Marv was perfect. He was a gifted sight-reader, and a patient teacher. He owned a enormous collection of sheet music, including an arrangement of "Thank You Very Much" that I had never seen before and which I eventually chose for the Shannon Lecture. He knew everything about the composer Leslie Bricusse, including the correct pronunciation of his last name, and to top it all, he and his wife Anita had seen the musical "Scrooge" with the original cast in London.

And finally if I may be permitted a brief generalization, Marv Simon was soft-spoken but self-confident, a creative thinker who combed his hair, proud of this family but aware of their faults; and a fine host.

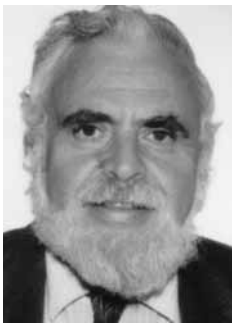
Atque in perpetuum, Marvin, ave atque vale.

[1] p. 22, November 2007 issue.

GOLOMB'S PUZZLE COLUMN™

Proofs by Dissections, Tiling, Etc.

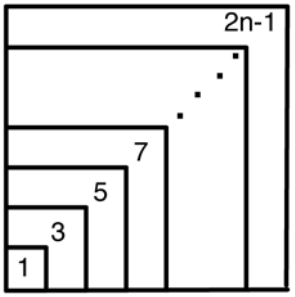
Solomon W. Golomb



Various geometric techniques can be used to give pictorial proofs of facts, formulas, and relationships. For example, the identity

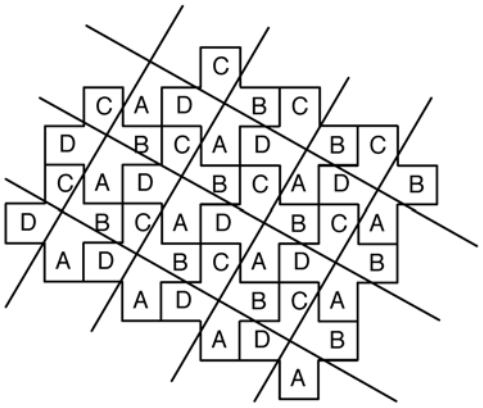
$$1+3+5+\cdots+(2n-1)=n^2$$

can be readily visualized from



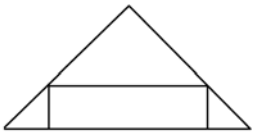
To cut the Greek cross (X-pentomino)  into four congruent pieces which can be reassembled to form a square, a tiling of the plane with these X's is superimposed on a tiling with squares of the same area.

The pieces A, B, C, D are congruent to each other, and can be assembled to form either a square of side $\sqrt{5}$, or an X composed of five squares of side 1.



Your task is to find geometric proofs of each of the following.

1. If a rectangle is inscribed on the base of an acute triangle, its area cannot exceed half the area of the triangle.



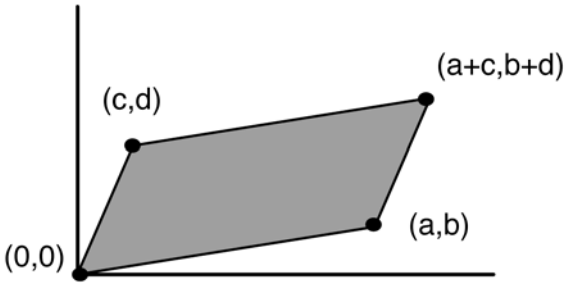
2. $1^3+2^3+3^3+\cdots+n^3=(1+2+3+\cdots+n)^2$.

3. For a right triangle with legs of length a and b, and a hypotenuse of length c, we know $c^2=a^2+b^2$. Prove this by superimposing a tiling of the plane with squares of side c on a periodic tiling using squares of side a and of side b. (Illustrate your tiling proof using $(a, b, c)=(3, 4, 5)$.)

4. The *partition function* $pt(n)$ counts the number of ways to write n as a sum of positive integers, without regard to order. (Since 4 can be written as: $4,\ 3+1,\ 2+2,\ 2+1+1,\ 1+1+1+1$, we have $pt(4)=5$.) We use $pt^k(n)$ to denote the number of partitions of n into at most k parts, and $pt_k(n)$ to denote the number of partitions of n into parts not exceeding k. (Thus $pt^2(4)=3$ since it counts $4, 3+1, 2+2$; and $pt_2(4)=3$ since it counts $2+2,\ 2+1+1,\ 1+1+1+1$.) Show geometrically that $pt^k(n)=pt_k(n)$, and illustrate your proof of the case $n=5,\ k=3$.

5. The determinant $ad-bc$ of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ equals (in absolute value) the area of the parallelogram whose vertices are $(0,0),\ (a,b),\ (c,d)$, and $(a+c,\ b+d)$.

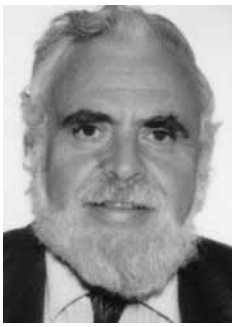
(This is the parallelogram based on the row vectors $\alpha=(a, b)$ and $\beta=(c, d)$.) Think of ad and bc as areas of rectangles, and give a geometric proof (by dissection) to show that the area of the parallelogram equals $|ad-bc|$.



GOLOMB'S PUZZLE COLUMN™

Some Problems About Primes Solutions

Solomon W. Golomb



1. "Every positive integer is in one and only one of the sequences $A_n = n + \pi(n)$ and $B_n = p_n + n - 1$ (with $n \geq 1$)" is an instance of what I call the "Sales Tax Theorem." We imagine a place where the sales tax starts at 0 and increases by 1¢ at each of the values p_n, with $n \geq 1$. Then the tax on an item listed at n¢ will be $\pi(n)$¢, for a *total price* of $n + \pi(n)$. When the list price (in cents) goes from $p_n - 1$ to p_n, the sales tax increases by 1¢, so the *total price* increases by 2¢. Thus, the *total prices* (in cents) that do not occur are precisely the numbers $p_n + \pi(p_n) - 1 = p_n + n - 1$.

2. The ratio $\frac{N}{\pi(N)}$ starts out, for $N > 1$, at $\frac{2}{\pi(2)} = 2$; and since $\lim_{x\to\infty} \frac{\pi(x)}{x} = 0$, we see that $\lim_{x\to\infty} \frac{x}{\pi(x)} = \infty$ That is, the ratio $\frac{N}{\pi(N)}$ becomes (and ultimately remains) arbitrarily large. Thus, to show that, for $N \geq 2$, $\frac{N}{\pi(N)}$ takes every integer value $k \geq 2$, we need only show that $\frac{N}{\pi(N)}$ cannot skip an integer when it is increasing, i.e., that $\frac{N}{\pi(N)} < k < \frac{N+1}{\pi(N+1)}$ cannot occur for integer $k \geq 2$. Now, either $\pi(N+1) = \pi(N)$ or $\pi(N+1) = \pi(N) + 1$. In the former case, from $\frac{N}{\pi(N)} < k < \frac{N+1}{\pi(N)}$ we get $N < k\pi(N) < N + 1$, putting the integer $k\pi(N)$ between two *consecutive* integers, an impossibility. But in the latter case, $\frac{N}{\pi(N)} < k < \frac{N+1}{\pi(N)+1} < \frac{N+1}{\pi(N)}$, again giving $N < k\pi(N) < N + 1$. (The first few integer values of $\frac{N}{\pi(N)} = k$ occur at least *three* times each. Does this pattern persist? Can such a result be proved?)

3. The largest N such that all $\phi(N) - 1$ values of k with $1 < k \leq N$ and *g.c.d.*$(k, N) = 1$ are prime is $N = 30$, where $\phi(30) - 1 = 7$, and these seven values of k, {7, 11, 13, 17, 19, 23, 29, are all the primes p with $5 < p < 30$.

4. The largest odd N such that all $\frac{1}{2}\phi(N) - 1$ odd values of k with $1 < k \leq N$ and *g.c.d.*$(k, N) = 1$ are prime is $N = 105$, where $\frac{1}{2}\phi(105) - 1 = 23$, and these 23 odd values of k, {11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103}, are all the primes p with $7 < p \leq 105$.

5. $Z(N) = N \prod_{p_i \leq y}(1 - \frac{1}{p_i})$, where $y = \sqrt{N}$, is an integer for exactly seventeen values of $N > 0$: {1, 2, 3, 4, 6, 8, 9, 12, 15, 18, 21, 24, 30, 45, 70, 105, 154. For $y = \sqrt{N} \geq 13$, and thus for $N \geq 169$, the denominator of $\prod_{p_i \leq y}(1 - \frac{1}{p_i})$ will contain prime factors that cannot be eliminated by multiplication times N.

6. For the same 17 values of N in the solution to Problem 5, and for no others, $\pi(N) = \pi(y) + Z(N) - 1$, where $y = \sqrt{N}$. It is a lucky coincidence that $N = 70$ and $N = 154$ are solutions, since these values of N are not multiples of *all* the odd primes p with $p \leq y = \sqrt{N}$.

References

[1] S. Golomb, "The Sales Tax Theorem," *Mathematics Magazine*, vol. 49, no. 4, pp. 187–189, Sep.–Oct., 1976.

[2] S. Golomb, "On the Ratio of N to $\pi(N)$," *American Math. Monthly*, vol. 69, no. 8, pp. 36–37, Jan. 1962.

Guest Column: News from the National Science Foundation

Sirin Tekinay

Dear reader,

I am sad to say, this is the last quarterly guest column I am writing in this series. It has been a privilege to fulfill part of my duty as your program officer by utilizing this space to write to you about relevant NSF programs and news, and facilitating our interaction on all that impact us as professionals in the communications community.

New and Upcoming Programs

Cyber-Enabled Discovery and Innovation, our NSF-wide multidisciplinary program, has completed its maiden voyage of 2008 with a fantastic group of projects, clearly poised to potentially produce transformative science. At the time of writing, the last of the thirty-six awards is being made by our grants division; therefore, by the time you read this, there will have been plenty of publicity about the results of CDI's first year.

We revised the CDI solicitation based on the lessons learned from year 1: http://www.nsf.gov/pubs/2008/nsf08604/nsf08604.htm . The solicitation is open, and of course, I am hoping for an even bigger participation by our community in unconventional intellectual partnerships that the program calls for.

News from the Communications Program

In July we received, as usual, an exciting batch of CAREER proposals. The CAREER panel for Communications and Information Foundations will be held, as scheduled, the first week in October. The review panel, which comprises experts who care deeply for our junior faculty and for the program, will have a hard time choosing between these terrific proposals.

As I proudly announced in the last installment, our Communications Program has become its own "Cluster;" i.e., we are no longer a "program element" under Theoretical Foundations, rather, a full blown program called "Communications and Information Foundations (CIF)" as part of the Computing and Communications Foundations Division, which will continue to serve the Communication, Information Theory, and Signal Processing communities. CIF is part of the coordinated CISE solicitation 2009, with its ongoing research foci, in addition to new frontiers in quantum information theory and foundations of secure communications.

Our community should also find that part of the directorate-wide cross-cutting program "Network Science and Engineering (NetSE)" is relevant to our interests in network information theory and coding, wireless networks, and sensor networks areas. There was a NetSE informational meeting held on September 5, where our community was well represented.

At the time of writing, the 2009 CISE solicitation which includes CIF, NetSE, and other areas of potential interest to our community, is open: http://www.nsf.gov/funding/pgm_list.jsp?org=CISE .

Due to changing rules and regulations of the NSF's Human Resource Management regarding rotators such as yours truly, it turned out I could not be offered a position here, despite being recommended for one. The worst part of this was that the representation of our community has been jeopardized. While I personally made plans to relocate to Istanbul, at the same time, I immediately applied for the volunteer status in order to make sure our program would enjoy seamless coverage. I have been broadcasting a message to our community, as widely as I can, suggesting that people get involved, communicate with NSF administration, and make sure our great community continues to have the champion it deserves at NSF. I'm glad to see NSF administration working hard to make up for their mistakes by working on the job announcement, and actively recruiting folks. I look forward to handing the torch over to another member of our community who will be our representative and advocate.

The "Social Scene"

The entire CDI team received the NSF Award for "Collaborative Integration" with a big luncheon on September 8. The award was announced on June 11; however, since the CDI team is so big and the CDI program is so special, the NSF Director gave this special party to celebrate its success. Ninety-nine plaques were given out: the team has grown, beyond the official working group of twenty four, comprising two representatives from each part of the foundation; we have an extended family of program officers who helped with the review process, acted as backup to the working group members; then we have the administrative staff, the unsung heroes who made it happen. It was a bittersweet farewell of sorts to the CDI chair, yours truly.

On a Personal Note

I have accepted the position of Vice Rector for Research and Technology at a new, privately funded research university in Istanbul, my hometown. The school is called Ozyegin University (http://ozyegin.edu.tr/eng/main/default.asp) and it boasts many firsts in Turkey, from student loans to having an executive academic administrator dedicated to research and technology. My second hat there will be the Director of School of Engineering. I am busy establishing the vision and putting together a portfolio of applied research. We already have many reverse-brain-drain hires we are proud of, and we are looking forward to more (http://ozyegin.edu.tr/eng/insanKaynaklari/isOlana klari.asp). The School of Business has started its classes with a very good bunch of students and the School of Engineering is gearing up to open its doors to students in September 2009.

As I leave my post at your service, I am getting back into our international research community, at a different level than I left. I look forward to our continued interaction in this new capacity.

… Always, dream big, and keep in touch!
Sirin Tekinay

Call for Nominations *continued from page 3*

IEEE Information Theory Society 2009 Aaron Wyner Distinguished Service Award

The IT Society Aaron D. Wyner Award honors individuals who have shown outstanding leadership in, and provided long standing exceptional service to, the Information Theory community. This award was formerly known as the IT Society Distinguished Service Award. Nominations for the Award can be submitted by anyone and are made by sending a letter of nomination to the President of the IT Society. The individual or individuals making the nomination have the primary responsibility for justifying why the nominee should receive this award.

NOMINATION PROCEDURE: Letters of nomination should

- Identify the nominee's areas of leadership and exceptional service, detailing the activities for which the nominee is believed to deserve this award;

- Include the nominee's current vita;

- Include two letters of endorsement.

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

Please send all nominations by March 15, 2008 to IT Society President, Andrea Goldsmith <andrea@systems.stanford.edu>.

IEEE Information Theory Society 2009 Paper Award

The Information Theory Society Paper Award is given annually for an outstanding publication in the fields of interest to the Society appearing anywhere during the preceding two calendar years. The purpose of this Award is to recognize exceptional publications in the field and to stimulate interest in and encourage contributions to fields of interest of the Society. The Award consists of a certificate and an honorarium of US$1,000 for a paper with a single author, or US$2,000 equally split among multiple authors. The award will be given for a paper published in the two preceding years.

NOMINATION PROCEDURE: By March 1, 2009, please email the name of the paper you wish to nominate, along with a supporting statement explaining its contributions, to the IT Transactions Editor-in-Chief, Ezio Biglieri, at <ezio.biglieri@gmail.com>.

IEEE Joint Comsoc/IT 2009 Paper Award

The Joint Communications Society/Information Theory Society Paper Award recognizes outstanding papers that lie at the intersection of communications and information theory. Any paper appearing in a ComSoc or IT Society publication during the years 2006-2008 is eligible for the 2009 award. A Committee with members from both societies will make the selection. The award consists of a plaque and cash prize presented at the Comsoc or IT symposium of the authors' chosing.

NOMINATION PROCEDURE: By March 1, 2009, please email the name of the paper you wish to nominate, along with a supporting statement explaining its contributions to both communications and information theory, to Frank Kschischang at <frank@comm.utoronto.ca>.

IEEE Fellow Program

For (s)he's a jolly good (IEEE) Fellow!

Do you have a friend or colleague who is a senior member of IEEE and is deserving of election to IEEE Fellow status? If so, consider submitting a nomination on his or her behalf to the IEEE Fellow Committee. The deadline for nominations is March 1st. IEEE Fellow status is granted to a person with an extraordinary record of accomplishments. The honor is conferred by the IEEE Board of Directors, and the total number of elected Fellows in any one year is limited to 0.1% of the IEEE voting membership. For further details on the nomination process please consult: http://www.ieee.org/web/membership/fellows/index.html

IEEE Awards

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

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

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

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

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

CALL FOR PAPERS AND FIRST ANNOUNCEMENT

Sixth International Workshop on Optimal Codes and Related Topics – OC 2009

Programme Committee	Stefan Dodunekov (Sofia), Marcus Greferath (Dublin), Tor Helleseth (Bergen), Ivan Landjev (Sofia), Juriaan Simonis (Delft), Leo Storme (Gent), Henk van Tilborg (Eindhoven), Wolfgang Willems (Magdeburg)
Organizing Committee	Silvia Boumova (Sofia), Tsonka Baicheva (V. Tarnovo), Peter Boyvalenkov (Sofia), Emil Kolev (Sofia), Ivan Landjev (Sofia), Nikolay Manev (Sofia)
Local Organizer	Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
Topics	• Optimal linear codes over finite fields and rings; • Bounds for codes; • Spherical codes and designs; • Covering problems for linear and nonlinear codes; • Optimization problems for nonlinear codes; • Sets of points in finite geometries; • Combinatorial configurations and codes; • Optimality problems in cryptography; • Graph theory and codes; • Related topics
Time	**June 16 – 22, 2009**
Location	The Workshop will take place in Varna (Bulgaria), hotel St. Elena (http://svetaelena.com/hotel.aspx?lang=en). Varna is the nearest airport. More information by `oc2009@math.bas.bg`.
Registration Fee (includes accommodation all inclusive, social events, workshop proceedings and materials)	EURO 550/600 (Double/Single room) prior to May 16, 2009; EURO 600/650 (Double/Single room) after May 16, 2009; EURO 400 for students (Double room); EURO 350 for spouses.
Deadlines	**March 31, 2009:** to inform the organizers if you intend to come; **April 15, 2009:** Deadline for submission of papers; **May 1, 2009:** Notification of acceptance (to be mailed out).
Language	The official language of the Workshop will be English.
Proceedings	The organizers intend to prepare a book of proceedings of the workshop. Authors are invited to submit at most six pages camera-ready papers in English, LaTeX format 132×190 mm, by e-mail to `oc2009@moi.math.bas.bg`.
Web site	`http://www.moi.math.bas.bg/oc2009/oc2009.html`.

Call for Papers

2009 IEEE International Symposium on Information Theory (ISIT2009)

COEX, Seoul, Korea / June 28-July 3, 2009 / http://www.isit2009.info

TPC Members

J. Andrews
A. Asikhmin
R. Baraniuk
A. Barg
J. C. Belfiore
C. Berrou
E. Biglieri
N. Cai
C. Carlet
M. Chiang
S. Diggavi
A. El Gamal
H. El Gamal
E. Erkip
C. Fragouli
T. Fujiwara
M. Gastpar
V. Goyal
A. Grant
B. Hajek
B. Hassibi
T. Helleseth
T. Ho
T. Javidi
N. Jindal
I. Kontoyiannis
V. Kumar
J. N. Laneman
T. Linder
H. A. Loeliger
H. Lu
G. Lugosi
S. Meyn
O. Milenkovic
U. Mitra
R. Nowak
D. Palomar
M. Parker
B. Prabhakar
B. Preneel
K. Ramchandran
R. Roth
S. Savari
A. Scaglione
G. Seroussi
S. Shamai
D. J. Shin
A. Shokrollahi
P. Siegel
E. Soljanin
H. Y. Song
R. Srikant
W. Szpankowski
E. Telatar
L. Tong
D. Tse
E. Tuncel
D. Tuninetti
S. Ulukus
R. Urbanke
P. Viswanath
P. Vontobel
T. Weissman
F. Willems
R. Yeung

The 2009 IEEE International Symposium on Information Theory will be held at COEX (Convention & Exhibition) in Seoul Korea, from Sunday June 28 through Friday July 3, 2009. Seoul, the capital of Korea for more than 600 years, boasts its unique, dynamic mixture of tradition and modernity, offering a wide spectrum of activities for travelers.
Previously unpublished contributions across a broad range of topics in information theory are solicited, including (but not limited to) the following areas:

- Channel and source coding
- Coding theory and practice
- Communication theory and systems
- Cryptography and security
- Data compression
- Detection and estimation
- Emerging applications of information theory
- Information theory and statistics
- Network and multi-user information theory
- Pattern recognition and learning
- Quantum information theory
- Sequences and complexity
- Signal processing

Submitted papers should be of sufficient detail for review by experts in the field. In addition to submitting new results in areas that form the core of information theory, researchers in related fields and researchers working on novel applications of information theory are encouraged to submit contributions. Final papers will be five pages in length. The submission deadline is January 7, 2009. Detailed information on paper submission, technical program, tutorials, travel, social programs, and travel grants will be announced on the ISIT 2009 web site: http://www.isit2009.info.

General Co-chairs:

Jong-Seon No (Seoul National University, Korea) jsno@snu.ac.kr
H. Vincent Poor (Princeton University, USA) poor@princeton.edu

TPC Co-chairs:

Robert Calderbank (Princeton University, USA)
Habong Chung (Hongik University, Korea)
Alon Orlitsky (UCSD, USA)

For general inquiries, please contact the General Co-chairs.

IEEE SOCIETY IEEE INFORMATION THEORY

IEEE

Conference Calendar

DATE	CONFERENCE	LOCATION	CONTACT/INFORMATION	DUE DATE
Feb 2-13 2009	**2009 Information Theory and Applications Workshop (ITA 2009)**	UCSD, San Diego, CA	http://ita.calit2.net/workshop.php	by invitation
March 18-20 1009	**2009 Conference on Information Sciences and Systems (CISS 2009)**	Johns Hopkins University, USA	http://ciss.jhu.edu/	January 5, 2009
April 19-25, 2009	**IEEE INFOCOM 2009**	Rio de Janeiro, Brazil	http://www.ieee-infocom.org/	August 29, 2008
May 10 - 15, 2009	**The 2009 International Workshop on Coding and Cryptography (WCC 2009)**	Ullensvang, Norway	http://wcc2009.org/	TBA
May 15-30, 2009	**XII International Symposium on problems of redundancy in information and control systems**	St. Petersburg, Russia	http://k36.0rg/redundancy2009/	March 1, 2009
June 10-12, 2009	**2009 IEEE Information Theory Workshop (ITW 2009)**	Volos, Greece	http://www.itw2009.org	Dec. 16, 2008
June 14-18, 2009	**IEEE International Conference of Communications (ICC 2009)**	Dresden, Germany	http://www.comsoc.org/confs/icc/2009/	Sept. 8, 2008
June 28 - July 2, 2009	**The 2009 IEEE International Symposium on Information Theory**	Seoul, Korea	http://www.isit2009.info	January 7, 2009
Oct. 11-16, 2009	**2009 IEEE Information Theory Workshop (ITW 2009)**	Taormina, Italy	http://www.deis.unical.it/itw2009	March 19, 2009

For other major ComSoc conferences: http://www.comsoc.org/confs/index.html