IEEE Information Theory Society Newsletter

Vol. 65, No. 4, December 2015 Editor: Michael Langberg ISSN 1059-2362

Editorial committee: Frank Kschischang, Giuseppe Caire, Meir Feder, Tracey Ho, Joerg Kliewer, Anand Sarwate, Andy Singer, and Sergio Verdú

President's Column

Michelle Effros

Last January, when I sat down to write my first column as President of the Information Theory Society, I asked for your help on a monumental task. In that column, I set forth the goal of increasing public awareness about information theory, argued for the importance of this mission, and noted that we are the right ones—perhaps the only ones—to make that goal a reality. I suggested that we teach children about Shannon so that they can be inspired by his example and understand that math is a powerful tool with which they too can transform the world. I proposed that we communicate our results beyond our Transactions and our own community's events so that researchers from other fields can learn enough about what we do to discover connections with their own interests. I advocated that we talk about our ideas in the popular press so that the public can understand the impact of our past accomplishments and the potential of our future work.

The goals that I set forth are long term endeavors. I write today, in mid-October, to report on how far we have come in the intervening months and to look forward to where we might go in the future.

The ad-hoc committee for Broader Outreach, led by Christina Fragouli, manned by Ruediger Urbanke, Lav Varshney, Sergio Verdú, and myself, and supported by a wide base of Society members, has been hard at work at making the goal of broad outreach a reality. Since Shannon's 100th birthday presents a rare and important opportunity to draw attention to the field and the people who make it happen, the committee's focus to date is on events and initiatives to mark Shannon's centennial.

We are moving forward on the Shannon Documentary, which we hope to have completed and signed for broad distribution by the end of 2016. The filmmaker Mark Levinson, working in consultation with Sergio Verdú, has written a first draft of a "treatment"—a sort of outline used in lieu of a script when making a documentary. The process of reviewing and revising that treatment with input from the committee and other experts is currently underway.

Our fund raising efforts for the documentary are in full swing. To date, we have raised over 2/3 of the projected budget. We would love to receive the support of individuals and institutions to bridge the gap and make this once-in-a-century opportunity a reality. **Donations for the project can be made at the IEEE Foundation's project webpage** https://ieeefoundation.org/ClaudeShannon. We look forward to publicly thanking the generous donors who are making this once-in-a-hundred-year opportunity a reality. (Anonymous donations are also possible. Please note the desire to maintain anonymity with your donation if that is your preference.)

The committee is working to seed and support Shannon Day events at institutions around the world. A variety of materials are being created and shared to amplify the efforts of all of our volunteers and make hosting a Shannon Day event at your institution as easy as possible. A Shannon Day logo, crowdsourced for these events, now graces our website and this Newsletter. Using this logo on publicity materials for events around the world will help unify these activities. Posters describing both Shannon's life and key technical problems in information theory are under preparation by a team of volunteers. Erdal Arikan, German Bassi, Vijay Gupta, Al Hero, Nicolas Macris, Anna Scaglione, Lalitha Sankar, Aslan Tchamkerten, Lav Varshney, and Aylin Yener are leading the efforts on posters with topics ranging from Shannon's biography to source coding, channel coding, and quantum

continued on page 39

From the Editor

Michael Langberg

Dear colleagues,

In our fourth issue for 2015 we open with Michelle Effos's last column as President of the IT Society. Please join me in thanking Michelle for her dedication and inspiring leadership over the past year, and in warmly welcoming our incoming President Alon Orlitsky. The upcoming year of 2016 is one of festivity for our community. I am eagerly looking forward to reporting on the workshops, seminars, and events now in preparation for The Shannon Centenary, events that will surely touch our society as well as reach out and influence societies beyond our own.

The current issue opens with a number of excellent technical contributions. We start with an outstanding contribution by this years Shannon Award winner Robert Calderbank summarizing his Shannon Lecture "The Art of Signaling", presented in ISIT over the summer, which beautifully links between binary and Euclidean geometry through the lens of Fourier analysis. We follow by a contribution from Alexander Barg and Itzhak Tamo, this years recipients of the IT Society Paper Award for the paper "A Family of Optimal Locally Recoverable Codes," which summarizes their elegant algebraic constructions alongside several extensions, connections, and applications. We conclude with two excellent surveys summarizing tutorial sessions presented in ISIT this summer. Mohammad Ali Maddah-Ali and Urs Niesen present the intriguing possibilities and challenges in the context of Cache Networks in their survey "Cache Networks: An Information-Theoretic View". Maxim Raginsky and Igal Sason present a comprehensive survey on concentration inequalities with a list of information theoretic applications in their article "Concentration of Measure Inequalities and Their Communication and Information-Theoretic Applications". I greatly thank the authors, on behalf of the newsletter editorial board, for their significant efforts in preparing these excellent contributions, which may expose us all to new and exciting fields of study.

The body of this issue also includes our recurring contributions alongside reports and announcements. Many thanks to Tony Ephremides for preparing the Historian's Column and to Sol Golomb for preparing his Puzzle Column. In continuation of our efforts to reach out to the students of our society, the "Students' Corner" is an attempt to bring forward contributions "by students-for students" allowing students to share their experiences and perspective on our community. Thanks to Jonathan Scarlett who generously wrote this issue's column. Thanks to Yuval Kochman, the chair of the IEEE Israel Section Chapter, for writing this issue's column "From the field" which regularly includes reports on exciting local events and initiatives form our chapters worldwide.

continued on page 23

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

The Art of Signaling

Robert Calderbank
Duke University

I had listened to many Shannon lectures without appreciating how hard it is to tell a technical story that is personal and broadly accessible. I know that I rarely follow a technical talk from start to finish, and so I wanted to tell a sequence of technical stories with a common thread but different enough to provide the audience with opportunities to reconnect. I was mindful that the audience would be quite diverse, but I thought all would be familiar with Fourier analysis, and that I could build out from Fourier analysis in the binary world.

1. Fourier Analysis in the Binary World

Philippe Delsarte is an extraordinary Belgian electrical engineer, he was employed by Philips, and his thesis [1], published in 1973, transformed the field of coding theory leading to the famous linear programming bounds on achievable rate.

If you are familiar with the duality between time and frequency in the continuous world, then you will know that sinusoids are eigenfunctions of time shifts. Delsarte's thesis opened the door to a parallel universe, where there are binary counterparts of time and frequency shifts that are interchanged by the Walsh-Hadamard transform, a matrix that plays the same role as the Fourier transform in classical analysis. The binary counterparts of sinusoids are called Walsh functions, known to coding theorists as the codewords in the 1st order Reed Muller code.

Set $N = 2^m$. The Heisenberg-Weyl group HW_N is a remarkable group of $N \times N$ matrices that provides the framework for Fourier analysis in the binary world. When $m = 1$, the group H_2 is just the symmetry group of the square, generated by the matrices $x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. The Walsh-Hadamard matrix $H_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} + & + \\ + & - \end{pmatrix}$ where +, – represent 1, –1.

We view the symmetry x as a binary time shift; indexing rows and columns by 0 and 1, this matrix represents adding 1 modulo 2. The lines B_1 and B_2 are eigenvectors of the time shift x, and we view them as binary sinusoids. The symmetry z then becomes a binary frequency shift. The lines A_1 and A_2 are eigenvectors of z, and we view them as Dirac spikes. The Walsh-Hadamard matrix H_2 acts like the Fourier transform in classical Fourier analysis, interchanging time and frequency in our binary world. Applying H_2 to vectors interchanges coordinate frames, and conjugating by H_2 interchanges x and z.

The matrices x and z are also known as Pauli matrices in quantum mechanics. When we discuss quantum error correction, we will need the concept of a quantum bit; this is a 2-dimensional Hilbert space and we will call the two basis states 0 and 1. In a quantum computer, these two basis elements will have a physical realization, perhaps as two states of a beryllium ion. The matrix x interchanges 0 and 1, and the matrix z changes their relative phase, hence we refer to x as a bit flip, and we refer to z as a phase flip. A quantum computer needs more than a single qubit, and we use m-fold Kronecker products to describe the evolution of an m-qubit quantum system.

Given a binary m-tuple $a = (a_0, \ldots, a_{m-1})$, we set $D(a, 0) = x^{a_{m-1}} \otimes \cdots \otimes x^{a_0}$. Thus

$$D(10,00) = \begin{pmatrix} + & \\ & + \end{pmatrix} \otimes \begin{pmatrix} & + \\ + & \end{pmatrix} = \begin{bmatrix} & + & & \\ + & & & \\ & & & + \\ & & + & \end{bmatrix} \begin{matrix} 00 \\ 10 \\ 01 \\ 11 \end{matrix}$$

The matrix $D(a, 0)$ is a permutation matrix that represents addition of the binary vector a modulo 2. We think of these matrices as binary time shifts.

Similarly, given a binary m-tuple $b = (b_0, \ldots, b_{m-1})$ we set $D(0, b) = z^{b_{m-1}} \otimes \cdots \otimes z^{b_0}$. Thus

$$D(00,10) = \begin{pmatrix} + & \\ & + \end{pmatrix} \otimes \begin{pmatrix} + & \\ & - \end{pmatrix} = \begin{bmatrix} + & & & \\ & - & & \\ & & + & \\ & & & - \end{bmatrix} \begin{matrix} 00 \\ 10 \\ 01 \\ 11 \end{matrix}$$

We think of these matrices as binary frequency shifts. Each frequency shift $D(0, b)$ is a diagonal matrix, and the diagonal entry

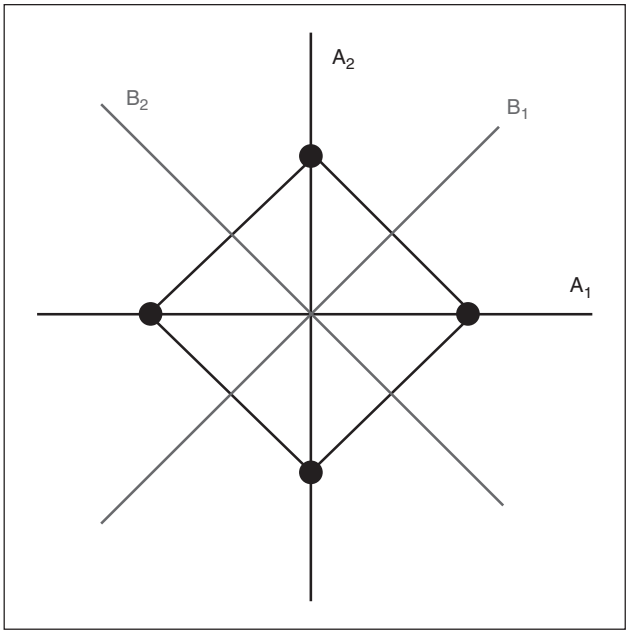


Figure 1. The square has 4 axes of symmetry: the lines A_1 and A_2 joining opposite vertices, and the lines B_1 and B_2 joining the midpoints of opposite sides. Any symmetry of the square permutes these four axes. The coordinate frame B_1, B_2 is fixed by x and the coordinate frame A_1, A_2 is fixed by z.

indexed by v is either 1 or -1 according as the binary inner product of b and v is 0 or 1.

The Walsh-Hadamard matrix H_N is the m-fold Kronecker product of the 2×2 matrix H_2. Conjugation by H_N interchanges the time shift $D(a, 0)$ and the frequency shift $D(0, a)$.

The elements of the Heisenberg-Weyl group HW_N are multiples of the Kronecker products $D(a, b) = D(a, 0)D(0, b)$ by the phases 1, -1, i, and $-i$. The notation $D(a, b)$ keeps track of how the Kronecker product is formed; the binary vector a captures appearances of x and the binary vector b captures appearances of z.

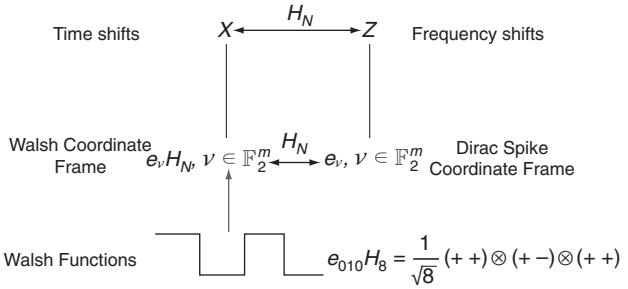
$$D(11010, 10110) = I_2 \otimes xz \otimes z \otimes x \otimes xz$$

There are $4N^2$ matrices, all of them square to I or $-I$, any pair of matrices commute or anticommute and a symplectic form governs which possibility occurs:

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

As before, the Walsh-Hadamard matrix H_N acts like the Fourier transform in classical Fourier analysis, interchanging time and frequency in our binary world. Applying H_N to vectors interchanges the Walsh and Dirac coordinate frames, and conjugating by H_N interchanges time shifts and frequency shifts.

The Heisenberg-Weyl framework makes it possible to construct a large number of coordinate frames, and to control the correlation between different frames. Every unitary matrix U that fixes HW_N by conjugation



maps the time shift group to a new commutative subgroup, and maps the Walsh vector coordinate frame to a new coordinate frame. What is special about HW_N is that there are many, many choices for the unitary matrix U.

2. Families of Quaternary Sequences with Low Correlation

Binary sequences generated by shift registers appear in spread spectrum communication and other applications. This subject has a long history, much of it associated with USC, starting with Sol Golomb and Lloyd Welch, and continuing with Vijay Kumar. This Section describes (in a new way) my 1994 paper with Vijay, his student Roger Hammons Jr., Neil Sloane, and Patrick Sole [2] on a code design framework that led to the short uplink scrambling codes in the 3G WCDMA standard.

Most coding theory emphasizes linear codes. They are simple to understand, to encode and decode, but most of all, they are simple to discover. However, the most efficient codes are sometimes nonlinear. For example the Nordstrom-Robinson and Preparata codes are twice as large as the best possible linear codes for the same parameters.

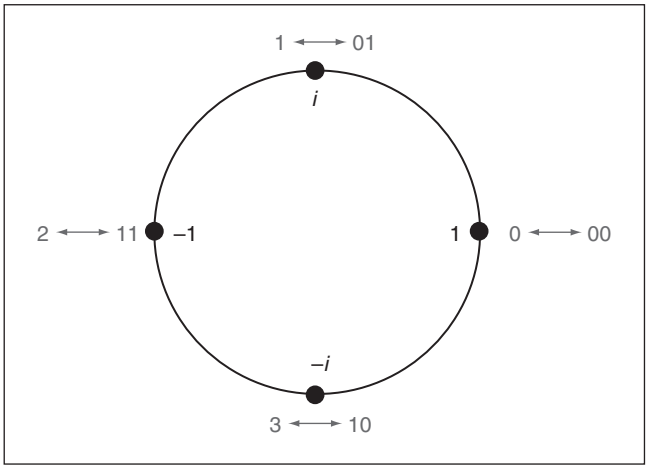


Figure 2. The quaternary world, where Lee distance is measured around the circle, and the binary world, where distance is Hamming distance. We give each QPSK phase a quaternary label and a binary label, and we observe that the two notions of distance coincide. The Gray map is an isometry.

When a code is linear the MacWilliams transform of the distance distribution yields the distance distribution of the dual. Kerdock and Preparata codes were known to be dual in this sense, even though as binary codes they were nonlinear. Why this was so was a mystery and it was resolved by showing that Kerdock and Preparata codes are in fact linear, if one views them in the right way over the ring of integers modulo 4, instead of the binary field. Over this larger ring the two codes ARE mathematical duals.

The Kerdock and Preparata codes turn out to be images under the Gray map of a quaternary linear code and its dual. Actually this is not strictly accurate. The Kerdock code is the one discovered by Kerdock, but Preparata discovered a slightly different code with identical distance properties. The MacWilliams identities also relate the weight enumerator of a quaternary code to that of its dual. Since Lee and Hamming weights coincide, the mystery is resolved.

We can think of the entries of a Kerdock codeword as QPSK phases. The Kerdock code is a union of coordinate frames, where the coherence between a basis vector in one frame and a basis vector in a different frame is as small as it can possibly be. Thus the Kerdock code is a union of mutually unbiased bases. If you ask for a maximal collection of unit vectors of length N with the property that any two vectors are orthogonal or have coherence $N^{-1/2}$, the answer is a Kerdock code.

We construct coordinate frames by constructing maximal commutative subgroups of the Heisenberg-Weyl group. The time-shift group determines the frame of Walsh functions. If P is an $m \times m$ symmetric matrix, then the unitary matrix $d_P = \text{diag}[i^{xPx^{\top}}]$ conjugates the Heisenberg-Weyl group to itself, and conjugates the time-shift group X to the maximal commutative subgroup X_P. The coordinate frame determined by X_P is formed by the rows

of the matrix $d_P H$, and if you are familiar with classical coding theory, you can start to see second order Reed Muller codes emerging.

It turns out that the coherence between coordinate frames determined by X_P and X_Q depends only on the rank of the binary matrix $P + Q$. In particular, if $P + Q$ is nonsingular, then these coordinate frames are mutually unbiased bases. It is possible to construct a binary vector space of N Hankel matrices of size m with the property that every non-zero matrix is non-singular. The vector space is called a Kerdock Set and the corresponding Kerdock code is just the set of coordinate frames determined by this set.

3. Space-Time Codes for Wireless Communication

This section focuses on three papers that I wrote with Vahid Tarokh [3, 4, 5], but he is representing many friends and collaborators - Nambi Seshadri, Suhas Diggavi, Naofal Al Dhahir, Ayman Naguib, and Hamid Jafarkhani in particular. AT&T Labs was an extraordinarily collaborative environment, one that I remember with great affection.

Why did we consider transmit diversity when, given a choice, information theorists know that it is better to have multiple antennas at the mobile? We were motivated by the fact that there are many more mobiles than there are base stations, making it easier to get innovation into base stations than into mobiles. We focused on small numbers of transmit antennas for reasons that have little to do with radio. The more antennas in a base station, the more it resembled a sail, so wind-qualification of the tower was a concern. Also, signal processing was taking place, not at the antenna, but in a hut at the base of the tower, and it was difficult to have too many thick wires going up the middle of the tower.

The simplest form of transmit diversity is the delay diversity scheme proposed by Wittneben for two transmit antennas. A signal is transmitted from the second antenna, then delayed one time slot and transmitted from the first antenna. It achieves a diversity gain but no coding gain, whereas our 1998 paper achieves both. Before we started thinking about transmit diversity, Nambi Seshadri had designed block codes for fading channels, including the following code for 8-PSK:

$$C = \{00, 15, 22, 37, 44, 51, 66, 73\}$$

It turns out that our first space-time trellis code results from applying delay diversity to this block code.

The most famous space-time code is the 2×2 block code discovered by Siavash Alamouti [6], where the columns represent different time slots, the rows represent different antennas, and the entries are the symbols to be transmitted. The encoding rule is

$$(c_1, c_2) = \begin{pmatrix} c_1 & c_2 \\ -\bar{c}_2 & \bar{c}_1 \end{pmatrix}$$

The signals r_1, r_2 received over two consecutive time slots are given by

$$\begin{pmatrix} r_1 \\ -\bar{r}_2 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 \\ -\bar{h}_2 & \bar{h}_1 \end{pmatrix} \begin{pmatrix} c_1 \\ -\bar{c}_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ -\bar{w}_2 \end{pmatrix}$$

Thus $r = Hc + w$ where the matrix H of channel gains that connects received to transmitted signals is a scalar multiple of a unitary matrix. When we apply H^* we obtain a scalar multiple of the transmitted codeword in Gaussian noise, and decoding is extremely simple.

Siavash Alamouti never thought about quaternions. William Hamilton, who discovered quaternions in 1843, certainly never thought about cellular phones. However he knew that quaternions were the answer to every problem presented by the physical world. a point of view that became known to some as the Irish madness. And perhaps he was right after all, because his representation of quaternions as pairs of complex numbers coincides with the Alamouti block space time code.

Quaternions lead to orthogonal designs, and to normed algebras and sums of squares. In fact the problem of finding full rate orthogonal designs is essentially that of finding all normed algebras, a question resolved by Hurwitz in 1896. Where can we find orthogonal designs? If we start with the quaternion orthogonal design we discover that the matrix that captures appearances of each variable is an element of the Heisenberg-Weyl group H_4.

$$\left[\begin{array}{cc|cc} x_0 & x_1 & x_2 & x_3 \\ -x_1 & x_0 & -x_3 & x_2 \\ \hline -x_2 & x_3 & x_0 & x_1 \\ -x_3 & -x_2 & x_1 & x_0 \end{array}\right] = x_0 I_4 - x_1 D(10, 11) - x_2 D(01, 01) - x_3 D(11, 10).$$

When we ask what properties these matrices must have, we discover that they must pairwise anticommute, and after a change of basis we are led to what is called a Hurwitz-Radon family of matrices. For more details, see my paper with Ayman Naguib [7] where we also describe interference cancellation of two Alamouti coded streams using a second antenna at the receiver. We used this idea in a project linking AT&T and Nokia where the aim was to achieve 1 Mb/s on GSM channels, and where Nokia was most insistent that we not exceed the capabilities of the DSP used in their 2nd generation systems. The Heisenberg-Weyl group structure makes it all possible. Texas Instruments liked this idea and pushed it in CDMA standards.

4. Quantum Error Correction

The focus of this final section is quantum computing. If we look back at 1946, a computer was essentially a physics experiment, and the same is true of quantum computers today. Peter Shor showed that if a quantum computer could be built then it would be possible to compromise public key cryptography by factoring integers exponentially faster than the best known classical computer. His discovery led to an explosion of interest in the field.

What makes quantum computing so challenging is decoherence; the environment interacts with the computer, introducing errors into a computation as it is happening. Quantum error correcting codes insulate computations from decoherence, and I am now going to describe the mathematical framework for quantum error correction that I developed with Peter and other colleagues [8, 9].

Classical bits can only take values 0 and 1, but a quantum bit is a two dimensional Hilbert space and can find itself in an arbitrary superposition of the two basis states e_0 and e_1. A quantum

computer employs m qubits and we use Kronecker products to describe all the basis states. When the quantum computer occupies a superposition of basis states, it is able to investigate in a uniform way exponentially many potential instances at the same time. That is the power of quantum computing.

We say that the environment is continually measuring the quantum system. What does that mean? Von Neumann formulated measurement in terms of a resolution of the identity, a collection of projection operators P_i that are pairwise orthogonal, and that sum to the identity operator. When we measure a state v, we project onto one of the subspaces, and learn the index i. The energy of the unit vector v is distributed across the different subspaces, and the probability of landing in subspace i is simply the fraction of energy in that subspace.

The error group of an m-qubit quantum system is the Heisenberg-Weyl group H_N, the matrix x represents a bit flip, and the matrix z a phase flip. This is the quantum analog of the classical binary symmetric channel and we can think of the probability of an error as Bernoulli p in the number of qubits that it touches. Note that we work with ixz rather than xz because we need Hermitian operators for measurement.

Given a commutative subgroup of H_N, the projection operators associated with the common eigenspaces form a resolution of the identity. More precisely, if we are given $e = (e_1, \ldots, e_k)$ with $e_j = 1$ or -1, and a commutative group of order $4 \cdot 2^k$ generated by iI_N and matrices $D(a_j, b_j)$ then the operators

$$P_e = \frac{1}{2^k} \prod_{j=1}^{k} \left(I_N + e_j i^{a_j b_j^\top} D(a_j, b_j) \right),$$

constitute a resolution of the identity. Note that the subspaces associated with the projection operators are defined by the property that every matrix in the commutative subgroup acts as I or $-I$ on each subspace. This means that the subspaces are independent of the choice of generators. Conjugation by an element g of the error group fixes the commutative subgroup, so g just permutes the projection matrices P_e.

Each subspace in the resolution of the identity has dimension 2^{N-k} and we can think of it as $N - k$ qubits. For example, the matrices $D(a, b)$ generated by the rows $(a \mid b)$ of the matrix

$$G = \left[\begin{array}{ccccc|ccccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{array}\right]$$

generate a commutative group of 32×32 matrices with order $4 \times 16 = 64$. There are 16 subspaces in the resolution of the identity and each is 2-dimensional.

We pick one of the subspaces, say P_{0000}, we think of it as a single qubit, and we now show that the construction protects against any single qubit error. There are 15 remaining subspaces in the resolution of the identity, and 15 possible single qubit errors. It is not difficult to show that different single qubit errors take us to different eigenspaces. We identify the eigenspace and simply reverse the most probable way of getting there – just like syndrome decoding in the classical world.

The codes constructed from commutative subgroups in this way are called stabilizer codes or CSS codes (for Calderbank, Shor and Steane).

5. Conclusion

I feel like I spent most of last year either procrastinating and not writing my Shannon lecture or worrying about the impression I would leave. Now that it is over I am happy with the story I chose, and I would like to thank all my friends and colleagues who made it possible.

References

[1] P. Delsarte, An algebraic approach to the association schemes of coding theory, Philips Research Reports Supplements, No. 10, 1973.

[2] A.R. Hammons Jr., P.V. Kumar, A.R. Calderbank, N.J.A. Sloane and P. Sole, The Z4-linearity of Kerdock, Preparata, Goethals and related codes, IEEE Transactions on Information Theory, Vol. 40 (2), pp. 301–319, March 1994.

[3] V. Tarokh, N. Seshadri, and A.R. Calderbank, Space-time codes for high data rate wireless communication: Performance criterion and code construction, IEEE Transactions on Information Theory, Vol. 44 (2), pp. 744–765, March 1998.

[4] A.F. Naguib, V. Tarokh, N. Seshadri and A.R. Calderbank, A space-time coding modem for high datarate wireless communications, IEEE Journal on Selected Areas in Communications, Vol. 16 (8), pp. 1459–1478, October 1998.

[5] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, Space-time block codes from orthogonal designs, IEEE Transactions on Information Theory, Vol. 45 (5), pp. 1456–1467, July 1999.

[6] S. Alamouti, A simple transmit diversity technique for wireless communications, IEEE Journal on Selected Areas in Communications, Vol. 16 (8), pp. 1451–1458, October 1998.

[7] A.R. Calderbank and A.F. Naguib, Orthogonal designs and third generation wireless communication, Surveys in Combinatorics, edited by J.W.P. Hirschfeld, London Mathematical Society Lecture Notes 288, pp. 75–107, 2001.

[8] A.R. Calderbank and P.W. Shor, Good quantum error correcting codes exist, Physical Review A, Vol. 54 (2), pp. 1098–1105, 1996.

[9] A.R. Calderbank, E.M. Rains, P.W. Shor and N.J.A. Sloane, Quantum error correction and orthogonal geometry, Physical Review Letters, Vol. 78. Pp. 405–409, 1997.

On Codes with the Locality Property

Alexander Barg and Itzhak Tamo

***Abstract*—We consider codes that support the local recovery property of each code symbol (LRC codes). Codes of this type were extensively studied in recent years because of their applications in distributed storage systems. We discuss algebraic constructions of LRC codes over small alphabets that attain the best possible distance-locality tradeoff and their extensions to cyclic codes and codes on algebraic curves. We also discuss examples of practical LRC codes used in large-scale storage systems and point out some open questions in this area.**

I. Introduction

Distributed and cloud storage systems have reached such a massive scale that recovery from several node failures is now part of regular operation of the system rather than a rare exception. To support reliable storage, system designers have turned to error correcting codes, introducing redundancy to recover the temporarily or permanently unavailable data. The simplest and to date the most frequently employed solution is to replicate the data several times, writing the copies of each data fragment to distinct physical locations. For example, *Apache Hadoop*, an open source software for distributed storage, uses a default method of 3-way replication. Another common solution, based on Reed-Solomon (RS) codes, provides stronger protection for the same or smaller storage overhead. For instance, the file systems of Facebook and Google use the (14,10) and (9,6) RS codes, respectively. RS codes have been also standardized as a part of the well-known RAID 6 data protection technology.

New challenges in the development of distributed storage systems are to a large extent driven by the exponential growth of the amount of stored data which makes exabyte data volumes today's new reality. One of the new tasks faced by such systems, but not addressed by current solutions, is recovery from a single node failure. Studies show that, although several concurrent failures are possible, and therefore the system should be able to protect against them, the most common scenario is the failure of a single node. Therefore, constructing codes that optimize the repair of a single node becomes an important problem for coding theorists and developers alike.

Recovery of the information stored on a single node, or the *repair problem*, can be carried out successfully because of the redundancy inserted in the information at the time of writing to the memory. The efficiency of the data repair can be measured in several ways. One of them, introduced in the foundational paper [4], proposes to optimize the amount of data transmitted in the system to accomplish the repair. This metric has become known as *repair bandwidth*. The second measure, called *locality*, is related to the total number of nodes accessed during the data recovery [8], [9], [7]. Both metrics have their own merits, and choosing between them is related to the type of the storage system and the underlying scope of applications. In this paper we focus on codes with locality, i.e., codes that in the course of repair of a single node access only a small number of other nodes.

An (n,k,r) locally recoverable (LRC) code encodes k data symbols into n symbols in such a way that the value of any symbol of the encoding can be found by accessing at most r other stored symbols. For example, a code of length $n = 2k$ in which every data symbol is repeated twice, is an LRC code with locality $r = 1$. As another extreme, consider an (n,k) MDS code with locality $r = k$ in which not only one symbol, but the entire encoding can be found by accessing k codeword symbols. Generally the value of locality r satisfies $1 \le r \le k$. Yet another simple example is provided by regular LDPC codes with $r+1$ nonzeros in every check equation, meaning that every single symbol of the codeword is a linear combination of some other r symbols. The study of LRC codes forms a new topic in coding theory that gives rise to questions ranging from limits to the maximum size of LRC codes to the constructions and structure of codes and their decoding algorithms. For instance, MDS codes which are optimal for the classical error/erasure correction problem, are far from being optimal in terms of locality because the repair task requires access to a large number of code symbols.

Bounds and constructions of LRC codes have been studied in a number of recent papers. A natural question to ask is as follows: given an (n,k,r) LRC code $\mathcal{C}$, what is the largest possible minimum distance $d(\mathcal{C})$? A useful generalization of the Singleton bound [7], discussed in Section II-A, Eq. (3) below, gave rise to both studies into code bounds and constructions of RS-type codes that form the main topic of this paper. While the LRC Singleton bound, like its classic counterpart, is independent of the code alphabet, another work [3] introduced a bound on the code's distance that accounts for the alphabet size, and more results of this kind appear in the recent paper [17].

Codes whose parameters satisfy the LRC Singleton bound with equality, are called *optimal LRC codes* in the literature. Among the constructions of LRC codes we note the results of [15], [19], [6] that combine some known code families to account for the LRC property. While these constructions are optimal by their parameters, they rely on alphabets of a large size, limiting their usefulness in applications.

Coding for distributed storage is currently an active research area. Codes that optimize the repair bandwidth and codes with locality appear in a large number of publications, too numerous to cite or overview in this article. In [16] we initiated a line of research in this area that begins with a construction of RS-type codes with the locality property and extends to constructions of cyclic codes and codes on algebraic curves, as well as to a study into bounds on the parameters of LRC codes. In this paper we present and discuss this work, apologizing to our

many colleagues whose contributions to this area we did not have a chance to mention.

II. Codes with the locality constraint (LRC codes)

We say that a code has locality r if the value of every coordinate of the codeword c is uniquely determined by the values of at most r other coordinates of c. In the context of storage applications, this enables the system to recover the data from a dysfunctional node by accessing at most r other nodes in the storage cluster. At the same time, if a group of more than one nodes become inaccessible, we would still like to be able to restore the data using the remaining storage nodes. In this case, it may not be possible to recover the missing symbols in a local fashion, but we would like to be able to recover them nevertheless by accessing the remaining available symbols of the codeword. Taken together, these conditions call for constructing codes with small locality and large distance d. A formal definition of an LRC code is as follows.

Definition 1: A code $\mathcal{C}$ of length n over a finite alphabet $\mathcal{Q}$ is said to have *locality* r if for every $i \in [n]$ there exists a subset $\mathcal{R}_i \subset [n]\backslash\{i\}, |\mathcal{R}_i| \leq r$ and a function ϕ_i such that for every codeword $c \in \mathcal{C}$

$$c_i = \phi_i(\{c_j, j \in \mathcal{R}_i\}). \tag{1}$$

As already remarked, simple examples of LRC codes are obtained by concatenating several copies of some code. For instance, replicating m times a single-parity-check code of length $r+1$, we obtain an $(m(r+1), mr, r)$ LRC code with distance $d = 2$, and repeating twice an $(n/2, k)$ MDS code yields an $(n, k, 1)$ LRC code with distance $d = 2(n/2 - k + 1)$.

Let us give a less trivial example of an LRC code. This example relies on the main LRC code construction discussed in the paper.

Example 1: We will construct an $(n = 9, k = 4, r = 2)$ LRC code $\mathcal{C}$ with distance $d = 5$, choosing $\mathbb{F}_{13}$ to be the code alphabet. Consider the space of polynomials

$$\mathcal{P} = \{f_a(x) = a_0 + a_1 x + a_3 x^3 + a_4 x^4\},$$

where $a = (a_0, a_1, a_3, a_4) \in \mathbb{F}_{13}^4$ denotes the message vector (the omission of x^2 is intended). Consider the linear code

$$\mathcal{C} = \{ev_A(f), f \in \mathcal{P}\},$$

defined by the set of points $A = \{1, 3, 9, 2, 6, 5, 4, 12, 10\}$ and the evaluation map $ev_A : \mathbb{F}_q[x] \to \mathbb{F}_q^n$ given as $ev_A(f) = (f(a), a \in A)$. For instance, taking $a = (1, 1, 1, 1)$, we evaluate the polynomial $f_a(x) = 1 + x + x^3 + x^4$ to find the codeword

$$c := ev_A(f_a) = (4, 8, 7, 1, 11, 2, 0, 0, 0). \tag{2}$$

Since the degree of $f_a(x)$ is at most 4, the distance of the code satisfies $d(\mathcal{C}) \geq 5$. It will be argued later that 5 is the maximum possible distance for any $(9, 4, 2)$ LRC code, so the code $\mathcal{C}$ is optimal. Note that an RS code with $n = 9$ and $k = 4$ has distance 6 which is only one greater than the distance of the code $\mathcal{C}$. Therefore by reducing the distance by one we managed to decrease the locality by a factor of two.

Although the code $\mathcal{C}$ is a subset of a $(n = 9, k = 5)$ RS code[1], we emphasize the special choice of the space $\mathcal{P}$ and the set A which account for the locality property of the code. Indeed, regardless of the exact values of the entries of the information vector a, there is a linear polynomial that passes through the points $f_a(1), f_a(3)$ and $f_a(9)$ of the codeword c. For instance, the polynomial $\delta_1(x) = a_0 + a_3 + (a_1 + a_4)x$ satisfies $\delta_1(i) = f_a(i), i = 1, 3, 9$, and in a similar way, $\delta_2(x) = a_0 + 8a_3 + (a_1 + 8a_4)x$ passes through the coordinates with locations $2, 6$, and 5. It is also possible to construct a linear polynomial $\delta_3(x)$ that passes through the locations $4, 12$, and 10. This property supports local recovery of any one symbol. Indeed, if the value $f_a(1)$ is unavailable, we can compute $\delta_1(x)$ from its values $\delta_1(3), \delta_1(9)$ and find $f_a(1) = \delta_1(1)$. For instance, for the codeword c in (2) we obtain $\delta_1(x) = 2x + 2$ and find the correct value $\delta_1(1) = 4$. This procedure is schematically shown in Fig. 1.

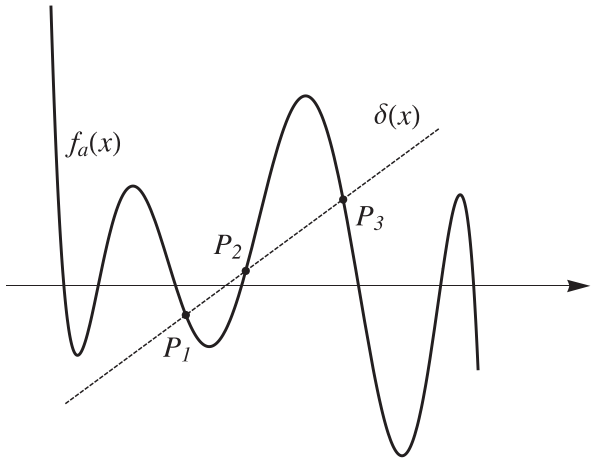


Fig. 1: Local recovery by polynomial interpolation

The local recovery described constitutes a saving compared to the standard decoding of RS codes which calls for computing the polynomial f_a of degree 4 from some of its 5 values. Note also a special property of the construction: the described linear polynomials pass through 3 points of the graph of f_a, which is one point more than is guaranteed by the general interpolation. That this becomes possible is an artifact of the special choice of the polynomial space $\mathcal{P}$ and the set of points A.

A. General Construction of Optimal LRC Codes

There are several classical bounds on the distance of the code in terms of its length and dimension. One of them is the Singleton bound, and a code that meets it is called an MDS code. Moreover, the MDS conjecture (partially proved recently in [1]) claims that, loosely speaking, in order to attain the Singleton bound, the code alphabet has to be of the order

[1]The RS code is obtained by evaluating *all the polynomials* of degree $\deg f \leq k - 1 = 4$.

of the length of the code. The Singleton bound was extended to codes with locality in [7] which showed that the distance $d(\mathcal{C})$ of an (n,k,r) LRC code $\mathcal{C}$ is bounded by

$$d(\mathcal{C}) \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 2. \quad (3)$$

An LRC code whose parameters meet this bound with equality is called optimal. Taking $r = k$ in (3), we recover the Singleton bound, so any (n,k) RS code is an optimal (n,k,k) LRC code. Likewise, the subcode of an RS code constructed in Example 1 is also an optimal LRC code. This suggests that to construct optimal LRC codes for a broad range of the parameters n,k,r, it suffices to take an alphabet of size q comparable to n, and RS codes and their subcodes form natural candidates for optimal LRC codes. We will show that this is indeed the case by providing such a construction which we call an RS-type LRC code.

In Example 1 we implicitly defined a partition of the set of locations into subsets $A_1 = \{1,3,9\}, A_2 = \{2,6,5\}, A_3 = \{4,12,10\}$ such that for each of them, there is a linear polynomial that passes through all the codeword coordinates in these locations. Building on this intuition, let us take a subset $A \subset \mathbb{F}_q$ of n points that label the coordinates of the code. Suppose that there is a partition $\mathcal{A} = \{A_1, ..., A_{\frac{n}{r+1}}\}$ of the set A into $n/(r+1)$ subsets of size $r+1$ and that there exists a polynomial $g(x)$ of degree $r+1$ such that g is constant on the blocks of the partition, i.e.,

$$g(\alpha) = g(\beta) \text{ for any } \alpha, \beta \in A_i, i = 1, \ldots, n/(r+1). \quad (4)$$

We aim at constructing a linear k-dimensional code $\mathcal{C} : \mathbb{F}_q^k \to \mathbb{F}_q^n$. Given a vector $a \in \mathbb{F}_q^k, a = (a_{ij}, i = 0, \ldots, r-1, j = 0, \ldots, \frac{k}{r} - 1)$, define the polynomial[2]

$$f_a(x) = \sum_{i=0}^{r-1} x^i \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(x)^j. \quad (5)$$

and note that $\deg f_a \leq k + \frac{k}{r} - 2$.

Definition 2: Let $\mathcal{P}$ be the set of polynomials of the form (5) and define the code

$$\mathcal{C} = \{ev_A(f_a), f_a \in \mathcal{P}\}. \quad (6)$$

The subsets A_i are called *recovery sets*. Once we specify a location α such that $A_i \ni \alpha$, the subset $A_i \backslash \{\alpha\}$ is called the recovery set of α. The main result about this code family is as follows.

Theorem 2.1: The code $\mathcal{C}$ defined in (6) is an optimal (n,k,r) LRC code. The local recovery of the symbol in location α is accomplished by computing a polynomial $\delta(x)$ of degree $r-1$ that passes through all the points of the recovery set of this location.

Sketch of the proof: The distance of $\mathcal{C}$ equals n minus the maximum number of zeros of $f_a(x)$, and is seen to meet the bound (3). The claim that $\dim \mathcal{C} = k$ becomes obvious once we observe that the k polynomials $g(x)^j x^i$ are all of different degrees and therefore span a k-dimensional subspace of $\mathbb{F}_q[x]$. Furthermore, the polynomials f_a are evaluated at $n > k$ distinct points of the field, therefore the evaluation mapping (6) is injective and the code is of dimension k.

The local recovery is accomplished as follows. Given the erased location $\alpha \in A_i$, find the unique polynomial $\delta(x)$ of degree at most $r-1$ that intersects the graph of $f_a(x)$ at all the other r points of the set A_i :

$$\delta(\beta) = f_a(\beta), \ \beta \in A_i \backslash \{\alpha\}.$$

Note that $g(x)$ is constant on A_i, and therefore $\delta(x)$ is the polynomial

$$\delta(x) = \sum_{i=0}^{r-1} x^i \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(\alpha)^j.$$

Hence, the symbol at location α equals to $\delta(\alpha) = f_a(\alpha)$. ■

Let us make a few observations about the features of the code family.

(i) CONSTRUCTING $g(x)$: The main ingredient of the construction is the polynomial $g(x)$ whose existence is a priori not so obvious. It is not difficult to prove by counting that the required $g(x)$ exists, but we would like to be able to construct it efficiently. This question will be discussed in the next subsection, and it will also enable us to establish relations between the code length n and the size of the alphabet q. The property that $g(x) = \text{const}$ on $A_j, 1 \leq j \leq n/(r+1)$ also has a natural geometric interpretation which provides a segue to constructing LRC codes on algebraic curves (more on this in Sect. IV-B below).

(ii) DIVISIBILITY ASSUMPTIONS: Both the assumptions $r|k$ and $(r+1)|n$ can be removed. To lift the first one, we simply modify the polynomials $f_a(x)$ by taking the inner sum in (5) to go to $\lfloor \frac{k}{r} \rfloor$ or $\lfloor \frac{k}{r} \rfloor - 1$ depending on whether $i < k \bmod r$ or not. As a result, the properties of the code do not change; in particular, it remains optimal. To construct codes of arbitrary length n, removing the constraint $(r+1)|n$, we take the last recovery set to be of a smaller size as needed. Most properties of the code again do not change, although its distance can be one less than the optimal value given by (3).

(iii) LRC REED-SOLOMON CODES: The codes introduced in this section form a direct extension of the classical *Reed-Solomon* codes; in particular, the code $\mathcal{C}$ is a k-dimensional subcode of an $(n, k + \frac{k}{r} - 1)$ RS code. Our construction also reduces to Reed-Solomon codes if r is taken to be k. Indeed, in this case the inner sum in (5) reduces to one term, so $g(x)$ is removed, and we recover the classical definition. Moreover, the set A in this case can be an arbitrary subset of $\mathbb{F}_q$, while the locality condition for $r < k$ imposes a restriction on the choice of the locations.

(iv) SYSTEMATIC ENCODING: Any linear code can be represented in a systematic way, but the described construction can be modified to make this systematic representation explicit and presented in algebraic terms. For $i = 1, ..., k/r$ let $B_i = \{\beta_{i,1}, ..., \beta_{i,r}\}$ be some subset of A_i of size r. For each set B_i define r polynomials $\phi_{i,j}, j = 1, ..., r$ of degree less than r such that $\phi_{i,j}(\beta_{i,l}) = \delta_{j,l}$, and similarly define

[2] We assume that both $\frac{n}{r+1}$ and $\frac{k}{r}$ are integer numbers and comment on the other possibilities in the remarks below.

$m = n/(r+1)$ polynomials $f_i(x)$ such that $f_i(A_j) = \delta_{i,j}$. For k information symbols $a = (a_{i,j}, i = 1,\ldots,k/r; j = 1,\ldots,r)$ construct the polynomial

$$f_a(x) = \sum_{i=1}^{k/r} f_i(x)\Big(\sum_{j=1}^{r} a_{i,j}\phi_{i,j}(x)\Big). \tag{7}$$

Define the evaluation code $\mathcal{C}^{(\mathrm{sys})} := \{ev_A(f_a), f_a \in \mathcal{P}\}$ where $\mathcal{P}$ is the set of all polynomials of the form (7). It is easily seen that this code is systematic, and the message symbols are written in the locations of the sets B_i.

A useful general view of these remarks as well as of the code construction itself is related to the study of properties of the polynomial algebra $\mathbb{F}_{\mathcal{A}}[x]$ spanned by the polynomials constant on the blocks of the partition $\mathcal{A}$. This approach and its connections to the code construction are further developed in [16].

B. Piecewise-constant polynomials

In this section we show how to construct a partition $\mathcal{A}$ of $A \subseteq \mathbb{F}_q$ and a polynomial $g(x)$ of degree $r+1$ that is constant on the blocks of the partition. Let $\mathbb{F}_q^*$ and $\mathbb{F}_q^+$ denote the multiplicative and the additive groups of $\mathbb{F}_q$ respectively. The main idea is expressed in the following simple observation.

Proposition 2.2: Let H be a subgroup of $\mathbb{F}_q^*$ or $\mathbb{F}_q^+$. The annihilator polynomial of the subgroup

$$g(x) = \prod_{h\in H}(x-h) \tag{8}$$

is constant on each coset of H.

Proof: Assume that H is a multiplicative subgroup and let $a, a\bar{h}$ be two elements of the coset aH, where $\bar{h} \in H$, then

$$\begin{aligned} g(a\bar{h}) = \prod_{h\in H}(a\bar{h}-h) &= \bar{h}^{|H|}\prod_{h\in H}(a-h\bar{h}^{-1}) \\ &= \prod_{h\in H}(a-h) \\ &= g(a). \end{aligned}$$

The proof for additive subgroups is completely analogous. ■

If H is a multiplicative subgroup of $\mathbb{F}_q^*$, then $g(x)$ in (8) can be written as $g(x) = x^{|H|} - 1$. Equivalently, we can take $g(x) = x^{|H|}$. Accordingly, the code length n can be any multiple of $r+1$ satisfying $n \le q-1$ (or $n \le q$ in the case of the additive group). In Example 1 we made the following choices: (i) H is the group of cube roots of unity modulo 13, (ii) $A = H \cup 2H \cup 4H$ a union of three cosets (note that we can take any three cosets of the full set of cosets), and (iii) $g(x) = x^3$ (instead of $g(x) = x^3 - 1$).

Example 2: In this example we construct an optimal $(12, 6, 3)$ LRC code with distance $d = 6$ using the additive group of the field. Let α be a primitive element of the field $\mathbb{F}_{2^4}$ and take the additive subgroup $H = \{x + y\alpha : x, y \in \mathbb{F}_2\}$. The polynomial $g(x)$ in (8) equals

$$\begin{aligned} g(x) &= x(x+1)(x+\alpha)(x+\alpha+1) \\ &= x^4 + (\alpha^2+\alpha+1)x^2 + (\alpha^2+\alpha)x. \end{aligned}$$

Let A be the union of any 3 out of the 4 cosets of H. For $a = (a_{i,j}, i = 0,1,2;\ j = 0,1) \in \mathbb{F}_{2^4}^6$ let

$$f_a(x) = \sum_{i=0}^{2}(a_{i,0} + a_{i,1}g(x))x^i.$$

Constructing a code $\mathcal{C}$ by evaluating the polynomials $f_a(x)$ at the points of A, we obtain an LRC code with locality $r = 3$. Note that any $(12, 6)$ MDS code over $\mathbb{F}_{2^4}$ has minimum distance $d = 7$ and locality $r = 6$. The distance of the code $\mathcal{C}$ is only one less than that, but at the same time the locality is decreased by a factor of two, from 6 to 3.

The method described above gives a way of constructing piecewise-constant polynomials, while at the same time constraining the possible values of the code length due to the natural divisibility constraints. We conclude by noting that the additive and multiplicative structures of the field can be combined into a more general method of constructing the polynomials, increasing the range of options for the code length [16, Section III.B].

III. Extensions: Multiple recovery sets; Correcting more than one erasure

A. Algebraic LRC codes with multiple recovery sets

In distributed storage applications there are fragments of the data that are accessed more often than the remaining contents (they are called "hot data"). In the case that such fragments are accessed simultaneously by many users of the system, it may be desirable to ensure that every symbol has several *disjoint recovery sets*, increasing the instantaneous availability of the data.

Using this as a motivation, let us generalize the definition of LRC codes as follows. A code over the alphabet $\mathcal{Q}$ is said to be *locally recoverable with two recovery sets* (an LRC(2) code) if for every $i \in \{1,\ldots,n\}$ there exist disjoint subsets $\mathcal{R}_{i,1}, \mathcal{R}_{i,2} \subset [n]\backslash\{i\}$ and functions $\phi_{i,j}, j = 1,2$ such that for every codeword $c \in \mathcal{C}$

$$c_i = \phi_{i,j}(c_\ell, \ell \in \mathcal{R}_{i,j}), \quad j = 1,2. \tag{9}$$

Suppose that $|\mathcal{R}_{i,1}| \le r_1, |\mathcal{R}_{i,2}| \le r_2$ for all i (we do not assume that $r_1 = r_2$). We write the parameters of an LRC(2) code of dimension k as $(n, k, \{r_1, r_2\})$.

Among the obvious ways to construct LRC(2) codes are various two-level constructions such as product codes or codes on bipartite graphs. We focus on algebraic constructions, extending the approach of the previous section to multiple recovery sets.

Suppose that $\mathcal{A}_1$ $(\mathcal{A}_2)$ is a partition of a set $A \subset [n]$ into subsets of size r_1+1 (resp., r_2+1). Call the partitions $\mathcal{A}_1, \mathcal{A}_2$ *orthogonal* if

$$|A_{1,i} \cap A_{2,j}| \le 1 \quad \text{for all } A_{1,i} \in \mathcal{A}_1, A_{2,j} \in \mathcal{A}_2.$$

If the partitions $\mathcal{A}_1$ and $\mathcal{A}_2$ are orthogonal, then it is possible to construct a code in which every symbol has two *disjoint* recovery sets of size r_1 and r_2. The construction relies on polynomial evaluation and is very similar to the construction of Section II-A. To give an example, consider the field $\mathbb{F}_{16}$. Its

additive group $\mathbb{F}_{16}^+$ contains several pairs of subgroups $G \cong H \cong (\mathbb{Z}_2)^2$ such that $G \cap H = 0$. For instance, take $G = \{0, 1, \alpha, \alpha^4\}$ and $H = \{0, \alpha^2, \alpha^3, \alpha^6\}$, where α is a primitive element that satisfies $\alpha^4 = \alpha + 1$. The subgroups G and H define a pair of orthogonal partitions of $\mathbb{F}_{16}^+$ given by

$$\mathcal{A}_G = \{G, \alpha^5 + G, \alpha^6 + G, \alpha^7 + G\}$$
$$\mathcal{A}_H = \{H, 1 + H, \alpha + H, \alpha^4 + H\}.$$

Using each of these partitions, we can construct an LRC code $\mathcal{C}$ with the parameters $(n = 16, k, \{r_1 = 3, r_2 = 3\})$ of dimension $k, 1 \le k \le 8$. Every coordinate of the codeword can be recovered in two independent ways: for instance, the coordinate c_α is found by computing the polynomial $\delta_1(x)$ of degree at most 2 that passes through the points c_0, c_1, c_{α^4} *as well as* the polynomial $\delta_2(x)$ that passes through $c_{\alpha^5}, c_{\alpha^9}, c_{\alpha^{11}}$. Then we have $c_\alpha = \delta_1(\alpha) = \delta_2(\alpha)$.

It is easy to identify a necessary and sufficient condition for two subgroups to generate orthogonal partitions.

Proposition 3.1: Let H and G be two subgroups of a finite group X, then the coset partitions $\mathcal{H}$ and $\mathcal{G}$ defined by H and G respectively are orthogonal iff the subgroups intersect trivially, namely

$$H \cap G = 1.$$

If the group X is cyclic, then it is equivalent to requiring that $\gcd(|H|, |G|) = 1$.

In the context of finite fields we can use both the multiplicative and the additive group of the field to construct LRC(2) codes. It is also easy to find several subgroups that intersect trivially; in particular, this is clearly possible for the additive group $\mathbb{F}_q^+$ in the case of a non-prime q. At the same time, constructing LRC(2) codes from a multiplicative subgroup of $\mathbb{F}_q, q = p^l$ requires one extra condition, namely, that $q - 1$ is not a power of a prime. In this case, we can find two subgroups of $\mathbb{F}_q^*$ of coprime orders that give rise to orthogonal partitions of $\mathbb{F}_q^*$.

Proposition 3.2: Let $\mathbb{F}_q$ be a finite field such that the $q - 1$ is not a power of a prime. Let $r_1, r_2 > 1, gcd(r_1, r_2) = 1$ be two factors of $q - 1$. Then there exists an LRC(2) code $\mathcal{C}$ of length $q - 1$ over $\mathbb{F}_q$ such that every code symbol has two disjoint recovery sets of sizes $r_1 - 1$ and $r_2 - 1$.

The discussed construction gives codes with distance close to the upper bound on LRC(2) codes derived in [17].

The definitions and constructions of this section extend straightforwardly to an arbitrary number $t \ge 2$ of recovery sets, giving rise to easily constructible LRC(t) codes with the parameters $(n, k, \{r_1, \ldots, r_t\})$, where $r_i + 1, i = 1, \ldots, t$ is the size of the blocks in the corresponding partition. At the same time, note that for $t \ge 3$ better parameters are obtained using random expanders; see [17, Theorem C]. Paper [17] also contains results on upper bounds for codes with an arbitrary number of recovery sets.

B. Correcting more than one erasure: $(r + \rho - 1, r)$ Local MDS Codes

A more general version of the local recovery problem calls for correcting more that one erasure within each recovery set. To address this task, we consider LRC codes in which the set of coordinates is partitioned into several subsets of cardinality $r + \rho - 1$ such that every local code is an $(r + \rho - 1, r)$ MDS code, where $\rho \ge 3$. Under this definition, every symbol of the codeword is a function of any r out of the $r + \rho - 2$ symbols, increasing the chances of successful recovery. A compact notation for such codes is (n, k, r, ρ) LRC codes, where n is the block length and k is the code dimension. A generalization of the bound on the distance (3) to the case of (n, k, r, ρ) LRC codes takes the form [11]

$$d \le n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\rho - 1). \tag{10}$$

As before, we will say that the LRC code is optimal if its minimum distance attains this bound with equality.

We assume that $(r + \rho - 1)|n$ and $r|k$, although the latter constraint is again unessential. To construct the code using the ideas of Sect. II-A, we begin with a partition $\mathcal{A} = \{A_1, \ldots, A_m\}, m = n/(r + \rho - 1)$ of the set $A \subset \mathbb{F}, |A| = n$, such that $|A_i| = r + \rho - 1, 1 \le i \le m$. Let $g \in \mathbb{F}[x]$ be a polynomial of degree $r + \rho - 1$ that is constant on each of the blocks A_i. Given a message vector $a \in \mathbb{F}^k$, let us write it as $a = (a_0, ..., a_{r-1}) \in \mathbb{F}^k$, where each $a_i = (a_{i,0}, ..., a_{i,\frac{k}{r}-1})$ is a vector of length k/r. In analogy to (5), define the polynomial

$$f_a(x) = \sum_{i=0}^{r-1} \sum_{j=0}^{\frac{k}{r}-1} a_{ij} g(x)^j x^i.$$

The properties of the obtained codes are summarized in the following theorem.

Theorem 3.3: Let $\mathcal{C} : \mathbb{F}_q^k \to \mathbb{F}_q^n$ be a linear code defined as the image of the evaluation map $a \mapsto ev_A(f_a)$. Then $\mathcal{C}$ is an optimal (n, k, r, ρ) LRC code in which local recovery of an erasure at location α can be performed by polynomial interpolation over any r locations of its recovery set.

IV. More algebraic connections: Cyclic LRC codes and LRC codes on curves

In classical coding theory there are two code families related to RS codes, namely cyclic codes and codes on algebraic curves. Since the codes considered above can be viewed as LRC analogs of RS codes, it is natural to consider these two families in relation to our construction. It turns out that both connections lead to new constructions of LRC codes as well as new problems in algebraic coding theory. Sections IV-A and IV-B below are based on [18] and [2] respectively.

A. Cyclic LRC codes

Cyclic codes form a well-established chapter in coding theory, important both theoretically and in applications. To construct cyclic LRC codes, we will rely on the multiplicative structure of the field $\mathbb{F}_q$. Let us choose the code length n to be a divisor of $q - 1$ and let us assume that the coordinates are labeled by n-th degree roots of unity in $\mathbb{F}_q$, i.e., $A = \{1, \alpha, \ldots, \alpha^{n-1}\}$, where α is a primitive root. Suppose that $(r + 1)|n$ and let $m = n/(r + 1)$ be their quotient. We rely on Proposition 2.2 to construct the polynomial $g(x)$.

Let H be a subgroup of the group $\mathbb{F}_q^*$ of order $|H|$ and let $r = |H| - 1$.

According to the discussion after Proposition 2.2, we can take $g(x) = x^{r+1}$. Examination of the expression (5) shows that the polynomial $f_a(x)$ can be written in the form

$$f_a(x) = \sum_{\substack{i=0 \\ i \neq r \bmod (r+1)}}^{\frac{k}{r}(r+1)-2} a_i x^i, \tag{11}$$

where the a_i's form the message vector. Following the construction (6), we obtain an LRC code, denoted by $\mathcal{C}$.

It is clear that the code $\mathcal{C}$ is cyclic, and it is easy to find its defining set of zeros. From the classical BCH bound it is well known that a set of $d-1$ consecutive zeros guarantees that $d(\mathcal{C}) \geq d$. The following theorem supplements this claim by identifying the set of zeros of $\mathcal{C}$ that supports the locality property.

Theorem 4.1: Consider the following sets of elements of $\mathbb{F}_q$:

$$L = \{\alpha^i, i \bmod (r+1) = l\}$$
$$D = \{\alpha^{j+s}, s = 0, \ldots, n - \tfrac{k}{r}(r+1)\},$$

where $0 \leq l \leq r$ and $\alpha^j \in L$. The cyclic code with the defining set of zeros $Z := L \cup D$ is an optimal (n,k,r) q-ary cyclic LRC code[3].

If the set Z contains cosets of two groups of roots of unity of coprime orders r_1+1 and r_2+1, then this gives rise to an LRC(2) code $(n,k,\{r_1,r_2\})$ which has two disjoint recovery sets for every coordinate.

The following obvious remark sometimes facilitates the analysis of cyclic LRC codes.

Proposition 4.2: Let $\mathcal{C}$ be a cyclic LRC code with locality r. Suppose that $d^\perp$ is the distance of the dual code $\mathcal{C}^\perp$, then $r = d^\perp - 1$.

So far we were interested in RS-type LRC codes. Subfield subcodes of these codes form a natural analog of the family of BCH codes. Their properties are not so easy to analyze in general, but one possibility has been suggested in [18]. Let $\mathcal{C}$ be an (n,k,r) LRC code over $\mathbb{F}_{q^m}$ and denote by $\mathcal{C}|_{\mathbb{F}_q}$ the subcode of $\mathcal{C}$ formed by the codewords whose coordinates are contained in $\mathbb{F}_q$. Suppose we attempt to construct LRC codes over $\mathbb{F}_q$ as subfield subcodes of RS-type LRC codes over $\mathbb{F}_{q^m}$. Since $\mathcal{C}|_{\mathbb{F}_q} \subset \mathcal{C}$, we have that $d(\mathcal{C}|_{\mathbb{F}_q}) \geq d(\mathcal{C})$. At the same time, the dual distance of a cyclic LRC code $(\mathcal{C}|_{\mathbb{F}_q})^\perp$ may, and often does, decrease from its original value $r+1$. Thus, studying subfield subcodes is an appropriate context for constructing cyclic LRC codes over small alphabets with good distance and small locality.

B. LRC codes on algebraic curves

The RS-type LRC codes constructed above solve the problem of local recovery for codes of length n that is on the order of the size of the alphabet q. Consider again the problem of constructing long LRC codes for a fixed alphabet size. In classical coding theory good codes of this kind are obtained using the Goppa construction of codes on algebraic curves. Here we show how this approach can be utilized for codes with the locality constraint.

We begin with another view of the construction of RS-type LRC codes (4)-(6), focusing on the polynomial $g(x)$. Let $\Bbbk = \mathbb{F}_q$ denote the code alphabet. Recall that $g : \Bbbk \to \Bbbk$ defines a mapping such that there are exactly $r+1$ points that are mapped to every point in the range of g. In other words, we have $|g^{-1}(P)| = r+1$ for all P in the range. Switching to geometric language, let $X = Y = \mathbb{P}^1$ denote the projective line over the field $\Bbbk$, then $g : X \to Y$ is a covering map of lines such that the fiber above any point of Y in its range contains exactly $r+1$ points of X. For instance, in Example 1 the range of $g : x \mapsto x^3$ is the set $\{P_1 = 1, P_2 = 8, P_3 = 12\}$ and $g^{-1}(P_i) = A_i, i = 1,2,3$.

This view of our construction suggests the following generalization to codes on curves. Let X and Y be smooth projective absolutely irreducible curves over $\Bbbk$ and let $g : X \to Y$ be a rational separable map of curves of degree $r+1$. For example, let $\Bbbk = \mathbb{F}_9$ and consider the Hermitian curve X of genus 3 given by the equation $x^3 + x = y^4$. The curve X has 27 points in the finite plane, shown in Fig. 2 below, and one point at infinity.

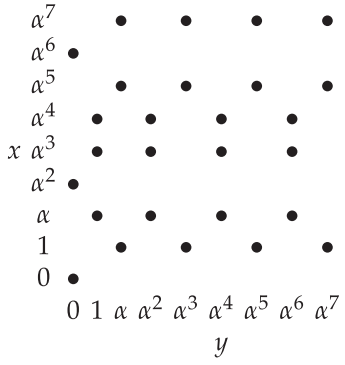


Fig. 2: 27 points of the Hermitian curve over $\mathbb{F}_9$; here $\alpha^2 = \alpha + 1$.

Take $Y = \mathbb{P}^1_{\Bbbk}$, then we can take g to be the map of degree $r+1 = 3$ given by the natural projection $g : (x,y) \mapsto y$. Another possibility is a degree-4 map $g : (x,y) \mapsto x$ whose range does not include the points $0, \alpha^2$, and α^6.

More generally, let $\Bbbk(X)$ and $\Bbbk(Y)$ denote the fields of rational functions on X and Y. By the primitive element theorem there exists a function $x \in \Bbbk(X)$ such that $\Bbbk(X) = \Bbbk(Y)(x)$ and that satisfies an algebraic equation of degree $r+1$ over $\Bbbk(Y)$. The function x can be considered as a map $x : X \to \mathbb{P}^1_{\Bbbk}$, and we denote its degree $\deg(x)$ by h.

The codes that we construct again belong to the class of evaluation codes. Let $S = \{P_1, \ldots, P_s\} \subset Y(\Bbbk)$ be a subset of $\Bbbk$-rational points of Y in the finite space, and let Q_∞ be a positive divisor of degree $\ell \geq 1$ such that $\operatorname{supp} Q_\infty \subset \pi^{-1}(\infty)$, where $\pi : Y \to \mathbb{P}^1_{\Bbbk}$ is a projection map. To construct our codes let us assume that

$$A := g^{-1}(S) = \{P_{ij}, i = 0, \ldots, r, j = 1, \ldots, s\} \subseteq X(\Bbbk); \tag{12}$$

[3]We note that the sets L and D have a nonempty intersection, but their common elements appear in Z only once.

$$g(P_{ij}) = P_j \text{ for all } i, j.$$

Let $\{f_1, \dots, f_m\}$ be a basis of the Riemann-Roch space $L(Q_\infty)$. Our codes will be constructed as evaluations of functions in the $\Bbbk$-subspace V of $\Bbbk(X)$ generated by the functions

$$\{f_j x^i, i = 0, \dots, r-1, j = 1, \dots, m\} \tag{13}$$

(note an analogy with (5)).

Definition 3: (*LRC codes on curves*). Consider the evaluation map

$$\begin{aligned} ev_A : V &\longrightarrow \Bbbk^{(r+1)s} \\ F &\mapsto (F(P_{ij}), i = 0, \dots, r, j = 1, \dots, s), \end{aligned} \tag{14}$$

and denote its image by $\mathcal{C}(Q_\infty, g)$. It is a linear code in the space $\mathbb{F}_q^n, n = (r+1)s$, and since $\operatorname{supp} Q_\infty \cap S = \emptyset$, the code is well defined.

The code coordinates are naturally partitioned into s subsets $A_j = \{P_{ij}, i = 0, ..., r\}, j = 1, \dots, s$ of size $r+1$ each; see (12).

Theorem 4.3: The subspace $\mathcal{C}(Q_\infty, g) \subset \mathbb{F}_q$ forms an (n, k, r) linear LRC code with the parameters

$$\begin{aligned} n &= (r+1)s \\ k &= rm \geq r(\ell - g_Y + 1) \\ d &\geq n - \ell(r+1) - (r-1)h, \end{aligned} \tag{15}$$

provided that the right-hand side of the inequality for d is a positive integer. Local recovery of an erased symbol $F(P_{ij})$ can be performed by polynomial interpolation through the points of the recovery set A_j.

In particular, let us specialize this construction for codes on Hermitian curves. Let $q = q_0^2$, where q_0 is a prime power, let $\Bbbk = \mathbb{F}_q$, and let X be the Hermitian curve, i.e., a plane smooth curve of genus $g_0 = q_0(q_0 - 1)/2$ with the affine equation

$$X : x^{q_0} + x = y^{q_0+1}.$$

The curve X has $q_0^3 = q\sqrt{q}$ rational points in the affine plane. By taking g to be the projection on y as discussed above we obtain a family of LRC codes with the parameters

$$\begin{gathered} n = q_0^3,\ k = (\ell+1)(q_0 - 1),\ r = q_0 - 1 \\ d \geq n - \ell q_0 - (q_0 - 2)(q_0 + 1). \end{gathered}$$

It is also possible to take g to be a projection on x, which gives a family of LRC codes with similar parameters and locality $r = q_0$.

Asymptotically good code families. As in classical coding theory, we obtain infinite families of codes with good parameters by taking asymptotically maximal curves such as, for instance, the Garcia-Stichtenoth towers of curves. These curves are constructed by successively extending the function fields, adding algebraic elements that satisfy equations similar to the equation that defines the Hermitian curves. Similarly to the Hermitian case, there are several variants of the code construction. For instance, it is possible to construct a family of q-ary LRC codes whose rate and relative distance satisfy the asymptotic inequality

$$R \geq \frac{r}{r+1}\left(1 - \delta - \frac{2\sqrt{q}}{q-1}\right), \tag{16}$$

where $r = \sqrt{q}$ and $q = q_0^2$ for some prime power q_0. For $q_0 \geq 23$ this bound improves upon the Gilbert-Varshamov type bound for LRC codes discussed in the next section (see an example in Fig. 3).

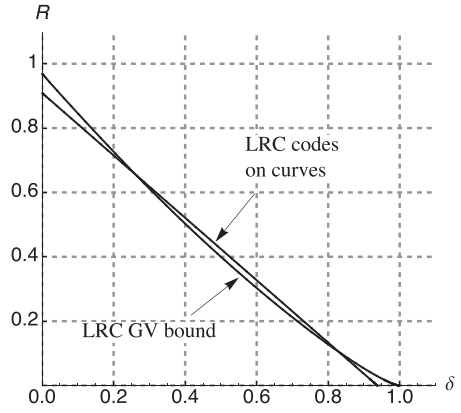


Fig. 3: The bound (16) shown together with the Gilbert-Varshamov type bound ($q_0 = 32$).

While this construction yields sequences of codes with asymptotically good parameters, its locality parameter r is fixed once we choose the code alphabet. In principle one would want to have flexibility in choosing r in a way similar to the construction of RS-type LRC codes. This is indeed possible by studying certain quotients of curves in the Garcia-Stichtenoth tower. As a result, we obtain asymptotically good codes over a fixed field $\mathbb{F}_q$ with a range of values of r with parameters similar to the ones mentioned above.

Concluding this section, note that the proposed approach generalizes to codes with more than one recovery set for every coordinate (the so-called availability problem). Indeed, when discussing the example in Fig. 2 we remarked that there are two natural maps from X to $\mathbb{P}^1$. A closer look confirms that together they define a pair of orthogonal partitions of the set of $n = q_0^3 - q_0 - 1$ affine points of the Hermitian curve, giving rise to an LRC code of length n with two disjoint recovery sets for each codeword symbol.

V. Bounds on the parameters of LRC codes

Here we discuss bounds on the rate and distance of LRC codes introduced in Definition 1. It can be easily seen that the rate of any LRC code $\mathcal{C}$ with locality r is at most $R(\mathcal{C}) \leq r/(r+1)$. Intuitively this is justified by the fact that any $r+1$ codeword symbols within a recovery set satisfy a functional relation, so they contain at most r information symbols.

How large can $d(\mathcal{C})$ be? Even in the classical coding problem, this question is addressed in more than one way, depending on whether we account for the value of q or not. The Singleton-type bound (3), discussed above, does not depend

on the size of the alphabet. A bound that accounts for the value of q, proved in [3], has the following form:

$$k \leq \min_{s \geq 1} \{sr + k_q(n - s(r+1), d)\}, \tag{17}$$

where $k_q(n,d)$ is the maximum dimension of a code of length n and distance d over $\mathbb{F}_q$ (with no locality assumptions). It is also possible to derive lower Gilbert-Varshamov-type bounds on the parameters of LRC codes using the probabilistic method, bringing the state of bounding the parameters of LRC codes to the same status as bounds on classical error correcting codes. In particular, sequences of codes of asymptotically positive rate exist if and only if the number of correctable errors does not exceed the $(q-1)/2q$ proportion of the code length. The results of [3], [17] imply that the same conclusion is valid once we add the locality constraint (for any constant r). Therefore, adding the locality constraint does not shift the "Plotkin point" for asymptotic relative distance from the value $(q-1)/q$. The best asymptotic lower and upper bounds on LRC codes are shown in Fig. 4.

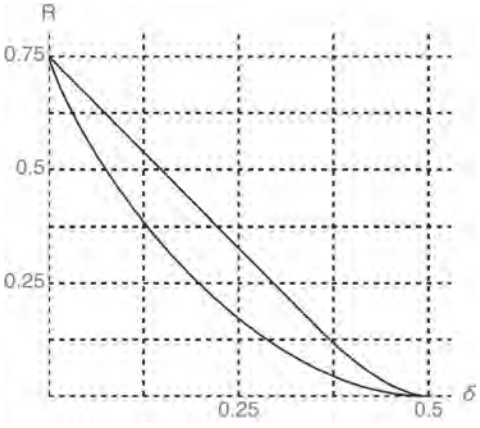


Fig. 4: Asymptotic bounds for the rate R of binary LRC codes as a function of the relative distance δ; $r = 3$. The upper curve is obtained from the bound (17), and the lower curve is a GV-type bound.

VI. Outlook

A. LRC codes in industry

Apart from their theoretical merits, LRC codes offer an efficient solution for data protection in large-scale distributed storage systems. Data encoding schemes employed by companies using or providing distributed storage solutions are based primarily on the ease of implementation, update, and maintenance. Driven by these metrics, companies are mostly interested in implementing LRC codes that provide the locality property only for the information part of the codeword. Codes with this property are said to have *information symbol* locality. It turns out that constructing such codes with good minimum distance is relatively simple, which is why these codes are popular in current industry solutions.

To construct an (n,k,r) LRC code with information symbol locality and good minimum distance, begin with an $(n - \frac{k}{r}, k)$ MDS code (typically an RS code). To account for locality, let us partition its k information symbols into k/r disjoint sets of size r and add one parity check symbol for each set.

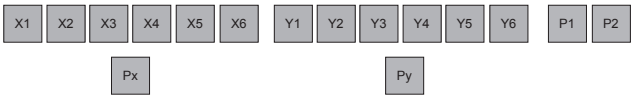


Fig. 5: $(16,12,6)$ LRC code used in Windows Azure storage [10].

This results in a code of length n with information locality r. Examples of LRC codes constructed in this way are already used in practice or have been tested by industry, and here we list a few of them.

The free software storage platform *Ceph* enables the users to protect their information by simple replication, RS code, or an LRC code. In another project [13], the authors constructed a $(16,10,5)$ LRC code based on the $(14,10)$ RS code and tested it on a cluster at Facebook's data warehouse. The construction proposed in [13] has in fact the all-symbol locality property. Finally, Windows Azure Storage (WAS), Microsoft's scalable cloud storage system that has been in use for some years [10], uses a $(16,12,6)$ LRC code shown in Fig. 5. Here P_1 and P_2 are the global parities found from all the 12 information symbols $X_i, Y_i, i = 1,...,6$. They are employed in cases of more than one failure among the nodes. The symbols P_x and P_y are the parities that provide local recovery for the information symbols by accessing 6 other symbols within the recovery set.

Encouraged by the fast embrace of LRC codes by large-scale users of distributed storage, we believe that there is room for implementation and testing of other code families with the locality property. Specific storage applications may benefit from all-symbol locality or large minimum distance. At the same time, the solutions should be tailored to the needs of the application, including update complexity, security and availability of the data, and other features.

B. LRC codes on graphs

An interesting generalization of the LRC coding problem is related to local recovery that is constrained by the topology of the computer network. Consider a graph on n storage nodes whose edges describe the available communication links between the nodes. Similarly to the problem studied above, we require that every node can recover its storage contents by reading the information stored in its neighbor nodes in the graph. A set of vectors over a finite alphabet that can be stored in the nodes to satisfy this constraint, forms an LRC code on the graph, and we seek such codes of the largest possible size. This problem was recently introduced in [12] (in a different form, it was also studied earlier in [5]). It is also shown to be (in some sense) a dual of the well-known index coding problem [12], [14]. Major open questions in this area include finding constructions of good codes for the graph LRC problem, for instance, for families of graphs with some structure, as well as advancing connections between LRC codes and index coding.

C. Maximally recoverable codes: Can a code be LRC and MDS?

MDS codes form a practically appealing family because they provide the best possible error resilience for a given

amount of storage overhead. In formal terms, this amounts to saying that any k symbols in a k-dimensional MDS code form an information set. At the same time, the locality constraint requires some dependence among the codeword symbols, so locality and the MDS property cannot be combined in one construction. How close to being both MDS and LRC can a code be? This question brings in the following natural definition: Call an (n,k,r) LRC code *maximally recoverable* [6] if every k coordinates that do not contain a full recovery set form an information set. Note any k-subset that contains a recovery set cannot be an information set.

For large sets of parameters (n,k,r) maximally recoverable codes have been constructed in [15], [19], [6]; however, none of these results yield code families over alphabets q of size comparable to the code length. At the same time, as shown above, it is possible to construct LRC codes over small alphabets. This gives rise to the following *open problem*: is it possible to construct maximally recoverable codes with small q, or does maximality necessarily require a superlinear alphabet size?

Observe that the maximality property is not resolved even for the RS-type LRC code family presented in this paper: we do not know if (apart from the trivial cases of $r = 1, k$) among the constructed codes there are maximally recoverable ones.

D. AG codes: Parameters and availability

The construction of LRC codes on curves in Sect. IV-B is rather general in the sense that it applies to any pair of curves equipped with a covering map. At the same time, the estimates of the parameters of the obtained codes derived using this general approach do not take into account specifics of individual families of curves, and for this reason may be somewhat crude. Thus, the initial results reported above could be specialized and improved in examples that rely on properties of specific curves and their maps.

Another problem, mentioned only very briefly in Sect. IV-B and in [2] concerns the availability problem for algebraic geometric codes. While we have explored the most natural approach to this problem, the parameters of the obtained codes are far from optimal. It may be possible to obtain better LRC codes relying on the automorphism groups of curves and their codes, and we envision this as another avenue for further studies.

Yet another topic of possible studies is related to decoding of the constructed codes. While we have focused on local erasure recovery, occasionally we will face the task of global decoding for the purpose of error and erasure correction. Here we tacitly rely on the existing decoding algorithms of algebraic geometric codes, although conceivably the structure of our codes could support decoding algorithms designed specifically for this family. It is also of interest to explore the connection of these codes with generic list decoding algorithms of codes on curves.

ACKNOWLEDGMENTS: We would like to thank our coauthors Robert Calderbank, Alexey Frolov, Sreechakra Goparaju, and Serge Vlăduţ for their collaboration on [18], [17] and [2], as well as the Editor for encouraging us to write this article. We are also grateful to our colleagues Alexandros G. Dimakis and P. Vijay Kumar for insightful discussions of the problems covered in this paper.

A.B. acknowledges support of the NSF grants CCF1422955 and CCF1217245.

REFERENCES

[1] S. Ball, "On sets of vectors of a finite vector space in which every subset of basis size is a basis," *J. Eur. Math. Soc.*, vol. 14, no. 3, pp. 733–748, 2012.

[2] A. Barg, I. Tamo, and S. Vlăduţ, "Locally recoverable codes on algebraic curves," in *Proc. IEEE Int. Sympos. Inform. Theory, Hong Kong*, 2015, pp. 1252–1256.

[3] V. Cadambe and A. Mazumdar, "Upper bounds on the size of locally recoverable codes," *IEEE Trans. Inform. Theory*, vol. 61, no. 11, pp. 5787–5794, 2015.

[4] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.

[5] M. Gadouleau and S. Riis, "Graph-theoretical constructions for graph entropy and network coding based communications," *IEEE Trans. Inform. Theory*, vol. 5 , no. 10, pp. 6703–6717, 2011.

[6] P. Gopalan, C. Huang, B. Jenkins, and S. Yekhanin, "Explicit maximally recoverable codes with locality," *IEEE Trans. Inform. Theory*, vol. 60, no. 9, pp. 5245–5256, 2014.

[7] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inform. Theory*, vol. 58, no. 11, pp. 6925–6934, 2011.

[8] J. Han and L. A. Lastras-Montano, "Reliable memories with subline accesses," in *Proc. IEEE Internat. Sympos. Inform. Theory*, 2007, pp. 2531–2535.

[9] C. Huang, M. Chen, and J. Li, "Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems," in *Sixth IEEE International Symposium on Network Computing and Applications*, 2007, pp. 79–86.

[10] C. Huang, H. Simitci, Y. Xu, A. Ogus, B. Calder, P. Gopalan, J. Li, and S. Yekhanin, "Erasure coding in windows Azure storage," in *Proc. USENIX Annual Technical Conference (USENIX ATC)*, 2012.

[11] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with local regeneration and erasure correction," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4637–4660, 2014.

[12] A. Mazumdar, "Storage capacity of repairable networks," *IEEE Trans. Inform. Theory*, vol. 61, no. 11, pp. 5810–5821, 2015.

[13] M. Sathiamoorthy, M. Asteris, D. Papailiopoulos, A. G. Dimakis, R. Vadali, S. Chen, and D. Borthakur, "XORing elephants: Novel erasure codes for big data," in *Proceedings of the VLDB Endowment*, 2013.

[14] K. Shanmugam and A. Dimakis, "Bounding multiple unicasts through index coding and locally repairable codes," in *Proc. IEEE Internat. Sympos. Inform. Theory, Honolulu, HI*, 2014, pp. 296–300.

[15] N. Silberstein, A. S. Rawat, O. Koyluoglu, and S. Vishwanath, "Optimal locally repairable codes via rank-metric codes," in *Proc. IEEE Int. Sympos. Inform. Theory, Boston, MA*, 2013, pp. 1819–1823.

[16] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 4661–4676, 2014.

[17] I. Tamo, A. Barg, and A. Frolov, "Bounds on the parameters of locally recoverable codes," Available online at arXiv:1506.07196, 2015.

[18] I. Tamo, A. Barg, S. Goparaju, and R. Calderbank, "Cyclic LRC codes and their subfield subcodes," in *Proc. IEEE Int. Sympos. Inform. Theory, Hong Kong, PRC*, 2015, pp. 1262–1266.

[19] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," in *Proc. 2013 IEEE Internat. Sympos. Inform. Theory*, 2013, pp. 1814–1818.

Cache Networks: An Information-Theoretic View

Mohammad Ali Maddah-Ali and Urs Niesen

***Abstract*—Caching is a popular technique that duplicates content in memories distributed across the network in order to enhance throughput and latency in a variety of applications. Cache systems were the subject of extensive study, mostly in the computer science community, in the 80s and 90s. However, the fundamental results derived during that period were mainly developed for systems with just a single cache and only heuristically extended to networks of caches. In this newsletter article, we argue that information theory can play a major role in establishing a fundamental understanding of such cache networks. In particular, we show that cache networks can be cast in an information-theoretic framework. Using this framework, we demonstrate that the aforementioned heuristics, which utilize caches to deliver part of the content locally, can be highly suboptimal when applied to cache networks. Instead, we identify cache memories as limited spaces to plan side information from among a fixed set of pre-recorded content (e.g., movies) to facilitate future communication. This new understanding of the role of caching creates various coding and signaling opportunities and can offer gains that scale with the size of the network.**

I. Introduction

Caching is an essential technique to improve throughput and latency in a vast variety of applications such as virtual memory hierarchies in CPU design, web caching for content delivery networks (CDNs), and inquiry caching in domain name systems. The core idea of caching is to use memories distributed across the network to duplicate data. This stored data can be then used to facilitate the delivery of future requests, thereby reducing network congestion and delivery delay. Companies like Akamai, Facebook, Netflix, Google, etc. are heavily investing in their cache networks to increase the performance of their systems.

There is a rich and beautiful theory, developed mostly in the computer science community during the 80s and 90s, for systems with a single cache. However, when it comes to networks of caches, the existing theory falls short, and engineers instead rely on heuristics and the intuition gained from the analysis of single-cache systems. Quoting Van Jacobson, one of the key contributors to TCP/IP and expert on content distribution:

> "ISPs are busily setting up caches and CDNs to scalably distribute video and audio. Caching is a necessary part of the solution, but there is no part of today's networking—from Information, Queuing, or Traffic Theory down to the Internet protocol specs—that tells us how to engineer and deploy it." [1, p. 302]

We argue that information theory can in fact provide the theoretical underpinnings for the deployment and operation of cache networks. Indeed, we show that the caching problem can be formulated as a network information theory problem. Applying information-theoretic tools for the analysis of cache networks reveals that the conventional way to operate these networks can be substantially suboptimal.

M. A. Maddah-Ali is with Bell Labs, Alcatel-Lucent. U. Niesen is with Qualcomm's New Jersey Research Center. Emails: mohammadali.maddah-ali@alcatel-lucent.com, urs.niesen@ieee.org.

Cache networks have two distinctive features that differentiate them from other problems in multi-user information theory:

- *Budget for side information:* In information theory, we often consider networks with side information, with the objective to characterize system performance in the presence of *given* side information. In contrast, in cache networks, the side information itself is subject to design and optimization. Each cache has a fixed memory limit, and the system designer is allowed to *choose* the side information in the cache subject to the memory constraint.
- *Pre-recorded content:* In network information theory, we usually assume that each source locally generates a message (e.g., voice) at transmission time for a particular user/destination. However, over the last decade or so, the bulk of traffic has shifted to content (e.g., movies), which is typically recorded centrally well ahead of transmission time, and which is not generated for a particular user/destination. It is this generation of messages ahead of transmission time that allows their duplication across the network.

In the remainder of this newsletter article, we discuss various opportunities and challenges in the area of cache networks with emphasis on the role of information theory in offering a fundamental view on this problem. We start in Section II with a canonical cache network, which provides an information-theoretic framework for the analysis of such systems. We then review an approximately optimal solution for this problem and compare it to conventional approaches. We proceed with recent results on cache networks in a variety of scenarios, comparing offline versus online caching, delay-tolerant versus delay-sensitive content (both in Section III), single layer versus hierarchical caching, server-oriented versus device-to-device settings (both in Section IV), among others. Throughout, we point out open problems motivated by real-life applications of caching.

II. Canonical Cache Network

We consider the following canonical cache network introduced in [2]. A server is connected through a shared bottleneck link to K users as shown in Fig. 1. The server has a database of N files $W_1, \dots, W_N$ each of size F bits. Each user k has an *isolated* private cache memory of size MF bits for some real number $M \in [0, N]$. In this article, we assume $N \geq K$ to simplify the exposition.

This setting can model a wireless network with an access point and several users, all sharing the common wireless

channel. It can also model a wireline network with several caches connected to a common server; here the shared link models a bottleneck along the path between the server and the users.

The system operates in two phases: a *content placement phase* and a *content delivery phase*. The placement phase occurs during a time of low network traffic, say in the early morning, so that network resources are abundant and cheap. The main constraint during this phase is the limited cache memory. We model this placement phase by giving each user access to the entire database $W_1, \ldots, W_N$ of files. Each user is thus able to fill its own cache subject only to the memory constraint of MF bits. Critically, in the placement phase, the system is not aware of users' future requests, so that the content cached by the users cannot depend on them.

The delivery phase occurs after the placement phase during a time of high network traffic, say in the evening. Network resources are now scarce and expensive and become the main constraint. We model this delivery phase as follows. Each user k requests one of the files W_{d_k} in the database. The server is informed of these requests and responds by sending a signal of size RF bits over the shared link for some fixed real number R called the *rate*. This signal sent from the server has to be constructed such that each user can recover its requested file from the signal received over the shared link and the contents of its own cache.

We need to design both the content placed in the users' caches during the placement phase and the signal sent by the server during the delivery phase. The objective is to minimize the rate R subject to the constraint that every possible set of user demands can be satisfied. We again emphasize that, while the signal sent over the shared link during the delivery phase is a function of the users' requests, the cache content designed during the earlier placement phase cannot depend on those requests (since they are unknown at the time). In addition, since R is determined with respect to the worst possible user requests, the cache content cannot be tuned for a specific set of requests.

Example 1 (*Uncoded Caching*)**.** As a baseline, let us review a conventional uncoded solution, where in the placement phase each user caches the same M/N fraction of each file. The motivation for this approach is that the system should be ready for any possible demand, therefore each user should give the same fraction of its memory to each file. Moreover, since there is no statistical difference in the user demands known during the placement phase, the content of the caches for different users should be the same.

In the delivery phase, the server simply transmits the remaining $1 - M/N$ fraction of any requested file over the shared link, and thus each user can recover its requested file. Since, there are K requests to be delivered, the worst-case delivery rate is

$$R_U(M) \triangleq K \cdot (1 - M/N). \tag{1}$$

The function $R_U(M)$ describes the trade-off between rate and memory for the baseline uncoded caching scheme. The factor K in (1) is the rate that we would achieve without access to any caches. The factor $1 - M/N$ arises because an M/N fraction of each file is locally cached at each user. We call this second factor in (1) the *local caching gain*.

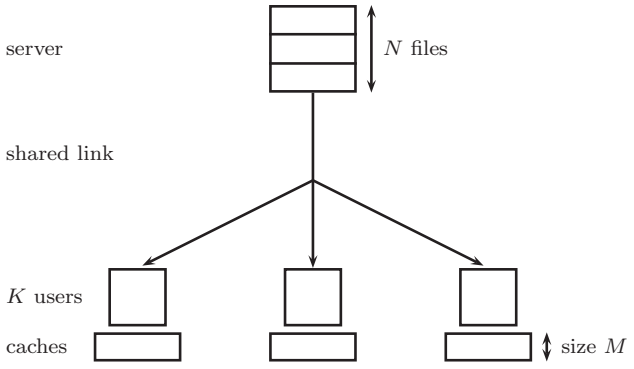


Fig. 1. Canonical cache network from [2]: A server containing N files of size F bits each is connected through a shared link to K users each with an isolated cache of size MF bits. The goal is to design the placement phase and the delivery phase such that the peak rate of the shared bottleneck link is minimized. In the figure, $N = K = 3$ and $M = 1$.

We refer to this caching strategy as *uncoded caching*, since both the content placement and delivery are uncoded. From the above discussion, we see that the role of caching in this uncoded scheme is to deliver part of the requested content *locally*. ◊

The uncoded caching scheme in Example 1 is just one among a long list of conventional uncoded approaches, developed for different applications, scenarios, and objectives. This includes popular schemes such as least-recently used (LRU) and least-frequently used (LFU) (see, e.g., [3]). All these conventional approaches share three main principles:

- The role of caching is to deliver part of the content locally.
- Users with statistically identical demands have the same cache contents.
- For isolated private caches, each user can only derive a caching gain from its own cache.

As we will see next, these three main principles, which are sensible for single-cache systems, do not carry over to networks of caches. Indeed, we argue that the role of caching goes well beyond local delivery and that local delivery only achieves a small fraction of the gain that cache networks can offer. We explain the main idea with two toy examples from [2].

Example 2 (*Coded Caching* $K = N = 3$, $M = 1$)**.** Consider a system with $K = 3$ users, each with a cache large enough to store one file, i.e., $M = 1$. Assume that the server has $N = 3$ files, A, B, and C. We split each file into three subfiles of equal size, i.e., $A = (A_1, A_2, A_3)$, $B = (B_1, B_2, B_3)$, and $C = (C_1, C_2, C_3)$. In the placement phase, instead of placing the same content in all caches, we place different content pieces at the users' caches as shown in Fig. 2. Formally, the cache of user k is populated with (A_k, B_k, C_k). Since the size of each subfile has $1/3$ of the size of a whole file, the size of (A_k, B_k, C_k) is equal to one file, satisfying the memory constraint of $M = 1$.

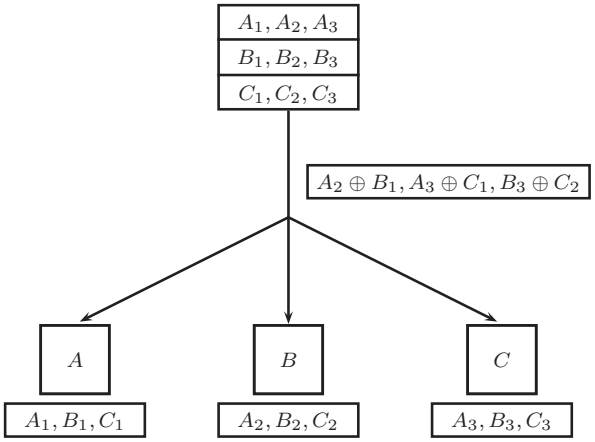


Fig. 2. Coded caching strategy for $K = 3$ users, $N = 3$ files, and cache size $M = 1$. Each file is split into three subfiles of size $1/3$, e.g., $A = (A_1, A_2, A_3)$. The content placement is not a function of the demands. The delivery phase uses coding to satisfy two user demands with a single transmission.

For the delivery phase, let us consider a generic case in which user one requests file A, user two requests file B, and user three requests file C. Then the missing subfiles are A_2 and A_3 for user one, B_1 and B_3 for user two, and C_1 and C_2 for user three. In other words, similar to the uncoded approach, $1/3$ of a user's requested file is available in that user's private cache and can therefore be delivered locally. The server could now transmit the remaining 6 subfiles, each with size of $1/3$, for a total rate of 2. This would be the same rate as for the uncoded scheme in Example 1. However, as we will see next, making use of the particular pattern of the content placement helps us achieve a better rate.

Note that user two has access to A_2, which user one needs, and user one has access to B_1, which user two needs. These two users would like to exchange this side information but cannot since their caches are isolated. Instead the server can exploit this situation by transmitting $A_2 \oplus B_1$ over the shared link, where $\oplus$ denotes bitwise XOR. Since user one already has B_1 from its local cache, it can recover A_2 from $A_2 \oplus B_1$. Similarly, since user two already has access to A_2, it can recover B_1 from $A_2 \oplus B_1$. Thus, the signal $A_2 \oplus B_1$ received over the shared link helps both users to effectively exchange the missing subfiles available in the cache of the other user. Similarly, the server transmits $A_3 \oplus C_1$ over the shared link to deliver A_3 to user one and C_1 to user three. Finally, the server transmits $B_3 \oplus C_2$ to deliver B_3 to user two and C_2 to user three as shown in Fig. 2. Since each server transmission is simultaneously useful for two users, the load of the shared link is reduced by a factor 2 compared to the uncoded approach. The resulting delivery rate is equal to 1.

Here we have focused on the demand tuple (A, B, C). It is straightforward to verify that the same rate is also achievable for all other 26 possible demand tuples. ◊

The above example highlights that, in addition to local delivery of content, caching offers another benefit. The content placed into the caches creates multicasting opportunities through coding that can further reduce the rate over the shared link compared to the uncoded scheme of Example 1. Both schemes enjoy the gain of local delivery as $1/3$ of the content is delivered locally, but the coded scheme enjoys an additional gain of a factor 2 due to coded multicasting.

How do these two caching gains, i.e., the gain of local delivery and the gain of coded multicasting, scale with the parameters of the problem? To get some insight, we increase the size of the cache from $M = 1$ to $M = 2$ and see how these two gains change.

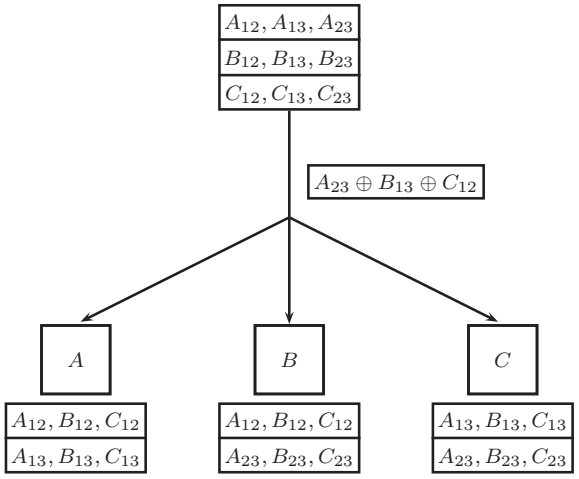


Fig. 3. Coded caching strategy for $K = 3$ users, $N = 3$ files, and cache size $M = 2$. Each file is split into three subfiles of size $1/3$, e.g., $A = (A_1, A_2, A_3)$. Here, the delivery phase uses coding to satisfy three user demands with a single transmission.

Example 3 (*Coded Caching* $K = N = 3$, $M = 2$)**.** In the placement phase, we again split each file into three subfiles of equal size. However, it will be convenient to label these subfiles differently, namely $A = (A_{12}, A_{13}, A_{23})$, $B = (B_{12}, B_{13}, B_{23})$, and $C = (C_{12}, C_{13}, C_{23})$. User k caches those content pieces that have k in the index set as shown in Fig. 3

For the delivery phase, let us again assume as an example that user one requests file A, user two requests file B, and user three requests file C (see again Fig. 3). In this case, each user can fetch $2/3$ of its requested file from the local cache and misses the remaining $1/3$ of the file. In particular, user one misses subfile A_{23}, which is available at both users two and three. User two misses subfile B_{13}, which is available at both users one and three. And user three misses subfile C_{12}, which is available at both users one and two. In other words, the three users would like to exchange the subfiles A_{23}, B_{13}, C_{12}, but are unable to do so because their caches are isolated. The server can remedy this situation by transmitting the signal $A_{23} \oplus B_{13} \oplus C_{12}$ over the shared link. Given its cache content, each user can then recover the missing subfile. Since the coded transmission is simultaneously useful for all three users, the coded caching approach reduces the load of the shared link by a factor of 3 compared to the uncoded scheme of Example 1, resulting in a rate of $1/3$. All other 26 possible requests can be satisfied in a similar manner. ◊

From the last example we see that, as we increase the size of the cache, both the local gain and the coded multicasting

gain improve. For the general case, it is shown in [2] that for arbitrary number N of files and $K \leq N$ users each with cache of size $M \in \{0, N/K, 2N/K, \ldots, N\}$, coded caching achieves a rate of

$$R_C(M) \triangleq K \cdot (1 - M/N) \cdot \frac{1}{1 + KM/N}. \qquad (2)$$

For general $0 \leq M \leq N$, the lower convex envelope of these points is achievable. The case $K > N$ can be handled similarly, but the resulting expression is a bit more complicated (see [2]). The function $R_C(M)$ describes the trade-off between rate and memory for the coded caching scheme.

We compare the three terms in the rate expression $R_C(M)$ in (2) achieved by coded caching with the two terms in the rate expression $R_U(M)$ in (1) achieved by uncoded caching.

- The first term K, representing the rate without caching, is the same in both rate expressions.
- The second term $1-M/N$, representing the local caching gain, is also the same in both rate expressions. Thus, both the coded and uncoded schemes enjoy the gain from having a fraction M/N of each file being locally available.
- On top of this, the coded scheme alone enjoys a second gain that is absent in the uncoded scheme. This gain is quantified by the extra factor $\frac{1}{1+KM/N}$, which captures the gain resulting from creating and exploiting coded multicasting opportunities. Perhaps surprisingly, we see that this gain is a function of the *cumulative* memory size, i.e., KM, even though the caches are isolated. We refer to this gain as the *global caching gain*. To attain this gain, we follow a particular pattern of content placement. In the delivery phase, this pattern allows the creation of coded packets each useful for $1 + KM/N$ users. This coded multicasting opportunity is available simultaneously for every one of the N^K possible set of user demands, i.e., it provides a *simultaneous coded multicasting opportunity*.

We next compare the two caching gains in more detail.

- The local caching gain $1 - M/N$ is significant if the *local* cache size M is comparable to the size of the entire content N.
- The global caching gain $\frac{1}{1+KM/N}$ is significant if the *cumulative* cache size KM is comparable to the size of the entire content N. As a result, the global caching gain can reduce the load of the shared link in the order of the number of caches K in the system.

Thus we see that, for networks of caches, the global gain can be much more important than the local gain.

The order difference between the local and global gains is illustrated in Fig. 4 for a system with $K = 30$ users. For example, if each user has space to cache half of the content, then uncoded caching reduces the load of the shared link from 30 files down to the equivalent of 15 files. On the other hand, coded caching reduces the load of the shared link to less than the equivalent of just a single file.

It can be shown that the rate $R_C(M)$ of the coded caching scheme is within a constant factor of the information-theoretic optimum for all values of the problem parameters [2]. This

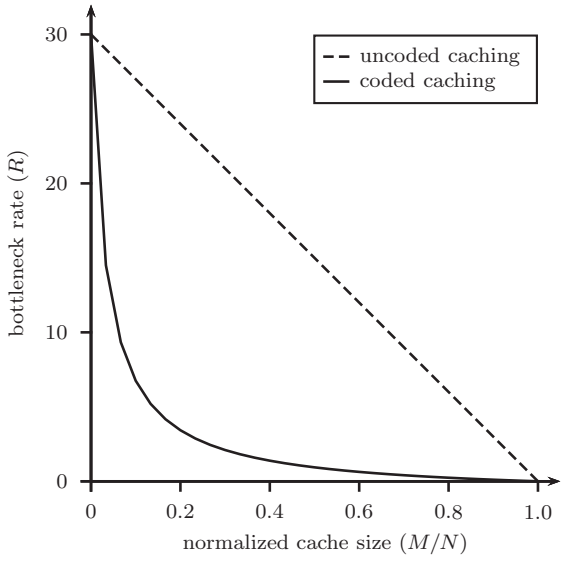


Fig. 4. Rate R required in the delivery phase as a function of normalized memory size M/N for $K = 30$ users from [2]. The figure compares the performance of the proposed coded caching scheme with that of conventional uncoded caching.

implies that the local and global gains identified above are *fundamental*, i.e., there are no other gains that scale with the system parameters.

Open Problem 1*:* Sharpening the approximation of the rate-memory trade-off is of both theoretical and practical interest. It is known that both the achievable scheme and the converse can be improved [2]. For achievability, the first question is if linear codes are sufficient for optimality or if nonlinear codes are needed. The second question is if, within the class of linear codes, larger field sizes can improve the performance. Finally, the content placement presented so far is uncoded and only the delivery is coded. It is known that coded content placement can improve system performance for small cache sizes [2]. Whether coded content placement can increase performance for larger cache sizes as well is unknown. There have also been some recent efforts to improve the converse part [4], [5].

Open Problem 2*:* The rate-memory trade-off is known exactly for a system with $K = 2$ users and $N = 2$ files [2]. Finding the exact trade-off for $K = 3$ and $N = 3$, the next-bigger case, is of interest. There has been some recent progress in this direction, and for some values of cache size M the optimal trade-off is known [6]. However, for general M, the $K = 3$ and $N = 3$ case is still open.

III. Other Service Requirements

Practical applications and constraints may necessitate different service requirements than the ones in the canonical model. We next discuss several of those requirements.

A. Decentralized Caching

In the canonical cache network both the number and the identity of the users in the delivery phase are already known in the prior placement phase. This is clearly not a realistic

assumption, because we would likely be unaware during the placement phase, say in the early morning, which users will be active in the following evening. In addition, users may join or leave the network asynchronously, so that the number of users in the delivery phase may also be time varying.

To deal with these issues, [7] develops a decentralized caching scheme, in which the placement phase is independent of the number and the identity of the users. In this scheme, cache stores a randomly selected subset of the bits. The rate of this decentralized scheme is shown to be within a constant factor of optimal universally for any number of users K. This universality property allows to address the problem of asynchronous user requests. In addition, this decentralized caching scheme is a key ingredient to handle online and nonuniform demands discussed below.

Open Problem 3*:* It is shown analytically in [7] that the rate of the decentralized caching scheme is within a constant factor of the centralized scheme. Numerically, this factor can be evaluated to be 1.6. This shows that there is at most a small price to be paid for a placement phase that is universal with respect to the number of users K. It is of interest to know if there is, in fact, a cost for this universality at all.

B. Nonuniform Demands

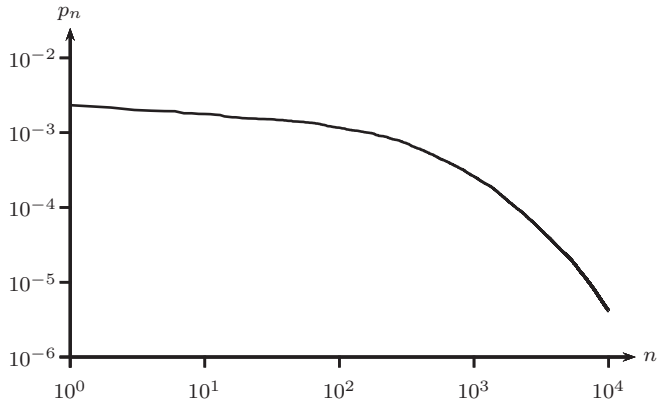


Fig. 5. File popularities p_n for the Netflix movie catalog from [8].

The canonical cache network focuses on the peak rate over the shared link, i.e., the rate for the worst user demands. In practice, the content files have different popularities, modeled as the probabilities of being requested by the users (see Fig. 5). Consequently, in some settings a more natural performance criterion is the expected rate over the shared link.

If the file popularity is uniform, the coded caching scheme from Section II also approximately minimizes the expected rate (as opposed to peak rate) [8]. For nonuniform popularity distributions, a different approach is needed. For such nonuniform distributions, [8] suggests to split the content files into several groups and to dedicate a fraction of the cache memory at each user to each group. The placement phase and delivery phase of the decentralized coded caching scheme are then applied within each group of files. Since the number and identity of users requesting files from each group is only known during the delivery phase but not during the placement phase, the universality of the decentralized caching scheme is critical for this file-grouping approach to work.

Subsequently, [9] proposed to use only two such file groups with all memory dedicated to the first group. [9] also showed that this approach is asymptotically within a constant factor from optimal for the important special case of Zipf popularity distributions in the limit as $K, N \to \infty$. Finally, [10] showed that this approach with only two groups is in fact optimal to within a constant multiplicative-plus-additive gap for all popularity distributions and all finite values of K and N (assuming $M \geq 2$). These two results thus show that, surprisingly, two groups are sufficient to adapt to the nonuniform nature of the popularity distribution.

This conclusion changes when, instead of a single user per cache, many users are attached to each cache. In this scenario, the grouping strategy with many groups is approximately optimal [11].

C. Online Caching

The canonical caching problem in Section II has two distinct phases: placement and delivery. The cache is updated only during the placement phase, but not during the delivery phase. In other words, caching is performed offline, meaning ahead of delivery.

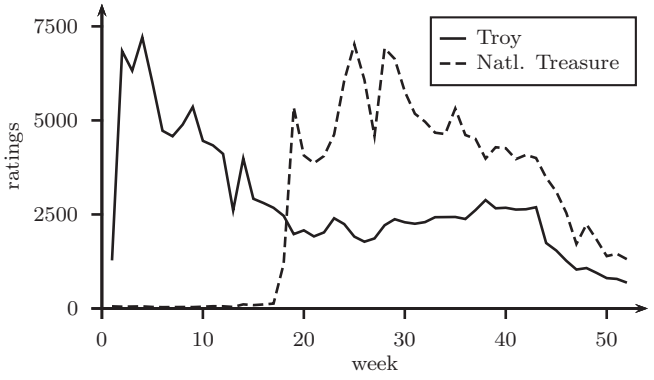


Fig. 6. Number of ratings in the Netflix database for two movies ("Troy" and "National Treasure") as a function of week in 2005 from [12]. Each movie was very popular upon release and then gradually reduced its popularity thereafter.

However, in many practical systems, the set of popular files is constantly changing. Some new popular files can be added to the content database, and some old files can become unpopular or be removed from the content database (see Fig. 6). In order to adapt to this dynamic content popularity, caching schemes that update their cache content online, i.e., during the delivery phase, are needed.

One popular cache update rule is least-recently used (better known by its abbreviation LRU), in which the least-recently requested file is evicted from the cache to open up space for a newly requested file. While LRU is proven to be efficient for single-cache systems [13], it is shown in [12] that for cache networks it can be significantly suboptimal. Instead, a coded version of LRU, in which the caches are updated during the delivery phase such as to preserve the coding gain, is proposed. For a probabilistic model of request dynamics, this update rule is shown to be approximately optimal in [12].

Open Problem 4*:* The approximate optimality result in [12] holds only under a probabilistic model of request dynamics. An open question is to develop schemes that have stronger competitive optimality guarantees valid for any individual sequence of users' requests as shown in [13] for the single-cache setting.

D. Delay-Limited Content

Fig. 7. Screenshot of video-streaming demo from [14]. The lower left window shows the server process. The upper left window shows the decoding process at a local cache. The three windows on the right show the reconstructed videos being played in real time.

Video streaming is a popular application for caching. In this setting each user sequentially requests small chunks of content. Each such chunk has to be delivered within a limited delay in order to enable continuous playback at the user. Thus the server can only exploit coding opportunities among the requested chunks within a given time window. In such scenarios, the ultimate gain of coded caching, as seen in the analysis of the canonical caching problem, is achievable only if the tolerable delay is very large. [14] investigates the trade-off between the performance of coded caching and delay tolerance, and proposes a computationally efficient, coded caching scheme that respects the delay constraint. This approach was demonstrated in a practical setting with a video-streaming prototype (see Fig. 7). The same approach also works for settings with small files.

Open Problem 5*:* Approximately characterizing the fundamental trade-off between the rate versus cache size under a delay constraint is of great interest.

Open Problem 6*:* The demo in [14] works for a small number of caches and users. Scaling the system up to say 100 caches with 100 users per cache is of interest. This will require addressing a significant number of systems issues such as how to maintain state and how to handle disk reads both at the server and at the caches, among others.

IV. Other Network and Channel Models

The canonical cache network has a noiseless broadcast channel topology. Here, we discuss other network and channel models.

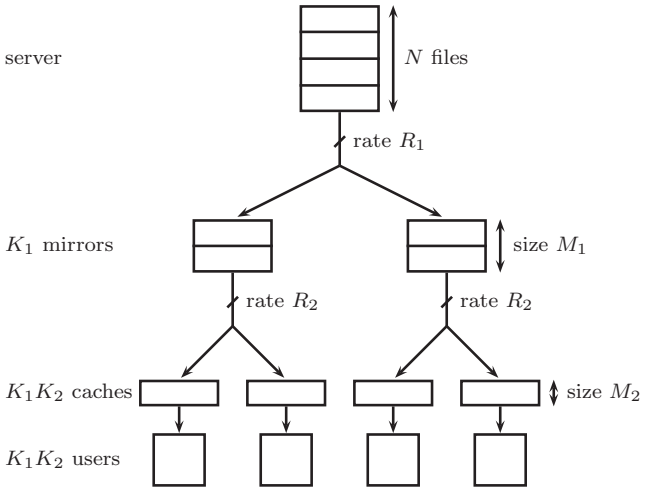


Fig. 8. System setup for the hierarchical caching problem from [15].

A. Other Network Topologies

In practice, many caching systems consist of not only one, but multiple layers of caches connected to each other to from a tree. The objective is to minimize the transmission rates in the various layers. [15] models this scenario as the network shown in Fig. 8 and approximately characterizes the rate-memory trade-off.

A different generalization of the broadcast topology is to allow each user to connect to several close-by caches. This scenario, particularly relevant for mobile users with caches located at femtocells, is analyzed in [16].

Scenarios with multiple servers have been considered in [17]. Coded caching for the device-to-device communication setting, where users help each other to deliver content, has been analyzed in [18].

Another topology arising in the context of distributed computation has been analyzed in [19]. This network topology models a data center with multiple servers, each performing part of a larger MapReduce job. Here the repetition in map assignments is used to create coding opportunities and to reduce the communication load of the shuffling phase.

Open Problem 7*:* An interesting open problem is to characterize the rate-memory trade-off for hierarchical cache networks with multiple levels within a constant factor independent of the number of levels.

Open Problem 8*:* Devising easily implementable and efficient algorithms for hierarchical cache networks with nonuniform file popularities and online cache updating is of practical interest.

Open Problem 9*:* Developing caching strategies with some optimality guarantee for general network topologies is a likely difficult but interesting open problem.

B. Noisy Channels

The noisy version of the noiseless broadcast channel in the canonical cache network is considered in [20]. Here, the noise is modeled as an erasure broadcast channel. The setting is

particularly interesting for asymmetric erasure probabilities, where unequal cache sizes can be used to improve system performance. A similar setting but with feedback was analyzed in [21].

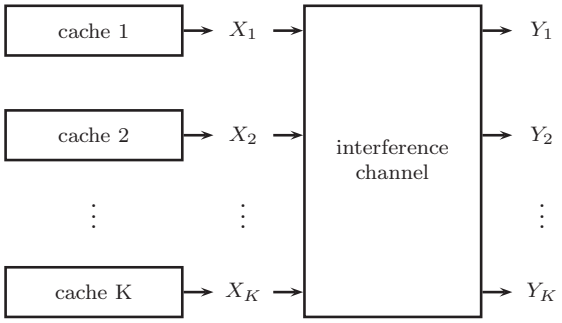


Fig. 9. K transmitters, each connected to a local cache, communicating to K receivers over a Gaussian interference channel from [22].

To reduce load and delay of the backhaul, cache-aided cellular base stations have received considerable attention [23]–[25]. This raises the question if caches at the base stations can also improve communication rate over the wireless links. This question is investigated in [22] using the interference channel model depicted in Fig. 9. It is shown that there are three distinct gains from caching at the transmitters of an interference channel: a load balancing gain, an interference cancellation gain, and an interference alignment gain. The load balancing gain is achieved through specific file placement, creating a particular pattern of content overlap in the caches. This overlap also enables interference cancellation through transmitters' cooperation. Finally, the cooperation among transmitters creates many virtual transmitters, which in turn increases interference alignment possibilities.

Open Problem 10*:* The rate-memory trade-off for cache-aided interference channels is still unknown. Even characterizing the degrees-of-freedom version of this trade-off is open.

Open Problem 11*:* Many multi-user channels could have a cache-aided version, where caches can be at the transmitters' side or at the receivers' side or both. Cataloguing what type of gains (similar to the coded multicasting, load balancing, interference cancellation, and alignment gains seen so far) caching can provide in these settings will be useful to guide the design and operation of noisy cache networks.

V. Connection with Network and Index Coding

Having surveyed the coded caching problem for various network topologies and service requirements, we now return to the basic canonical cache network and explore its connection to network and index coding.

The canonical caching problem is related to the network coding problem [26]. Indeed, the canonical cache network with K users and N files can be expressed as a single-source multiple-multicast problem with KN^K sinks and N multicast groups (see Fig. 10). Unlike the single-source *single*-multicast problem, the single-source *multiple*-multicast problem is a hard problem in general [27]. It is the special structure of

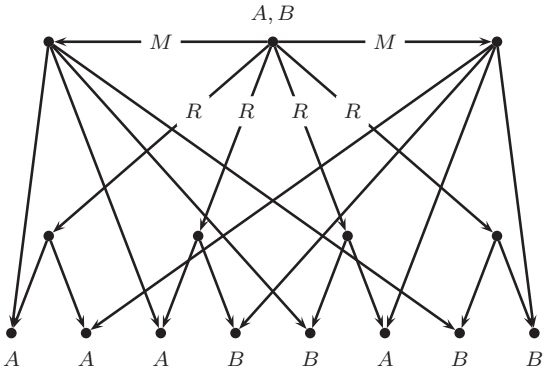


Fig. 10. The $K = 2$-user, $N = 2$-file canonical cache network expressed as a single-source multiple-multicast network coding problem.

this network coding problem induced by the caching setting that allows for the constant-factor approximation in [2].

The canonical caching problem is also related to the index coding problem [28], [29]. Consider again the canonical cache network with K users and N files. Then for *fixed* and *uncoded* cache content chosen in the placement phase and for *fixed* user demands, the delivery phase of the caching problem is exactly a K-user index coding problem. Since there are N^K possible user demands, the complete delivery phase consists of N^K parallel such index coding problems. Unfortunately, the general index coding problem is hard to solve even approximately [30]. The main difference with the canonical caching problem is that here we are tasked with also designing the side information (which may not be uncoded) subject to a memory constraint. In other words, instead of fixed side information as in index coding, we have a budget for side information. Moreover, we have to be able to handle any possible user demands. Interestingly, it is exactly this additional freedom to design the side information that renders the canonical caching problem more tractable.

VI. Concluding Remarks

In this newsletter article, we have argued that information theory can play an important role in providing a fundamental understanding of how to design and operate cache networks. Many open questions remain to complete this understanding, and we have pointed out a number of them.

References

[1] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*. Pearson, sixth ed., 2012.
[2] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, pp. 2856–2867, May 2014.
[3] A. Silberschatz, P. B. Galvin, and G. Gagne, *Operating System Concepts*. Wiley, eighth ed., 2008.
[4] A. Sengupta and R. Tandon, "Improved approximation of storage-rate tradeoff for caching via new outer bounds," in *Proc. IEEE ISIT*, June 2015.
[5] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," in *Proc. IEEE ISIT*, June 2015.
[6] C. Tian, "A note on the fundamental limits of coded caching," *arXiv:1503.00010 [cs.IT]*, Feb. 2015.
[7] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, pp. 1029–1040, Aug. 2015.

[8] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," *arXiv:1308.0178 [cs.IT]*, Aug. 2013.
[9] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands," *arXiv:1502.03124 [cs.IT]*, Feb. 2015.
[10] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," in *Proc. ITA*, Feb. 2015.
[11] J. Hachem, N. Karamchandani, and S. Diggavi, "Effect of number of users in multi-level coded caching," in *Proc. IEEE ISIT*, June 2015.
[12] R. Pedarsani, M. A. Maddah-Ali, and U. Niesen, "Online coded caching," *arXiv:1311.3646 [cs.IT]*, Nov. 2013. To appear in IEEE/ACM Trans. Netw.
[13] D. D. Sleator and R. E. Tarjan, "Amortized efficiency of list update and paging rules," *Communications ACM*, vol. 28, pp. 202–208, Feb. 1985.
[14] U. Niesen and M. A. Maddah-Ali, "Coded caching for delay-sensitive content," in *Proc. IEEE ICC*, June 2015.
[15] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. Diggavi, "Hierarchical coded caching," *arXiv:1403.7007 [cs.IT]*, Mar. 2014.
[16] J. Hachem, N. Karamchandani, and S. Diggavi, "Content caching and delivery over heterogeneous wireless networks," *arXiv:1404.6560 [cs.IT]*, Apr. 2014.
[17] S. Shariatpanahi, A. S. Motahari, and B. H. Khalaj, "Multi-server coded caching," *arXiv:1503.00265 [cs.IT]*, Mar. 2015.
[18] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless D2D networks," *arXiv:1405.5336 [cs.IT]*, May 2014.
[19] S. Li, M. A. Maddah-Ali, and S. Avestimehr, "Coded MapReduce," in *Proc. Allerton Conf.*, Sept. 2015.
[20] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," *arXiv:1505.01016 [cs.IT]*, May 2015.
[21] A. Ghorbel, M. Kobayashi, and S. Yang, "Cache-enabled broadcast packet erasure channels with state feedback," *arXiv:1509.02074 [cs.IT]*, Sept. 2015.
[22] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *Proc. IEEE ISIT*, June 2015.
[23] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, pp. 1107–1115, Mar. 2012.
[24] A. Liu and V. K. N. Lau, "Exploiting base station caching in MIMO cellular networks: Opportunistic cooperation for video streaming," *IEEE Trans. Signal Process.*, vol. 63, pp. 57–69, Jan. 2015.
[25] K. Poularakis, G. Iosifidis, and L. Tassiulas, "Approximation algorithms for mobile data caching in small cell networks," *IEEE Trans. Commun.*, vol. 62, pp. 3665–3677, Oct. 2014.
[26] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1204–1216, Apr. 2000.
[27] A. R. Lehman and E. Lehman, "Complexity classification of network information flow networks," in *Proc. ACM-SIAM SODA*, pp. 142–150, Jan. 2004.
[28] Y. Birk and T. Kol, "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2825–2830, June 2006.
[29] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Proc. IEEE FOCS*, pp. 197–206, Oct. 2006.
[30] M. Langberg and A. Sprintson, "On the hardness of approximating the network coding capacity," *IEEE Trans. Inf. Theory*, vol. 57, pp. 1008–1014, Feb. 2011.

From the Editor *continued from page 2*

Suhas Diggavi, Vijay Kumar, Pierre Moulin, David Tse, and Raymond Yeung have prepared a report summarizing the new initiatives/experiments conducted in the organization of our flagship conference ISIT that took place over the summer. Georg Bcherer, Gianluigi Liva, and Gerhard Kramer have prepared a report on the Munich Workshop on Coding and Modulation (MCM 2015). Also from Munich, Stefan Dierks, Markus Jger, Gerhard Kramer, and Roy Timo have prepared a report on the Munich Workshop on Massive MIMO (MMM 2015). Vincent Tan, Matthieu Bloch, and Merouane Debbah report on the Mathematical Tools of Information-Theoretic Security Workshop that took place recently in the Huawei Mathematical and Algorithmic Sciences Lab, Paris, France. Many thanks for all the contributors for their efforts!

With sadness, we conclude this issue with tributes to two prominent members of our community, Oscar Moreno de Ayala who passed away on July 14, and Victor Wei who passed away on October 17th. Thanks to Heeralal Janwa, P. Vijay Kumar and Andrew Z. Tirkel; and to Lolita Chuang, Yu Hen Hu, Yih-Fang Huang, and Ming-Ting Sun for preparing the tributes.

Please help to make the newsletter as interesting and informative as possible by sharing with me any ideas, initiatives, or potential newsletter contributions you may have in mind. I am in the process of searching for contributions outside our community, which may introduce our readers to new and exciting problems and, in such, broaden the influence of our society. Any ideas along this line will also be very welcome.

Announcements, news and events intended for both the printed newsletter and the website, such as award announcements, calls for nominations and upcoming conferences, can be submitted at the IT Society website http://www.itsoc.org. Articles and columns can be e-mailed to me at mikel@buffalo.edu with a subject line that includes the words IT newsletter.

The next few deadlines are: January 10, 2016 for the issue of March 2016. April 10, 2016 for the issue of June 2016.

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

I look forward to hearing your suggestions and contributions.

With best wishes,
Michael Langberg.

Concentration of Measure Inequalities and Their Communication and Information-Theoretic Applications

Maxim Raginsky and Igal Sason

***Abstract*—During the last two decades, concentration of measure has been a subject of various exciting developments in convex geometry, functional analysis, statistical physics, high-dimensional statistics, probability theory, information theory, communications and coding theory, computer science, and learning theory. One common theme which emerges in these fields is probabilistic stability: complicated, nonlinear functions of a large number of independent or weakly dependent random variables often tend to concentrate sharply around their expected values. Information theory plays a key role in the derivation of concentration inequalities. Indeed, both the entropy method and the approach based on transportation-cost inequalities are two major information-theoretic paths toward proving concentration.**

This brief survey is based on a recent monograph of the authors in the *Foundations and Trends in Communications and Information Theory*, and a tutorial given by the authors at ISIT 2015. It introduces information theorists to three main techniques for deriving concentration inequalities: the martingale method, the entropy method, and the transportation-cost inequalities. Some applications in information theory, communications, and coding theory are used to illustrate the main ideas.

I. Introduction

Concentration inequalities bound from above the probability that a random variable Z deviates from its mean, median or some other typical value by a given amount. These inequalities have been studied for several decades, with some fundamental and substantial contributions during the last two decades. Very roughly speaking, the concentration-of-measure phenomenon can be stated in the following simple way: "A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant" [1]. Informally, this amounts to saying that such a random variable Z concentrates around its expected value, $\mathbb{E}[Z]$, in such a way that the probability of the event $\{|Z-\mathbb{E}[Z]| \geq t\}$, for a given $t > 0$, decays exponentially in some power of t. Detailed treatments of the concentration-of-measure phenomenon, including historical accounts, can be found, e.g., in [2]–[9].

In recent years, concentration inequalities have been intensively studied and used as a powerful tool in various areas. These include convex geometry, functional analysis, statistical physics, probability theory, statistics, information theory, communications and coding theory, learning theory, and computer science. Several techniques have been developed so far to prove concentration inequalities. This survey paper focuses on three such techniques which are studied in our tutorial [9] and references therein:

- The martingale method (see, e.g., [6], [10], [11], [8, Chapter 7], [12], [13]), and its information-theoretic applications (see, e.g., [14] and references therein, [15]).
- The entropy method and logarithmic Sobolev inequalities (see, e.g., [3, Chapter 5], [4] and references therein).
- Transportation-cost inequalities which originated from information theory (see, e.g., [3, Chapter 6], [16], [17] and references therein).

Our goal here is to give the reader a quick preview of the vast field of concentration inequalities and their applications in information theory, communications and coding. Therefore, we state most of the theorems and lemmas without proofs; occasionally, we provide sketches or brief outlines. More details can be found in our monograph [9] and the slides of our ISIT'15 tutorial.[1]

II. The Basic Toolbox

Our objective is to derive tight upper bounds on the tail probabilities

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + t] \text{ and } \mathbb{P}[Z \leq \mathbb{E}[Z] - t], \qquad \forall t > 0$$

where $Z = f(X_1, \ldots, X_n)$ is an arbitrary function of n independent random variables $X_1, \ldots, X_n$. To get an idea of what we can expect, let us first recall Chebyshev's inequality:

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \geq t] \leq \frac{\mathrm{Var}[Z]}{t^2}, \qquad \forall t > 0.$$

This inequality shows that the tail probability decays with t, and that the rate of decay is proportional to the variance of Z. Thus, the variance of Z gives an idea about how tightly Z concentrates around its mean. In fact, if Z takes values in a bounded interval, then we can upper-bound the variance of Z only in terms of the length of this interval:

Lemma 1. *Let Z be a random variable taking values in an interval $[a, b]$. Then*

$$\mathrm{Var}[Z] \leq \tfrac{1}{4}(b-a)^2. \tag{1}$$

This bound is sharp: if Z only takes the two values a and b with equal probability, then $\mathrm{Var}[Z] = \frac{1}{4}(b-a)^2$.

Maxim Raginsky is with Department of Electrical and Computer Engineering, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (e-mail: maxim@illinois.edu).

I. Sason is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: sason@ee.technion.ac.il).

[1] Part 1 (The martingale method): http://webee.technion.ac.il/people/sason/raginsky_sason_ISIT_2015_tutorial_part_1.pdf.

Part 2 (The entropy method and transportation-cost inequalities): http://webee.technion.ac.il/people/sason/raginsky_sason_ISIT_2015_tutorial_part_2.pdf.

Proof: Recall that $\mathrm{Var}[Z] \le \mathbb{E}[(Z-c)^2]$ for all $c \in \mathbb{R}$. Letting $c = \frac{a+b}{2}$, we obtain (1). The case of equality is an easy calculation. ∎

Thus, for a bounded Z in an interval $[a,b]$, Chebyshev's inequality gives

$$\mathbb{P}[|Z - \mathbb{E}[Z]| \ge t] \le \frac{(b-a)^2}{4t^2}.$$

Much stronger concentration inequalities can be derived, however, for bounded random variables. Using Markov's inequality, for every $\lambda > 0$ we have

$$\begin{aligned}\mathbb{P}\left[Z - \mathbb{E}[Z] \ge t\right] &= \mathbb{P}\left[e^{\lambda(Z-\mathbb{E}[Z])} \ge e^{\lambda t}\right] \\ &\le e^{-(\lambda t - \psi(\lambda))},\end{aligned}$$

where $\psi(\lambda) \triangleq \log \mathbb{E}[e^{\lambda(Z-\mathbb{E}[Z])}]$ is the *logarithmic moment-generating function* of Z. Optimizing over λ, we get the *Chernoff bound*

$$\mathbb{P}\left[Z \ge \mathbb{E}[Z] + t\right] \le e^{-\psi^\star(t)},$$

where $\psi^\star(t) \triangleq \sup_{\lambda \ge 0} [\lambda t - \psi(\lambda)]$ is the Legendre dual of ψ. For example, if $Z \sim N(0,\sigma^2)$ (Gaussian with mean 0 and variance σ^2), we have $\psi(\lambda) = \lambda^2\sigma^2/2$, and $\psi^\star(t) = t^2/2\sigma^2$. With this in mind, we say that a random variable Z is σ^2-*subgaussian* if $\psi(\lambda) \le \lambda^2\sigma^2/2$. For a subgaussian random variable, we obtain $\psi^\star(t) \ge t^2/2\sigma^2$, which gives the tail bound

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + t] \le e^{-t^2/2\sigma^2}, \qquad \forall t > 0.$$

Thus, the whole affair hinges on our ability to prove that the random variable Z of interest is subgaussian.

To start with, a bounded random variable is subgaussian:

Lemma 2 (Hoeffding [11])**.** *For a random variable Z taking values in an interval $[a,b]$, we have*

$$\log \mathbb{E}[e^{\lambda(Z-\mathbb{E}[Z])}] \le \tfrac{1}{8}\lambda^2(b-a)^2. \tag{2}$$

Proof: We give a simple probabilistic proof, which has the additional benefit of highlighting the role of the tilted distribution. Let $P = \mathcal{L}(Z)$,[2] and introduce its *exponential tilting* $P^{(t)}$: for an arbitrary sufficiently regular function $f\colon \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}_{P^{(t)}}[f(Z)] \triangleq \frac{\mathbb{E}_P[f(Z)e^{tZ}]}{\mathbb{E}_P[e^{tZ}]}.$$

Since Z is supported on $[a,b]$ under P, the same holds under $P^{(t)}$ as well. Therefore, by Lemma 1,

$$\mathrm{Var}_{P^{(t)}}[Z] \le \tfrac{1}{4}(b-a)^2.$$

On the other hand,

$$\begin{aligned}\mathrm{Var}_{P^{(t)}}[Z] &= \frac{\mathbb{E}_P[Z^2e^{tZ}]}{\mathbb{E}_P[e^{tZ}]} - \left(\frac{\mathbb{E}_P[Ze^{tZ}]}{\mathbb{E}_P[e^{tZ}]}\right)^2 \\ &= \psi''(t).\end{aligned}$$

Therefore,

$$\psi''(t) \le \tfrac{1}{4}(b-a)^2$$

for all t. Integrating and using the fact that

$$\psi(0) = \psi'(0) = 0,$$

we get (2). ∎

Both the martingale method and the entropy method are just elaborations of these basic tools, which are applicable to an arbitrary bounded real-valued random variable. However, one should keep in mind that concentration of measure is a *high-dimensional* phenomenon: we are interested in situations when Z is a function of many independent random variables $X_1, \ldots, X_n$, and we can often quantify the "sensitivity" of f to changes in each of its arguments while the others are kept fixed. This suggests that we may get a handle on the high-dimensional concentration properties of Z by breaking up the problem into n one-dimensional subproblems involving only one of the X_i's at a time. Whenever such a divide-and-conquer approach is possible, we speak of *tensorization*, by which we mean that some quantity involving the distribution of

$$Z = f(X_1, \ldots, X_n)$$

(e.g., variance or relative entropy) can be related to the sum of similar quantities involving the *conditional* distribution of each X_i given

$$\overline{X}^i \triangleq (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

III. The Martingale Method

The basic idea behind the martingale method is to start with the *Doob martingale decomposition*

$$Z - \mathbb{E}[Z] = \sum_{k=1}^n \xi_k, \tag{3}$$

where

$$\xi_k \triangleq \mathbb{E}[Z|X^k] - \mathbb{E}[Z|X^{k-1}] \tag{4}$$

with

$$X^k \triangleq (X_1, \ldots, X_k)$$

and then to exploit any information about the sensitivity of f to local changes in its arguments in order to control the sizes of the increments ξ_k. As a warm-up, consider the following inequality, first obtained in a restricted setting by Efron and Stein [18] and generalized by Steele [19]:

Lemma 3 (Efron–Stein–Steele)**.** *Let $Z = f(X^n)$ where $X_1, \ldots, X_n$ are independent, then*

$$\mathrm{Var}[Z] \le \sum_{k=1}^n \mathbb{E}\left[\mathrm{Var}[Z|\overline{X}^k]\right]. \tag{5}$$

Proof: We exploit the fact that $\{\xi_k\}_{k=1}^n$ in (4) is a *martingale difference sequence* with respect to X^n, i.e.,

$$\mathbb{E}[\xi_k|X^{k-1}] = 0 \tag{6}$$

[2] The notation $\mathcal{L}(Z)$ stands for the law, or probability distribution, of the random variable Z.

for all $k \in \{1, \ldots, n\}$. Hence, since $\mathbb{E}[\xi_k \xi_l] = 0$ for $k \neq l$,

$$\mathrm{Var}[Z] = \sum_{k=1}^{n} \mathbb{E}[\xi_k^2]. \tag{7}$$

The independence of $X_1, \ldots, X_n$ in (4) yields

$$\xi_k = \mathbb{E}\big[Z - \mathbb{E}[Z|\overline{X}^k] \,|\, X^k\big]$$

and, from Jensen's inequality,

$$\xi_k^2 \leq \mathbb{E}\big[\big(Z - \mathbb{E}[Z|\overline{X}^k]\big)^2 \,|\, X^k\big].$$

Due to the independence of $X_1, \ldots, X_n$, this in turn yields

$$\begin{aligned}\mathbb{E}[\xi_k^2] &\leq \mathbb{E}[\big(Z - \mathbb{E}[Z|\overline{X}^k]\big)^2] \\ &= \mathbb{E}[\mathrm{Var}[Z|\overline{X}^k]]. \end{aligned} \tag{8}$$

Substituting (8) into (7) yields (5). ■

The Efron–Stein–Steele inequality is our first example of tensorization: it upper-bounds the variance of $Z = f(X_1, \ldots, X_n)$ by the sum of the expected values of the conditional variances of Z given all but one of the variables. In other words, we say that $\mathrm{Var}[f(X_1, \ldots, X_n)]$ tensorizes. This fact has immediate useful consequences. For example, we can use any convenient technique for upper-bounding variances to control each term on the right-hand side of (7), and thus obtain many useful variants of the Efron–Stein–Steele inequality:

1) For every random variable U with a finite second moment,

$$\mathrm{Var}[U] = \tfrac{1}{2}\,\mathbb{E}[(U - U')^2]$$

where U' is an i.i.d. copy of U. Thus, if we let

$$Z_k' = f(X_1, \ldots, X_{k-1}, X_k', X_{k+1}, \ldots, X_n),$$

where X_k' is an i.i.d. copy of X_k, then Z and Z_k' are i.i.d. given $\overline{X}^k$. This implies that

$$\mathrm{Var}[Z|\overline{X}^k] = \tfrac{1}{2}\,\mathbb{E}\Big[(Z - Z_k')^2 \Big| \overline{X}^k\Big]$$

for $k \in \{1, \ldots, n\}$, yielding the following variant of the Efron–Stein–Steele inequality:

$$\mathrm{Var}[Z] \leq \tfrac{1}{2} \sum_{i=1}^{n} \mathbb{E}[(Z - Z_k')^2]. \tag{9}$$

This inequality is sharp: if $Z = \sum_{k=1}^n X_k$, then

$$\mathbb{E}[(Z - Z_k')^2] = 2\,\mathrm{Var}[X_k],$$

and (9) holds with equality. This shows that sums of independent random variables $X_1, \ldots, X_n$ are the least concentrated among all functions of X^n.

2) For every random variable U with a finite second moment and for all $c \in \mathbb{R}$,

$$\mathrm{Var}[U] \leq \mathbb{E}[(U - c)^2].$$

Thus, if we condition on $\overline{X}^k$, we can let $Z_k = f_k(\overline{X}^k)$ for arbitrary functions f_k ($k \in \{1, \ldots, n\}$) of $n - 1$ variables, and obtain another variant of the Efron–Stein–Steele inequality:

$$\mathrm{Var}[Z] \leq \sum_{i=1}^{n} \mathbb{E}[(Z - Z_k)^2]. \tag{10}$$

3) Suppose we know that, by varying just one of the arguments of f while holding all others fixed, we cannot change the value of f by more than some bounded amount. More precisely, suppose that there exist finite constants $c_1, \ldots, c_n \geq 0$, such that

$$\begin{aligned}&\sup_x f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \\ &- \inf_x f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \leq c_i \end{aligned} \tag{11}$$

for all i and all $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. Then

$$\mathrm{Var}[Z|\overline{X}^k] \leq \tfrac{1}{4}\, c_k^2$$

by Lemma 1, and therefore from (5), (8)

$$\mathrm{Var}[Z] \leq \tfrac{1}{4} \sum_{k=1}^{n} c_k^2. \tag{12}$$

Example: Kernel Density Estimation

As an example of Efron–Stein–Steele inequalities in action, let us look at *kernel density estimation* (KDE), a nonparametric procedure for estimating an unknown pdf ϕ of a real-valued random variable X based on observing n i.i.d. samples $X_1, \ldots, X_n$ drawn from ϕ [20, Chap. 9]. A *kernel* is a function $K \colon \mathbb{R} \to \mathbb{R}^+$ satisfying the following conditions:

1) It is integrable and normalized: $\int_{-\infty}^{\infty} K(u)\mathrm{d}u = 1$.
2) It is even: $K(u) = K(-u)$ for all $u \in \mathbb{R}$.
3) $\lim_{h \downarrow 0} \frac{1}{h} K(\frac{x-u}{h}) = \delta(x - u)$, where δ is the Dirac function.

The KDE is given by

$$\phi_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where $h > 0$ is a parameter called the *bandwidth*. From the properties of K, for each $x \in \mathbb{R}$ we have

$$\mathbb{E}[\phi_n(x)] = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x-u}{h}\right) \phi(u)\mathrm{d}u \xrightarrow{h \downarrow 0} \phi(x).$$

Thus, we expect the KDE ϕ_n to concentrate around the true pdf ϕ; to quantify this, let us examine the L_1 error

$$Z_n = f(X_1, \ldots, X_n) = \int_{-\infty}^{\infty} |\phi_n(x) - \phi(x)| \mathrm{d}x.$$

A simple calculation shows that f satisfies (11) with

$$c_1 = \ldots = c_n = \frac{2}{n},$$

and therefore (12) yields

$$\mathrm{Var}[Z_n] \leq \frac{1}{n}.$$

Now, to take full advantage of the martingale method, we need to combine the martingale decomposition (3) with the Chernoff bound. To proceed, we first note that the sequence of random variables $Z_k \triangleq \mathbb{E}[Z|X^k]$, for $k = 0, 1, \ldots, n$, is a martingale with respect to $X_1, \ldots, X_n$, i.e., $\mathbb{E}[Z_{k+1}|X^k] = Z_k$ for each k. Here is one frequently used concentration result:

Theorem 1 (Azuma–Hoeffding inequality [10], [11])**.** *Let $\{Z_k\}_{k=0}^n$ be a real-valued martingale sequence. Suppose that the martingale increments $\xi_k = Z_k - Z_{k-1}$, for $k = 1, \ldots, n$, are almost surely bounded, i.e., $|\xi_k| \leq d_k$ a.s. for some constants $d_1, \ldots, d_n \geq 0$. Then*

$$\mathbb{P}\left[|Z_n - Z_0| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2\sum_{k=1}^n d_k^2}\right), \quad \forall t > 0. \quad (13)$$

The main idea behind the proof is to apply Hoeffding's lemma to each term ξ_k in the Doob martingale decomposition (3), conditionally on X^{k-1}: for all $\lambda > 0$

$$\begin{aligned}\mathbb{E}[e^{\lambda(Z_n - Z_0)}] &= \mathbb{E}\left[\prod_{k=1}^n e^{\lambda\xi_k}\right] \\ &= \mathbb{E}\left[\prod_{k=1}^{n-1} e^{\lambda\xi_k}\mathbb{E}[e^{\lambda\xi_n}|X^{n-1}]\right].\end{aligned}$$

Since $|\xi_n| \leq d_n$, we have $\ln\mathbb{E}[e^{\lambda\xi_n}|X^{n-1}] \leq \frac{\lambda^2 d_n^2}{2}$, by Hoeffding's lemma. Continuing in this manner and peeling off the terms ξ_k one by one, we can apply the Chernoff bound and obtain (13). However, the Azuma-Hoeffding inequality is not tight in general (e.g., if $t > \sum_{k=1}^n d_k$, then the probability in the left side of (13) is zero, due to the boundedness of the ξ_k's, whereas its bound in the right side of (13) is strictly positive). One way to tighten it is to make use of additional information on the conditional variances along the martingale sequence [21]:

Theorem 2 (McDiarmid)**.** *Let $\{Z_k\}_{k=0}^\infty$ be a martingale satisfying the following two conditions for some constants $d, \sigma > 0$:*

- *$|\xi_k| \leq d$ for all k.*
- *$\mathrm{Var}[Z_k|X^{k-1}] = \mathbb{E}[|\xi_k|^2|X^{k-1}] \leq \sigma^2$ for all k.*

Then, for every $\alpha \geq 0$,

$$\mathbb{P}\left[|Z_n - Z_0| \geq \alpha n\right] \leq 2\exp\left(-nd\Big(\frac{\delta+\gamma}{1+\gamma}\Big\|\frac{\gamma}{1+\gamma}\Big)\right),$$

where $\gamma = \sigma^2/d^2$, $\delta = \alpha/d$, and $d(p\|q) \triangleq p\ln\frac{p}{q} + (1-p)\ln\frac{1-p}{1-q}$ is the binary relative entropy function.

Note that, in contrast to Theorem 1, the martingale increments $\{\xi_k\}$ in Theorem 2 should be bounded by a constant d which is independent of k.

A prominent application of the martingale method is a powerful inequality due to McDiarmid [21], also known as the bounded difference inequality:

Theorem 3 (McDiarmid's inequality)**.** *If f satisfies the bounded difference property* (11)*, and $X_1, \ldots, X_n$ are independent random variables, then for all $t > 0$*

$$\mathbb{P}\left[|f(X^n) - \mathbb{E}[f(X^n)]| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{\sum_{k=1}^n c_k^2}\right). \quad (14)$$

The strategy of the proof is similar to the one used to derive the Azuma–Hoeffding inequality. In fact, we could have used the Azuma–Hoeffding inequality to bound the tail probability in (14); however, McDiarmid's inequality provides a factor of 4 improvement in the exponent of the bound when f is a function of n independent random variables.

Here is a nice information-theoretic application of McDiarmid's inequality [22]. Consider a discrete memoryless channel (DMC) with input alphabet X, output alphabet Y, and strictly positive transition probabilities $T(y|x)$. Fix an arbitrary distribution P_{X^n} of the input n-block X^n, and let P_{Y^n} denote the resulting output distribution. Then, for every input n-block $x^n \in \mathsf{X}^n$,

$$\begin{aligned}\mathbb{P}_{Y^n|X^n=x^n}\Big[\log\frac{P_{Y^n|X^n=x^n}(Y^n)}{P_{Y^n}(Y^n)} \\ \geq D(P_{Y^n|X^n=x^n}\|P_{Y^n}) + t\Big] \leq \exp\left(-\frac{2t^2}{nc(T)}\right),\end{aligned} \quad (15)$$

where

$$c(T) \triangleq 2\max_{x,x'\in\mathsf{X}}\max_{y\in\mathsf{Y}}\log\frac{T(y|x)}{T(y|x')}. \quad (16)$$

Proof: Let us consider the function

$$f(y_1, \ldots, y_n) \triangleq \log\frac{P_{Y^n|X^n=x^n}(y^n)}{P_{Y^n}(y^n)}$$

(recall that the input block x^n is fixed). A simple calculation shows that this f has bounded differences with

$$c_1 = \ldots = c_n = c(T).$$

Moreover, since the channel is memoryless, $Y_1, \ldots, Y_n$ are independent random variables under $P_{Y^n|X^n=x^n}$ (although not under P_{Y^n}, unless P_{X^n} is a product distribution). Applying McDiarmid's inequality, we get (15). ■

The martingale method has also been used successfully to analyze concentration properties of random codes around their ensemble averages. The performance analysis of a particular code is usually difficult, especially for codes of large block lengths. Availability of a concentration result for the performance of capacity-approaching code ensembles under low-complexity decoding algorithms, as it is the case with low-density parity-check (LDPC) codes [14], validates the use of the density evolution technique as an analytical tool to assess the performance of individual codes from a code ensemble whose block length is sufficiently large, and to assess their asymptotic gap to capacity. However, it should be borne in mind that the current concentration results for codes defined on graphs, which mainly rely on the Azuma–Hoeffding inequality, are weak since in practice concentration is observed at much shorter block lengths.

Here are two illustrative examples of the use of martingale concentration inequalities in the analysis of code performance. The first result, due to Sipser and Spielman [23], is useful for assessing the performance of bit-flipping decoding algorithms for expander codes:

Theorem 4 (Sipser and Spielman)**.** *Let $\mathcal{G}$ be a bipartite graph that is chosen uniformly at random from the ensemble of bipartite graphs with n vertices on the left, a left degree l, and a right degree r. Let $\alpha \in (0,1)$ and $\delta > 0$ be fixed numbers. Then, with probability at least $1 - \exp(-\delta n)$, all sets of αn vertices on the left side of $\mathcal{G}$ are connected to at least*

$$n\left[\frac{l\big(1-(1-\alpha)^r\big)}{r} - \sqrt{2l\alpha\left(h(\alpha)+\delta\right)}\right]$$

vertices (neighbors) on the right side of $\mathcal{G}$, where h is the binary entropy function to base e (i.e., $h(x) = -x\ln(x) - (1-x)\ln(1-x)$, $x \in [0,1]$).

The proof revolves around the analysis of the so-called *neighbor exposure martingale* via the Azuma–Hoeffding inequality to bound the probability that the number of neighbors deviates significantly from its mean value.

Let $\mathrm{LDPC}(n,\lambda,\rho)$ denote an LDPC code ensemble of block length n, respectively, and with left and right degree distributions λ and ρ from the edge perspective (i.e., λ_i designates the fraction of edges which are connected to a variable node of degree i, and ρ_i designates the fraction of edges which are connected to parity-check nodes of degree i).

The second result, due to Richardson and Urbanke [24], concerns the performance of message-passing decoding algorithms for LDPC codes.

Theorem 5 (Richardson–Urbanke)**.** *Let $\mathcal{C}$, a code chosen uniformly at random from the ensemble $\mathrm{LDPC}(n,\lambda,\rho)$, be used for transmission over a memoryless binary-input output-symmetric (MBIOS) channel. Assume that the decoder performs ℓ iterations of message-passing decoding, and let $P_{\mathrm{b}}(\mathcal{C},\ell)$ denote the resulting bit error probability. Then, for every $\delta > 0$, there exists some $\alpha = \alpha(\lambda,\rho,\delta,\ell) > 0$ (independent of the block length n), such that*

$$\mathbb{P}\left[\left|P_{\mathrm{b}}(\mathcal{C},\ell) - \mathbb{E}_{\mathrm{LDPC}(n,\lambda,\rho)}[P_{\mathrm{b}}(\mathcal{C},\ell)]\right| \geq \delta\right] \leq e^{-\alpha n}$$

The proof also applies the Azuma–Hoeffding inequality to a certain martingale sequence. Some additional references on the use of the martingale method in the context of codes include [14], [23]–[29]. For more details, we refer the reader to our monograph [9].

IV. The entropy method and logarithmic Sobolev inequalities

The entropy method, as its name suggests, relies on information-theoretic techniques to control the logarithmic moment-generating function ψ directly in terms of certain relative entropies. Recall our roadmap for proving a concentration inequality for $Z = f(X)$, where X is an arbitrary random variable:

- Derive a tight quadratic bound on ψ:

$$\psi(\lambda) = \log \mathbb{E}[e^{\lambda(Z-\mathbb{E}[Z])}] \leq \frac{\lambda^2\sigma^2}{2}.$$

- Use the Chernoff bound to get

$$\mathbb{P}[Z \geq \mathbb{E}[Z] + t] \leq e^{-t^2/2\sigma^2}.$$

Let $P = \mathcal{L}(X)$, and introduce the tilted distribution $P^{(\lambda f)}$:

$$\mathrm{d}P^{(\lambda f)} = \frac{e^{\lambda f}\mathrm{d}P}{\mathbb{E}_P[e^{\lambda f}]}.$$

The entropy method revolves around the relative entropy $D(P^{(\lambda f)}\|P)$, and has two ingredients: (1) the Herbst argument, and (2) tensorization.

We start with the Herbst argument (the name refers to an unpublished note by I. Herbst that proposed the use of such an argument in the context of mathematical physics of quantum fields). Let us examine the relative entropy:

$$\begin{aligned} D(P^{(\lambda f)}\|P) &= \int \mathrm{d}P^{(\lambda f)} \log \frac{\mathrm{d}P^{(\lambda f)}}{\mathrm{d}P} \\ &= \mathbb{E}^{(\lambda f)}\left[\lambda f(X) - \psi(\lambda)\right] \\ &= \lambda\psi'(\lambda) - \psi(\lambda), \end{aligned}$$

where $\mathbb{E}^{(\lambda f)}[\cdot]$ denotes expectation with respect to the tilted distribution $P^{(\lambda f)}$. Now, with a bit of foresight, we rewrite the last expression as

$$\lambda\psi'(\lambda) - \psi(\lambda) = \lambda^2 \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\psi(\lambda)}{\lambda}\right).$$

Thus, we end up with the identity

$$D(P^{(\lambda f)}\|P) = \lambda^2 \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\psi(\lambda)}{\lambda}\right).$$

Integrating and using the fact that $\lim_{\lambda\to 0}\frac{\psi(\lambda)}{\lambda} = 0$ (which can be proved using l'Hopital's rule), we get

$$\psi(\lambda) = \lambda\int_0^\lambda \frac{D(P^{(tf)}\|P)}{t^2}\mathrm{d}t. \tag{17}$$

Appealing to the Chernoff bound, we end up with the following:

Lemma 4 (The Herbst argument)**.** *Suppose that $Z = f(X)$ is such that*

$$D(P^{(\lambda f)}\|P) \leq \frac{\lambda^2\sigma^2}{2}, \qquad \forall \lambda \geq 0. \tag{18}$$

Then Z is σ^2-subgaussian, and therefore

$$\mathbb{P}\left[f(X) \geq \mathbb{E}[f(X)] + t\right] \leq e^{-t^2/2\sigma^2}, \qquad \forall t \geq 0. \tag{19}$$

In fact, it can be shown that the reverse implication holds as well, but with some loss in the constants [30]: if $Z = f(X)$ is $\sigma^2/4$-subgaussian, then

$$D(P^{(\lambda f)}\|P) \leq \frac{\lambda^2\sigma^2}{2}, \qquad \lambda \geq 0.$$

In other words, subgaussianity of $Z = f(X)$ is equivalent to $D(P^{(\lambda f)}\|P) = O(\lambda^2)$. It seems, therefore, that we have not really accomplished anything, apart from arriving at an equivalent characterization of subgaussianity. However, the relative entropy has one crucial property: it tensorizes. Recall that we are interested in the high-dimensional setting, where $X = (X_1, \ldots, X_n)$ is a tuple of n independent random variables. Thus, $P = \mathcal{L}(X)$ is a product distribution: $P_X = P_{X_1} \otimes \ldots \otimes P_{X_n}$. Using this fact together with the chain rule for relative entropy, we arrive at the following:

Lemma 5 (Tensorization of the relative entropy)**.** *Let P and Q be two probability distributions of a random n-tuple $X = (X_1, \ldots, X_n)$, such that the coordinates of X are independent under P. Then*

$$D(Q\|P) \le \sum_{i=1}^{n} D(Q_{X_i|\overline{X}^i}\|P_{X_i}|Q_{\overline{X}^i}). \tag{20}$$

The quantity on the right-hand side of (20) is the *erasure divergence* between Q and P [31]. We now particularize this general bound to our problem, where Q is given by the tilted distribution $P^{(\lambda f)}$. In that case, using Bayes' rule and the fact that the X_i's are independent, we can express the conditional distributions $P^{(\lambda f)}_{X_i|\overline{X}^i}$ as follows: for each $\bar{x}^i$,

$$\mathrm{d}P^{(\lambda f)}_{X_i|\overline{X}^i=\bar{x}^i} = \frac{e^{\lambda f(x_1,\ldots,x_{i-1},\cdot,x_{i+1},\ldots,x_n)}}{\mathbb{E}\left[e^{\lambda f(x_1,\ldots,x_{i-1},X_i,x_{i+1},\ldots,x_n)}\right]}\mathrm{d}P_{X_i}.$$

This looks formidable; nevertheless, it reveals that the conditional distribution $P^{(\lambda f)}_{X_i|\overline{X}^i=\bar{x}^i}$ is the exponential tilting of the marginal distribution P_{X_i} with respect to the random variable $f_i(X_i) = f(x_1, \ldots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n)$, which depends only on X_i because $\bar{x}^i$ is fixed. Thus, we arrive at the following bound:

$$D(P^{(\lambda f)}\|P) \le \sum_{i=1}^{n} \widetilde{\mathbb{E}}\left[D(P^{(\lambda f_i)}_{X_i}\|P_{X_i})\right],$$

where the expectation on the right-hand side is with respect to the tilted distribution.

We can now distill the entropy method into a series of steps:

1) We wish to derive a subgaussian tail bound

$$\mathbb{P}\left[f(X^n) \ge \mathbb{E}[f(X^n)] + t\right] \le e^{-t^2/2\sigma^2}, \qquad t \ge 0,$$

where $X_1, \ldots, X_n$ are independent random variables.

2) Suppose that we can prove that there exist constants $c_1, \ldots, c_n \ge 0$, such that

$$D(P^{(\lambda f_i)}_{X_i}\|P_{X_i}) \le \frac{\lambda^2 c_i^2}{2}, \qquad \forall i. \tag{21}$$

3) Then, by the tensorization lemma,

$$D(P^{\lambda f}\|P) \le \frac{\lambda^2 \sum_{i=1}^n c_i^2}{2},$$

and therefore, by the Herbst argument, $Z = f(X^n)$ is σ^2-subgaussian with $\sigma^2 = \sum_{i=1}^n c_i^2$.

The main benefit of passing to the relative-entropy characterization of subgaussianity is that now, via tensorization, we have broken up a difficult n-dimensional problem into n presumably easier 1-dimensional problems, each of which boils down to analyzing the behavior of the function $f_i(X_i) \equiv f(x_1, \ldots, x_{i-1}, X_i, x_{i+1}, \ldots, x_n)$, where only the ith input coordinate is random, and the remaining ones are fixed at some arbitrary values.

Of course, the problem now reduces to showing that (21) holds. One route, which often yields tight constants, is via so-called *logarithmic Sobolev inequalities*. In a nutshell, a logarithmic Sobolev inequality (or LSI, for short) ties together a probability distribution P, some function class $\mathcal{A}$ that contains the function f of interest, and an "energy" functional $E : \mathcal{A} \to \mathbb{R}$ with the property

$$E(\alpha f) = \alpha E(f), \qquad \forall \alpha \ge 0, f \in \mathcal{A}.$$

With these ingredients in place, a log-Sobolev inequality takes the form

$$D(P^{(f)}\|P) \le \frac{c}{2}E^2(f), \qquad \forall f \in \mathcal{A}.$$

Now suppose that $E(f) \le L$. Then we readily get the bound

$$D(P^{(\lambda f)}\|P) \le \frac{c}{2}E^2(\lambda f) = \frac{\lambda^2}{2}cE^2(f) \le \frac{\lambda^2 cL^2}{2},$$

so $f(X)$, $X \sim P$, is σ^2-subgaussian with $\sigma^2 = cL^2$.

There is a vast literature on log-Sobolev inequalities, and an interested reader may consult our monograph for more details and additional references. Here we will give the two classic examples: the Bernoulli LSI and the Gaussian LSI, due to Gross [32].

Theorem 6 (Bernoulli LSI)**.** *Let $X_1, \ldots, X_n$ be i.i.d.* $\mathrm{Bern}(1/2)$ *random variables. Then, for every function $f\colon \{0,1\}^n \to \mathbb{R}$, we have*

$$D(P^{(f)}\|P) \le \frac{1}{8}\frac{\mathbb{E}\left[|Df(X^n)|^2 e^{f(X^n)}\right]}{\mathbb{E}[e^{f(X^n)}]}, \tag{22}$$

where $P = \mathrm{Bern}(1/2)^{\otimes n}$,

$$Df(x^n) \triangleq \sqrt{\sum_{i=1}^{n} |f(x^n) - f(x^n \oplus e_i)|^2},$$

and $x^n \oplus e_i$ is the XOR of x^n with the bit string of all zeros, except for the ith bit. In other words, $x^n \oplus e_i$ is x^n with the ith bit flipped.

The proof, which we omit, is to first establish the $n = 1$ case via a straightforward if tedious exercise in calculus, and then to extend to an arbitrary n by tensorization. Note that the mapping $f \mapsto Df$ has the desired scaling property: $D(\alpha f) = \alpha D(f)$ for all $\alpha \ge 0$.

Theorem 7 (Gaussian LSI)**.** *Let $X_1, \ldots, X_n$ be i.i.d. $N(0,1)$ random variables. Then, for an arbitrary smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$,*

$$D(P^{(f)}\|P) \le \frac{1}{2}\frac{\mathbb{E}\left[\|\nabla f(X^n)\|_2^2 e^{f(X^n)}\right]}{\mathbb{E}[e^{f(X^n)}]}. \tag{23}$$

Note that the mapping $f \mapsto \|\nabla f\|_2$ has the scaling property: $\|\nabla(\alpha f)\|_2 = \alpha\|\nabla f\|_2$ for all $\alpha \geq 0$. By now, there are at least fifteen different ways in the literature for proving the Gaussian LSI. The original proof by Gross was to apply the Bernoulli LSI to the function

$$f\left(\frac{X_1 + \ldots + X_n - n/2}{\sqrt{n/4}}\right), \qquad X_i \overset{\text{i.i.d.}}{\sim} \text{Bern}(1/2),$$

and then pass to the Gaussian limit by appealing to the Central Limit Theorem.

The Gaussian LSI can be used to give a short proof of the following concentration inequality for Lipschitz functions of Gaussians, which was originally obtained by Tsirelson, Ibragimov, and Sudakov [33] using different methods:

Theorem 8 (Tsirelson–Ibragimov–Sudakov)**.** *Let $X_1, \ldots, X_n$ be i.i.d. $N(0,1)$ random variables, and let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a function which is L-Lipschitz:*

$$|f(x^n) - f(y^n)| \leq L\,\|x^n - y^n\|_2\,.$$

Then, $f(X^n)$ is L^2-subgaussian, which yields

$$\mathbb{P}\left[f(X^n) \geq \mathbb{E}[f(X^n)] + t\right] \leq e^{-\frac{t^2}{2L^2}} \tag{24}$$

for all $t > 0$.

Proof: By a standard approximation argument, we may assume that f is differentiable. Since it is also L-Lipschitz, $\|\nabla f\|_2^2 \leq L^2$ everywhere. Substituting this bound into the Gaussian LSI for λf, we obtain

$$D(P^{(\lambda f)}\|f) \leq \frac{\lambda^2 L^2}{2}.$$

By the Herbst argument, $Z = f(X^n)$, $X^n \sim N(0, I_n)$, is L^2-subgaussian, and we are done. ∎

This result is remarkable in two ways: It only assumes Lipschitz continuity of f, and gives *dimension-free* concentration (i.e., the exponent in (24) does not depend on n).

Deriving log-Sobolev inequalities, especially with tight constants, is a subtle art. A commonly used method is to realize P as an invariant distribution of some continuous-time reversible Markov process and to extract a suitable energy functional E from the structure of the infinitesimal generator of the process. In many cases, however, it is possible to derive a log-Sobolev inequality via tensorization and a nice and simple variance-based representation of the relative entropy due to A. Maurer [34]:

Theorem 9 (Maurer)**.** *Let X be a random variable with law P. Then, for every real-valued function f and all $\lambda \geq 0$*

$$D(P^{(\lambda f)}\|P) = \int_0^\lambda \int_t^\lambda \text{Var}^{(sf)}[f(X)]\mathrm{d}s\,\mathrm{d}t,$$

where $\text{Var}^{(sf)}[f(X)]$ is the variance of $f(X)$ under the tilted distribution $P^{(sf)}$.

Proof: As before, let $\psi(\lambda) = \log \mathbb{E}[e^{\lambda(f(X)) - \mathbb{E}[f(X)]}]$ be the logarithmic moment-generating function of $f(X)$. Then

$$\begin{aligned} D(P^{(\lambda f)}\|P) &= \lambda\psi'(\lambda) - \psi(\lambda) \\ &= \int_0^\lambda [\psi'(\lambda) - \psi'(t)]\,\mathrm{d}t \\ &= \int_0^\lambda \int_t^\lambda \psi''(s)\mathrm{d}s\,\mathrm{d}t, \end{aligned}$$

where we have used the fact that $\psi(0) = \psi'(0) = 0$ and the fundamental theorem of calculus. Recalling that $\psi''(s) = \text{Var}^{(sf)}[f(X)]$, we are done. ∎

The following result is a direct consequence of Theorem 9:

Theorem 10. *Let $\mathcal{A}$ be a class of functions of X, and suppose that there is a mapping $\Gamma : \mathcal{A} \to \mathbb{R}$, such that:*

1) *For all $f \in \mathcal{A}$ and $\alpha \geq 0$, $\Gamma(\alpha f) = \alpha\Gamma(f)$.*
2) *There exists a constant $c > 0$, such that*

$$\text{Var}^{(\lambda f)}[f(X)] \leq c|\Gamma(f)|^2, \qquad \forall f \in \mathcal{A},\ \lambda \geq 0.$$

Then

$$D(P^{(\lambda f)}\|P) \leq \frac{\lambda^2 c|\Gamma(f)|^2}{2}, \qquad \forall f \in \mathcal{A}, \lambda \geq 0.$$

To illustrate Maurer's method, let's use it to derive the Bernoulli LSI. It suffices to prove the $n = 1$ case, and then to scale up to an arbitrary n by tensorization. Thus, let $P = \text{Bern}(1/2)$, and for every function $f\colon \{0,1\} \to \mathbb{R}$ define $\Gamma(f) \triangleq |f(0) - f(1)|$. By Lemma 1,

$$\text{Var}^{(\lambda f)}[f(X)] \leq \tfrac{1}{4}\,|f(0) - f(1)|^2 = \tfrac{1}{4}\,|\Gamma(f)|^2.$$

Thus, the conditions of Theorem 10 are satisfied with $c = 1/4$, and we get precisely the Bernoulli LSI. One can also use Maurer's method to prove McDiarmid's inequality (see Theorem 3).

V. Transportation-cost inequalities

At this point, we notice a common theme running through the above examples of concentration phenomena:

- Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be 1-Lipschitz with respect to the Euclidean norm $\|\cdot\|_2$, and let $X_1, \ldots, X_n$ be i.i.d. $N(0,1)$ random variables. Then

$$\mathbb{P}[f(X^n) \geq \mathbb{E}[f(X^n)] + t] \leq e^{-t^2/2}.$$

- Let X be an arbitrary space, and consider a function $f\colon \mathsf{X}^n \to \mathbb{R}$, which is 1-Lipschitz with respect to the weighted Hamming metric

$$d_{\mathbf{c}}(x^n, y^n) \triangleq \sum_{i=1}^n c_i \mathbf{1}_{\{x_i \neq y_i\}},$$

where $c_1, \ldots, c_n \geq 0$ are some fixed constants. It is easy to see that such a Lipschitz property is equivalent to the bounded difference property (11), and in that case McDiarmid's inequality tells us that

$$\mathbb{P}[f(X^n) \geq \mathbb{E}[f(X^n)] + t] \leq e^{-2t^2/\sum_{i=1}^n c_i^2}$$

for every tuple $X_1, \ldots, X_n$ of independent X-valued random variables.

Thus, metric spaces and Lipschitz functions seem to be a natural setting to study concentration. To make this statement more precise, let (X, d) be a metric space. We say that a function $f\colon \mathsf{X} \to \mathbb{R}$ is L-Lipschitz (with respect to d) if

$$|f(x) - f(y)| \le L d(x, y), \qquad \forall x, y \in \mathsf{X}.$$

Denoting by $\mathrm{Lip}_L(\mathsf{X}, d)$ the class of all L-Lipschitz functions, we can pose the following question: What conditions does a probability distribution P on X have to satisfy, so that $f(X)$ with $X \sim P$ is σ^2-subgaussian for every $f \in \mathrm{Lip}_1(\mathsf{X}, d)$?

Through the pioneering work of Katalin Marton [17], [35]–[39], the answer to the above question has deep links to information theory via the notion of so-called *transportation-cost inequalities* [40]. In order to introduce them, we first need some definitions. A *coupling* of two probability distributions P and Q on X is a probability distribution π on the Cartesian product $\mathsf{X} \times \mathsf{X}$, such that for $(X, Y) \sim \pi$ we have $X \sim P$ and $Y \sim Q$. Let $\Pi(P, Q)$ denote the set of all couplings of P and Q. For $p \ge 1$, the L^p *Wasserstein distance* between P and Q is defined as

$$W_p(P, Q) \triangleq \inf_{\pi \in \Pi(P,Q)} \left(\mathbb{E}_\pi[d^p(X, Y)]\right)^{1/p}.$$

The name "transportation cost" comes from the following interpretation: Let P (resp., Q) represent the initial (resp., desired) distribution of some matter (say, sand) in space, such that the total mass in both cases is normalized to one. Thus, both P and Q correspond to sand piles of some given shapes. The objective is to rearrange the initial sand pile with shape P into one with shape Q with minimum cost, where the cost of transporting a grain of sand from location x to location y is given by $d^p(x, y)$. If we allow randomized transportation policies, i.e., those that associate with each location x in the initial sand pile a conditional probability distribution $\pi(\mathrm{d}y|x)$ for its destination in the final sand pile, then the minimum transportation cost is given by $W_p(P, Q)$. We say that P satisfies an L^p *transportation-cost inequality* with constant c, or $T_p(c)$ for short, if

$$W_p(P, Q) \le \sqrt{2cD(Q\|P)}, \qquad \forall Q.$$

The well-known Pinsker's inequality is, in fact, a transportation-cost inequality: If we take X to be an arbitrary space and equip it with the metric $d(x, y) = \mathbf{1}_{\{x \ne y\}}$, then the L^1 Wasserstein distance $W_1(P, Q)$ is simply the total variation distance

$$\|P - Q\|_{\mathrm{TV}} = \sup_A |P(A) - Q(A)|,$$

and Pinsker's inequality

$$\|P - Q\|_{\mathrm{TV}} \le \sqrt{\tfrac{1}{2} D(Q\|P)}$$

(in nats) is then a $T_1(\frac{1}{4})$ inequality, which is satisfied by all probability measures P, Q where $Q \ll P$ (i.e., Q is absolutely continuous with respect to P). Various distribution-dependent refinements of Pinsker's inequality where the constant is optimized for a fixed P while varying only Q [41], [42] can be interpreted in the same vein as well. Another well-known transportation-cost (TC) inequality is due to Talagrand [43]: Let X be the Euclidean space $\mathbb{R}^n$, equipped with the Euclidean metric $d(x, y) = \|x - y\|_2$. Then $P = N(0, I_n)$ satisfies the $T_2(1)$ inequality: $W_2(P, Q) \le \sqrt{2D(Q\|P)}$. The remarkable thing here is that the constant is independent of the dimension n.

With these preliminaries out of the way, we can now state the theorem, due to Bobkov and Götze [44], which provides an answer to the question posed above:

Theorem 11 (Bobkov–Götze)**.** *Let X be a random variable taking values in a metric space (X, d) according to a probability distribution P. Then, the following are equivalent:*

1) *$f(X)$ is σ^2-subgaussian for every $f \in \mathrm{Lip}_1(\mathsf{X}, d)$.*
2) *P satisfies $T_1(\sigma^2)$, i.e.,*

$$W_1(P, Q) \le \sqrt{2\sigma^2 D(Q\|P)}$$

for all Q.

At this point, one may wonder what we have gained – verifying that a given P satisfies a TC inequality, let alone determining tight constants, is a formidable challenge. However, once again, tensorization comes to the rescue. Marton's insight was that TC inequalities tensorize [40]:

Theorem 12. *Let (X_i, P_i, d_i), $1 \le i \le n$, be probability metric spaces. If for some $1 \le p \le 2$ each P_i satisfies $T_p(c)$ on (X_i, d_i), then the product measure $P = P_1 \otimes \ldots \otimes P_n$ on $\mathsf{X} = \mathsf{X}_1 \times \ldots \times \mathsf{X}_n$ satisfies $T_p(cn^{2/p-1})$ w.r.t. the metric*

$$d_p(x^n, y^n) \triangleq \left(\sum_{i=1}^n d_i^p(x_i, y_i)\right)^{1/p}.$$

In particular, if each P_i satisfies $T_1(c)$, then $P = P_1 \otimes \ldots \otimes P_n$ satisfies $T_1(cn)$ with respect to the metric $\sum_i d_i$. Note that the constant deteriorates with n. On the other hand, if each P_i satisfies $T_2(c)$, then P satisfies $T_2(c)$ with respect to $\sqrt{\sum_i d_i^2}$. Note that the latter constant is independent of n.

To give a simple illustration of all these concepts, let us outline yet another proof of McDiarmid's inequality. Consider a product probability space $(\mathsf{X}_1 \times \ldots \times \mathsf{X}_n, P_1 \otimes \ldots \otimes P_n)$. For a fixed choice of constants $c_1, \ldots, c_n \ge 0$, equip X_i with the metric $d_i(x_i, y_i) = c_i \mathbf{1}_{\{x_i \ne y_i\}}$. Then, by rescaling Pinsker's inequality, we see that P_i satisfies a $T_1(c_i^2/4)$ inequality with respect to the metric d_i:

$$W_{1,d_i}(P_i, Q_i) \le \sqrt{\tfrac{1}{2} c_i^2 D(Q_i\|P_i)}, \qquad \forall Q_i. \tag{25}$$

By the tensorization theorem for TC inequalities, the product distribution P satisfies a $T_1(c)$ inequality with $c = (1/4)\sum_{i=1}^n c_i^2$ with respect to the weighted Hamming metric $d_{\mathbf{c}}$. By the Bobkov–Götze theorem, this is equivalent to the subgaussianity of all $f(X_1, \ldots, X_n)$ with $f \in \mathrm{Lip}_1(\mathsf{X}, d)$ and mutually independent $X_i \in \mathsf{X}_i$, $1 \le i \le n$. But this is precisely McDiarmid's inequality.

VI. Some applications in information theory

We end this survey by briefly describing some information-theoretic applications of concentration inequalities.

A. The Blowing-up Lemma and Information-Theoretic Consequences

The first explicit appeal to the concentration phenomenon in information theory dates back to the 1970s work by Ahlswede and collaborators, who used the so-called *blowing-up lemma* for deriving strong converses for a variety of communications and coding problems.

Consider a product space Y^n equipped with the Hamming metric $d(y^n, z^n) = \sum_{i=1}^n \mathbf{1}_{\{y_i \neq z_i\}}$. For $r \in \{0, 1, \ldots, n\}$, define the *r-blowup* of a set $A \subseteq \mathsf{Y}^n$ as

$$[A]_r \triangleq \left\{ z^n \in \mathsf{Y}^n : \min_{y^n \in A} d(z^n, y^n) \leq r \right\}$$

The following result, in a different (asymptotic) form was first proved by Ahlswede, Gács, and Körner [45]; a simple proof, which we sketch below, was given by Marton [35]:

Lemma 6 (Blowing-up)**.** *Let $Y_1, \ldots, Y_n$ be independent random variables taking values in Y. Then for every set $A \subseteq \mathsf{Y}^n$ with $P_{Y^n}(A) > 0$*

$$P_{Y^n}\left\{[A]_r\right\} \geq 1 - \exp\left[-\frac{2}{n}\left(r - \sqrt{\frac{n}{2}\log\frac{1}{P_{Y^n}(A)}}\right)_+^2\right],$$

where $(u)_+ \triangleq \max\{0, u\}$.

Proof: We sketch the proof in order to highlight the role of TC inequalities. For each $i \in \{1, \ldots, n\}$, let $P_i = \mathcal{L}(Y_i)$. By tensorization, the product distribution $P = P_{Y^n}$ satisfies the TC inequality

$$W_1(P, Q) \leq \sqrt{\frac{n}{2} D(Q\|P)}, \qquad \forall Q, \tag{26}$$

where

$$W_1(P, Q) = \inf_{\pi \in \Pi(P,Q)} \mathbb{E}_\pi \left[\sum_{i=1}^n \mathbf{1}_{\{X_i \neq Y_i\}}\right].$$

Now, for an arbitrary $B \subseteq \mathsf{Y}^n$ with $P(B) > 0$, consider the conditional distribution $P_B(\cdot) \triangleq \frac{P(\cdot \cap B)}{P(B)}$. Then $D(P_B\|P) = \log\frac{1}{P(B)}$, and in that case using (26) with $Q = P_B$, we get

$$W_1(P, P_B) \leq \sqrt{\frac{n}{2}\log\frac{1}{P(B)}}. \tag{27}$$

Applying (26) to $B = A$ and $B = [A]_r^c$, we get

$$W_1(P, P_A) \leq \sqrt{\frac{n}{2}\log\frac{1}{P(A)}},$$

$$W_1(P, P_{[A]_r^c}) \leq \sqrt{\frac{n}{2}\log\frac{1}{1 - P([A]_r)}}.$$

Adding up these two inequalities, we obtain

$$\begin{aligned}
&\sqrt{\frac{n}{2}\log\frac{1}{P(A)}} + \sqrt{\frac{n}{2}\log\frac{1}{1 - P([A]_r)}} \\
&\quad \geq W_1(P_A, P) + W_1(P_{[A]_r^c}, P) \\
&\quad \geq W_1(P_A, P_{[A]_r^c}) \\
&\quad \geq \min_{x^n \in A, y^n \in [A]_r^c} d(x^n, y^n) \\
&\quad \geq r,
\end{aligned}$$

where the first step holds due to (27), the second step is verified by the triangle inequality, and the remaining steps follow from definitions. Rearranging, we obtain the lemma. ∎

Informally, the lemma states that every set in a product space can be "blown up" to engulf most of the probability mass. Using this fact, one can prove strong converses for channel coding in single-terminal and multiterminal settings. Here is the simplest consequence of the blowing-up lemma in the context of channel codes: Consider a DMC with input alphabet X, output alphabet Y, and transition probabilities $T(y|x)$, $x \in \mathsf{X}, y \in \mathsf{Y}$. An (n, M, ε)-code for T consists of an encoder $f\colon \{1, \ldots, M\} \to \mathsf{X}^n$ and a decoder $g : \mathsf{Y}^n \to \{1, \ldots, M\}$, such that

$$\max_{1 \leq j \leq M} \mathbb{P}[g(Y^n) \neq j | f(X^n) = j] \leq \varepsilon.$$

Lemma 7. *Let $u_j = f(j)$, $1 \leq j \leq M$, denote the M codewords of the code, and let $D_j \triangleq g^{-1}(j)$ be the corresponding decoding regions in Y^n. There exists some $\delta_n > 0$, such that*

$$T^n\left([D_j]_{n\delta_n} \middle| X^n = u_j\right) \geq 1 - \frac{1}{n}, \qquad j = 1, \ldots, M.$$

Informally, this corollary of the blowing-up lemma says that "any bad code contains a good subcode." Using this result, Ahlswede and Dueck [46] established a strong converse for channel coding as follows: Consider an (n, M, ε)-code $\mathcal{C} = \{(u_j, D_j)\}_{j=1}^M$. Each decoding set D_j can be "blown up" to a set $\tilde{D}_j \subseteq \mathsf{Y}^n$ with

$$T^n(\tilde{D}_j | u_j) \geq 1 - \frac{1}{n}.$$

The object $\tilde{\mathcal{C}} = \{(u_j, \tilde{D}_j)\}_{j=1}^M$ is not a code (since the sets $\tilde{D}_j$ are no longer disjoint), but a random coding argument can be used to extract an (n, M', ε') "subcode" with M' slightly smaller than M and $\varepsilon' < \varepsilon$. Then one can apply the usual (weak) converse to the subcode. Similar ideas have found use in multiterminal settings, starting with the work of Ahlswede–Gács–Körner [45].

B. Empirical distribution of good channel codes with non-vanishing error probability

Another recent application of concentration inequalities to information theory has to do with characterizing stochastic behavior of output sequences of good channel codes. On a conceptual level, the random coding argument originally used by Shannon (and many times since) to show the existence of

good channel codes suggests that the input/output sequence of such a code should resemble, as much as possible, a typical realization of a sequence of i.i.d. random variables sampled from a capacity-achieving input/output distribution. For capacity-achieving sequences of codes with asymptotically vanishing probability of error, this intuition has been analyzed rigorously by Shamai and Verdú [47], who have proved the following remarkable statement [47, Theorem 2]: given a DMC T, any capacity-achieving sequence of channel codes with asymptotically vanishing probability of error (maximal or average) has the property that

$$\lim_{n\to\infty}\frac{1}{n}D(P_{Y^n}\|P^*_{Y^n})=0, \tag{28}$$

where, for each n, P_{Y^n} denotes the output distribution on Y^n induced by the code (assuming that the messages are equiprobable), while $P^*_{Y^n}$ is the product of n copies of the single-letter capacity-achieving output distribution. In a recent paper [48], Polyanskiy and Verdú extended the results of [47] for codes with *nonvanishing* probability of error.

To keep things simple, we will only focus on channels with finite input and output alphabets. Thus, let X and Y be finite sets, and consider a DMC T with capacity C. Let $P^*_X\in\mathcal{P}(\mathsf{X})$ be a capacity-achieving input distribution (which may be nonunique). It can be shown [49] that the corresponding output distribution $P^*_Y\in\mathcal{P}(\mathsf{Y})$ is unique. Consider any (n,M)-code $\mathcal{C}=(f,g)$, let $P^{(\mathcal{C})}_{X^n}$ denote the distribution of $X^n=f(J)$, where J is uniformly distributed in $\{1,\ldots,M\}$, and let $P^{(\mathcal{C})}_{Y^n}$ denote the corresponding output distribution. The central result of [48] is that the output distribution $P^{(\mathcal{C})}_{Y^n}$ of any (n,M,ε)-code satisfies

$$D\big(P^{(\mathcal{C})}_{Y^n}\big\|P^*_{Y^n}\big)\le nC-\log M+o(n); \tag{29}$$

moreover, the $o(n)$ term was refined in [48, Theorem 5] to $O(\sqrt{n})$ for any DMC, except those that have zeroes in their transition matrix. Using McDiarmid's inequality, this result is sharpened as follows [22]:

Theorem 13. *Consider a DMC T with positive transition probabilities. Then any (n,M,ε)-code $\mathcal{C}$ for T, with $\varepsilon\in(0,1/2)$, satisfies*

$$D\Big(P^{(\mathcal{C})}_{Y^n}\Big\|P^*_{Y^n}\Big)\le nC-\log M+\log\frac{1}{\varepsilon}+c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}},$$

where $c(T)$ is defined in (16).

Proof (Sketch): Using the inequality (15) with $P_{Y^n}=P^{(\mathcal{C})}_{Y^n}$ and $t=c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}$, we get

$$P_{Y^n|X^n=x^n}\left[\log\frac{P_{Y^n|X^n=x^n}(Y^n)}{P^{(\mathcal{C})}_{Y^n}(Y^n)}\ge D\Big(P_{Y^n|X^n=x^n}\Big\|P^{(\mathcal{C})}_{Y^n}\Big)+c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}\right]\le 1-2\varepsilon$$

Now, just like Polyanskiy and Verdú, we can appeal to a strong converse result due to Augustin [50] to get

$$\log M\le\log\frac{1}{\varepsilon}+D\Big(P_{Y^n|X^n}\Big\|P^{(\mathcal{C})}_{Y^n}\Big|P^{(\mathcal{C})}_{X^n}\Big)+c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}}. \tag{30}$$

Therefore,

$$\begin{aligned}&D\Big(P^{(\mathcal{C})}_{Y^n}\Big\|P^*_{Y^n}\Big)\\&=D\Big(P_{Y^n|X^n}\Big\|P^*_{Y^n}\Big|P^{(\mathcal{C})}_{X^n}\Big)-D\Big(P_{Y^n|X^n}\Big\|P^{(\mathcal{C})}_{Y^n}\Big|P^{(\mathcal{C})}_{X^n}\Big)\\&\le nC-\log M+\log\frac{1}{\varepsilon}+c(T)\sqrt{\frac{n}{2}\log\frac{1}{1-2\varepsilon}},\end{aligned}$$

where the first step is by the chain rule, the second follows from the properties of the capacity-achieving output distribution, and the last step uses (30). ■

A useful consequence of this result is that a broad class of functions evaluated on the output of a good code concentrate sharply around their expectations with respect to the capacity-achieving output distribution:

Theorem 14. *Consider a DMC T with $c(T)<\infty$. Let d be a metric on Y^n, and suppose that $P_{Y^n|X^n=x^n}$, $x^n\in\mathsf{X}^n$, as well as $P^*_{Y^n}$, satisfy $T_1(c)$ for some $c>0$. Then, for every $\varepsilon\in(0,1/2)$, every (n,M,ε)-code $\mathcal{C}$ for T, and every function $f\colon\mathsf{Y}^n\to\mathbb{R}$ which is L-Lipschitz on (Y^n,d), we have*

$$\begin{aligned}&P^{(\mathcal{C})}_{Y^n}\Big(|f(Y^n)-\mathbb{E}[f(Y^{*n})]|\ge t\Big)\\&\le\frac{4}{\varepsilon}\exp\left(nC-\ln M+a\sqrt{n}-\frac{t^2}{8cL^2}\right),\ \forall\, r\ge 0\end{aligned} \tag{31}$$

where $Y^{*n}\sim P^*_{Y^n}$, and $a\triangleq c(T)\sqrt{\frac{1}{2}\ln\frac{1}{1-2\varepsilon}}$.

As pointed out in [48], concentration inequalities like (31) can be very useful for gaining insight into the performance characteristics of good channel codes without having to explicitly construct such codes: all one needs to do is to find the capacity-achieving output distribution P^*_Y and evaluate $\mathbb{E}[f(Y^{*n})]$ for an arbitrary f of interest. Consequently, the above theorem guarantees that $f(Y^n)$ concentrates tightly around $\mathbb{E}[f(Y^{*n})]$, which is relatively easy to compute since $P^*_{Y^n}$ is a product measure.

References

[1] M. Talagrand, "A new look at independence," *Annals of Probability*, vol. 24, no. 1, pp. 1–34, January 1996.

[2] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[3] M. Ledoux, *The Concentration of Measure Phenomenon*, ser. Mathematical Surveys and Monographs. American Mathematical Society, 2001, vol. 89.

[4] G. Lugosi, "Concentration of measure inequalities - lecture notes," 2009, available at http://www.econ.upf.edu/~lugosi/anu.pdf.

[5] P. Massart, *The Concentration of Measure Phenomenon*, ser. Lecture Notes in Mathematics. Springer, 2007, vol. 1896.

[6] C. McDiarmid, "Concentration," in *Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, 1998, pp. 195–248.

[7] M. Talagrand, "Concentration of measure and isoperimteric inequalities in product space," *Publications Mathématiques de l'I.H.E.S*, vol. 81, pp. 73–205, 1995.

[8] N. Alon and J. H. Spencer, *The Probabilistic Method*, 3rd ed. Wiley Series in Discrete Mathematics and Optimization, 2008.

[9] M. Raginsky and I. Sason, *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, 2nd ed. Foundations and Trends in Communications and Information Theory, Now Publishers, 2014. [Online]. Available: http://arxiv.org/abs/1212.4663.

[10] K. Azuma, "Weighted sums of certain dependent random variables," *Tohoku Mathematical Journal*, vol. 19, pp. 357–367, 1967.

[11] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, March 1963.

[12] F. Chung and L. Lu, *Complex Graphs and Networks*, ser. Regional Conference Series in Mathematics. Wiley, 2006, vol. 107.

[13] ——, "Concentration inequalities and martingale inequalities: a survey," *Internet Mathematics*, vol. 3, no. 1, pp. 79–127, March 2006, available at http://www.math.ucsd.edu/~fan/wp/concen.pdf.

[14] T. J. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[15] Y. Seldin, F. Laviolette, N. Cesa-Bianchi, J. Shawe-Taylor, and P. Auer, "PAC-Bayesian inequalities for martingales," *IEEE Trans. on Information Theory*, vol. 58, no. 12, pp. 7086–7093, December 2012.

[16] N. Gozlan and C. Leonard, "Transport inequalities: a survey," *Markov Processes and Related Fields*, vol. 16, no. 4, pp. 635–736, 2010.

[17] K. Marton, "Distance-divergence inequalities," *IEEE Information Theory Society Newsletter*, vol. 64, no. 1, pp. 9–13, March 2014.

[18] B. Efron and C. Stein, "The jackknife estimate of variance," *Annals of Statistics*, vol. 9, pp. 586–596, 1981.

[19] J. M. Steele, "An Efron–Stein inequality for nonsymmetric statistics," *Annals of Statistics*, vol. 14, pp. 753–758, 1986.

[20] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*. Springer, 2001.

[21] C. McDiarmid, "On the method of bounded differences," in *Surveys in Combinatorics*. Cambridge University Press, 1989, vol. 141, pp. 148–188.

[22] M. Raginsky and I. Sason, "Refined bounds on the empirical distribution of good channel codes via concentration inequalities," in *Proceedings of the 2013 IEEE International Workshop on Information Theory*, Istanbul, Turkey, July 2013, pp. 221–225.

[23] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Trans. on Information Theory*, vol. 42, no. 6, pp. 1710–1722, November 1996.

[24] T. J. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[25] M. G. Luby, Mitzenmacher, M. A. Shokrollahi, and D. A. Spielmann, "Efficient erasure-correcting codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 569–584, February 2001.

[26] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager bounds, density evolution, and code performance bounds," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1636–1652, July 2003.

[27] A. Montanari, "Tight bounds for LDPC and LDGM codes under MAP decoding," *IEEE Trans. on Information Theory*, vol. 51, no. 9, pp. 3247–3261, September 2005.

[28] C. Méasson, A. Montanari, and R. Urbanke, "Maxwell construction: the hidden bridge between iterative and maximum apposteriori decoding," *IEEE Trans. on Information Theory*, vol. 54, no. 12, pp. 5277–5307, December 2008.

[29] I. Sason and R. Eshel, "On concentration of measures for LDPC code ensembles," in *Proceedings of the 2011 IEEE International Symposium on Information Theory*, Saint Petersburg, Russia, August 2011, pp. 1273–1277.

[30] R. van Handel, "Probability in high dimension," ORF 570 lecture notes, Princeton University, June 2014.

[31] S. Verdú and T. Weissman, "The information lost in erasures," *IEEE Trans. on Information Theory*, vol. 54, no. 11, pp. 5030–5058, November 2008.

[32] L. Gross, "Logarithmic Sobolev inequalities," *American Journal of Mathematics*, vol. 97, no. 4, pp. 1061–1083, 1975.

[33] B. S. Tsirelson, I. A. Ibragimov, and V. N. Sudakov, "Norms of Gaussian sample functions," in *Proceedings of the Third Japan-USSR Symposium on Probability Theory*, ser. Lecture Notes in Mathematics. Springer, 1976, vol. 550, pp. 20–41.

[34] A. Maurer, "Thermodynamics and concentration," *Bernoulli*, vol. 18, no. 2, pp. 434–454, 2012.

[35] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. on Information Theory*, vol. 32, no. 3, pp. 445–446, May 1986.

[36] ——, "A measure concentration inequality for contracting Markov chains," *Geometric and Functional Analysis*, vol. 6, pp. 556–571, 1996, see also erratum in *Geometric and Functional Analysis*, vol. 7, pp. 609–613, 1997.

[37] ——, "Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration," *Annals of Probability*, vol. 24, no. 2, pp. 857–866, 1996.

[38] ——, "Measure concentration for Euclidean distance in the case of dependent random variables," *Annals of Probability*, vol. 32, no. 3B, pp. 2526–2544, 2004.

[39] ——, "Correction to 'Measure concentration for Euclidean distance in the case of dependent random variables'," *Annals of Probability*, vol. 38, no. 1, pp. 439–442, 2010.

[40] C. Villani, *Topics in Optimal Transportation*. Providence, RI: American Mathematical Society, 2003.

[41] E. Ordentlich and M. Weinberger, "A distribution dependent refinement of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1836–1840, May 2005.

[42] D. Berend, P. Harremoës, and A. Kontorovich, "Minimum KL-divergence on complements of L_1 balls," *IEEE Trans. on Information Theory*, vol. 60, no. 6, pp. 3172–3177, June 2014.

[43] M. Talagrand, "Transportation cost for Gaussian and other product measures," *Geometry and Functional Analysis*, vol. 6, no. 3, pp. 587–600, 1996.

[44] S. G. Bobkov and F. Götze, "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities," *Journal of Functional Analysis*, vol. 163, pp. 1–28, 1999.

[45] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, pp. 157–177, 1976, see correction in vol. 39, no. 4, pp. 353–354, 1977.

[46] R. Ahlswede and G. Dueck, "Every bad code has a good subcode: a local converse to the coding theorem," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, pp. 179–182, 1976.

[47] S. Shamai and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. on Information Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

[48] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," *IEEE Trans. on Information Theory*, vol. 60, no. 1, pp. 5–21, January 2014.

[49] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 291–292, 1967.

[50] U. Augustin, "Gedächtnisfreie Kanäle für diskrete Zeit," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 6, pp. 10–61, 1966.

The Historian's Column

Anthony Ephremides

By the time you read today's column we are probably already in the year 2016, although, formally, this issue of the Newsletter is still the last one of 2015. The reason this matters is that 2016 is the Centennial Anniversary of Claude E. Shannon's birthday. As explained elsewhere in this issue, the Society has numerous plans for observing this milestone. Shannon was born on April 29, 1916 and died in February of 2001. His life and accomplishments left an indelible mark on the History of Science and Technology. In particular, starting with his monumental paper of 1948 ("A Mathematical Theory of Communication"), he established the field of Information Theory.

The number of people who have known him personally and who have learned from him directly keeps dwindling. Most of our readers have only heard of him indirectly through the words of their mentors and, of course, have learned from Shannon's own works. In 1998, when we celebrated the Golden Anniversary of the founding of our Society, among the many festive activities there was a special issue of this Newsletter that Jim Massey and I co-edited that was dedicated to Shannon and which is available online and contains a wealth of retrospective information about him provided by many of the most prominent members of our Society. Also, a few years earlier, the complete works of Shannon were collected and published in a volume edited by Aaron Wyner and Neil Sloane. So, the technical side of Shannon's heritage has been well-covered and never ceases to be invoked time-and-again by more and more people working in several diverse fields. The depth and impact of his work have been simply formidable.

Here, today, I would like to only recall and offer a few glimpses of Shannon, the man who, even to those who knew him relatively well, was a reclusive and inscrutable individual. Thus, any portrait of Shannon will be fuzzy and incomplete and, hence, like many works of Art, it will allow for different interpretations, extensions, and appreciation of his persona.

I met Shannon in 1973 at the landmark ISIT in Ashkelon, Israel. He gave the first Shannon Lecture there in observation of the 25th Anniversary of the birth of Information Theory. The idea to establish the Shannon Lecture (and subsequent Shannon Award) was the brainchild of Jacob Ziv and Aaron Wyner and proved to be a visionary one. On that occasion, Shannon was rather nervous but this was only detected by those who were physically in close proximity to him and chatted with him before his talk. His lecture, which was on the subject of feedback with a fanciful twist to it, was actually delightful, informative, and entertaining. Although he was already famous he had not yet acquired the aura of a genius and a legend. That happened gradually after he slowly withdrew and receded from the spotlight of the professional arena, where he had been active for only a little over a decade or two. We know now of some of his famous quotes from that period. When he left Bell labs to go to MIT he is said to have exclaimed "Oh, God, it feels good to be back in Industry!" And when he visited the Technion in Haifa for the first time he quipped that MIT was the Technion of the United States, reversing the popular slogan that the Technion was the MIT of Israel.

We would only hear about Shannon rather rarely after that. For example, he had given some famous and broadly quoted interviews, like the one to Robert Price in 1982 and some others in which he revealed that he loved juggling, liked to build and ride monocycles, and liked drinking beer. Then we heard of his winning the prestigious 1985 Kyoto Prize. Unfortunately the Nobel Prize does not have a category that includes Information Theory, or else he would have been a slam-dunk candidate for winning it.

It was in 1985 and 1986 that the Society reached out to him again and invited him to attend the ISIT's in Brighton, England, and in Ann Arbor, Michigan respectively. In 1985 he was already not recognized by most of the attendees. It is not a fable that before the banquet a young researcher told him in confidence that apparently Claude Shannon was in attendance at the Symposium. He was introduced to a tumultuous applause by Bob McEliece, a talented Thespian in our midst, who was then President of the Society. And in 1986 he was actually called upon to deliver the Award plaques to the various recipients during the award ceremony at the banquet. I had the privilege of handing him the plaques and diplomas as I was at the time the 1st Vice-President of the Society and hence responsible for handling the Awards. In retrospect, I could already see some early signs of confusion and forgetfulness on his part that, as we know, developed later to a full-blown case of Alzheimer's disease. This unfortunate development drew slowly (but in a dignified, albeit protracted, fashion) the curtain over what had been an amazing lifetime of contributions.

Shannon had a complex and very private personality and led an equally private life. Only small fragments of that life and character became ever visible to his close associates and friends. He remained a mystery to most. His friendly and amiable expression with a trace of benevolent smile was the "brand" of his image. But there was also remoteness in that expression. He always downplayed the significance of his work and was always a champion of modesty and self-effacing behavior. All these traits together compose an appropriate figure for the founder and leader of our fascinating field.

My closing comment, as the year of celebration of Shannon begins, is that the impact of his work on technology and applications has not been fully appreciated. Shannon was NOT a mathematician. Even Doob had said that in his landmark paper, "his mathematical intentions were not honorable!" Above all he was an engineer. In fact he was a truly ultimate type of engineer who could dissect the complexity of a problem into simple pieces and who would solve these pieces in order to then put them together to reconstruct the full solution to the original problem. This is the essence of Shannon's legacy.

GOLOMB'S PUZZLE COLUMN™

Numerical Oddities

Solomon W. Golomb

1) Find integers a, b, c with $1 < a < b < c$ such that $a!\ b! = c!$ (Two solutions.)

2) Find integers k and n with $1 < k < n$ such that $k! + 1 = n^2$. (Three solutions.)

3) In each of these cases, find two primes p and q, both <17, and different in each case, that satisfy:

   a) $p^2 - q^3 = 1$
   b) $p^3 - q^2 = 2$
   c) $p^7 - q^3 = 3$
   d) $p^3 - q^2 = 4$
   e) $p^3 - q^7 = 10$
   f) $2p^2 - q^2 = 1$
   g) $p^5 - 2q^2 = 1$

4) Find an integer $a > 1$ where $a^5 + 1$ is a k-digit number, $k > 1$, with all k of its digits the same. (Two solutions.)

5) Find $1 < k < n$ with $\binom{n}{k+2} = \frac{3}{2}\binom{n}{k+1} = 3\binom{n}{k}$.

6) Find $1 < k < n$ with $\binom{n}{k} = \binom{n-1}{k+1}$.

7) Find $1 < k < n$ with $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} = 2^k$.

8) Find $1 < k < n$ with $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3} = 2^k$.

9) Find integers $a > 1$ and $b > 1$ such that $(2^a - 1) + (3^b - 1) = m$, where m is a 5-digit prime such that $2^m - 1$ is a (Mersenne) prime.

10) Find a set $S = \{n + 1, n + 2, \ldots, n + k\}$ of k consecutive integers, $1 < k < n$, such that each element of S has a prime factor in common with at least one other element of S.

GOLOMB'S PUZZLE COLUMN™

Simple Theorems About Prime Numbers Solutions

Solomon W. Golomb

1) "The number of primes of the form $4n - 1$ is infinite."

Proof by contradiction. Suppose not. Then the primes of this form are a finite set, say $S = \{q_1, q_2, \ldots, q_n\}$ is all of them. Let $Q = q_1, q_2, \ldots, q_n$, and consider the number $K = 4Q - 1$. Since each q_i divides Q, none of them can divide K. If K is prime, since it is of the form $4n - 1$ and is bigger than any q_i, this contradicts the assumption that the list S is complete. If K,which is odd, is not prime, it must be a product of odd primes, each of which is either of form $4a - 1$ or $4a + 1$. If *all* the prime factors of K had form $4a + 1$, their product would also have form $4a + 1$, so at least one prime factor of K must have form $4a - 1$, a prime of form $4n - 1$ not on the assumed complete list S. □

2) "The number of primes of the form $6n - 1$ is infinite."

Proof by contradiction. Suppose not. Then the primes of this form are a finite set, say $T = \{r_1, r_2, \dots, r_m\}$ is all of them. Let $R = r_1, r_2, \dots, r_m$, and consider the number $L = 6R - 1$. Since each r_i divides R, none of them can divide L. If L is prime, since it is of the form $6n - 1$ and is bigger than any r_i, this contradicts the assumption that the list T is complete. If L is not prime, it can be divisible by neither 2 nor 3, and all other primes are either of the form $6a - 1$ or $6a + 1$. If *all* prime factors of L had form $6a + 1$, their product would also have form $6a + 1$, so at least one prime factor of L must have form $6a - 1$, a prime of the form $6n - 1$ which is not on the assumed complete list L. □

NOTE. These two are special cases of a far more general, and far deeper theorem of Dirichlet, which states that the arithmetic progression $\{an + b\}$, as n runs through all positive integers, takes on infinitely many prime values, so long as the obvious necessary condition, that a and b have no common prime factor, is satisfied.

3) "There are infinitely many "twin primes" if and only if there are infinitely many positive integers n NOT of the form are $6ab \pm a \pm b$, where a and b are positive integers, and all combinations of the ± signs are allowed."

Proof. Every twin prime except (3, 5) must be of the form $(6n - 1, 6n + 1)$. If either $6n - 1$ or $6n + 1$ factors, it must be a product of two numbers, say $6a \pm 1$ and $6b \pm 1$, since all numbers divisible by neither 2 nor 3 are of these forms. Thus, $6n \pm 1$ fails to be a twin prime if and only if $6n \pm 1 = (6a \pm 1)(6b \pm 1) = 36ab \pm 6a \pm 6b \pm 1$. Considering the two sides of this equation modulo 6, the ±1 terms must match, so $6n = 36ab \pm 6a \pm 6b$, and $n = 6ab \pm a \pm b$, for some choices of a and b, in order for $(6n - 1, 6n + 1)$ to fail to be a twin prime. □

NOTE. This was my first published result. It appeared as a Problem I submitted in the May, 1951, issue of the *American Mathematical Monthly*.

4) "Every positive integer is either of the form $n + \pi(n)$ or $p_n + n - 1$ (but not both), as n takes on all positive integers."

Proof. Imagine a land where the sales tax on n (cents) is $\pi(n)$, so the "total price" is $n + \pi(n)$. This tax increases by 1 (cent) whenever n is a prime, so if the pre-tax price is p_n, then the total price is $p_n + n$. But when the pre-tax price was $p_n - 1$, the tax was $n - 1$, for a total price of $(p_n - 1) + (n - 1) = p_n + n - 2$. So the numbers that never appear as the total price are precisely the numbers $p_n + n - 1$. □

NOTE. This result was the substance of my paper "The 'Sales Tax' Theorem", *Mathematics Magazine*, September–October, 1976.

5) "The ratio $\frac{n}{\pi(n)}$ takes on every positive integer value > 1 at least once, as $n > 1$ runs through the positive integers."

Proof. This result depends on only two properties of $\pi(n)$;

i) That $\pi(n + 1)$ is either $\pi(n)$ or $\pi(n) + 1$; and ii) that $\lim_{n\to\infty} \frac{\pi(n)}{n} = 0$. Now $\frac{2}{\pi(2)} = \frac{2}{1} = 2$, so the value 2 occurs and since $\lim_{n\to\infty} \frac{n}{\pi(n)} = \infty$, arbitrarily large value of $\frac{n}{\pi(n)}$ occur. Our task is to prove that no integer values of this ratio are skipped. For any $m \geq 2$, there is a unique largest prime $p_k = p_{k(m)}$ for which $\pi(p_k) = k \geq p_k/m$. Thus $m\pi(p_k) = mk \geq p_k$. Either $mk < p_{k+1}$ or $mk \geq p_{k+1}$. If $m_k < p_{k+1}$, since $p_k \leq mk$, $\pi(p_k) \leq \pi(mk) < \pi(p_{k+1})$, from which $\pi(mk) = k$, and $mk/\pi(mk) = m$, so that $n = mk$ is an integer for which $n/\pi(n) = m$.

If $mk \geq p_{k+1}$, then $\pi(p_{k+1}) = k + 1 > k = mk/m \geq p_{k+1}/m$, which contradicts the choice of p_k as the largest prime for which $\pi(p) \geq p/m$. □

NOTE. This result was the substance of my paper "On the ratio of N to $\pi(N)$", American Mathematical Monthly, January, 1962. I suspect (but have not proved) that the ratio $n/\pi(n)$ takes on every integer value $m \geq 2$ at least three times, for all $n > 1$.

The Students' Corner

Jonathan Scarlett

I recently came across an article on the arXiv that I believe could be of interest to students looking for new topics to work on, or thinking of pursuing a long-term career in information theory—http://arxiv.org/abs/1507.05941. It is written by a diverse range of leading experts, and overviews a number of future directions of information theory, including in particular applications to areas such as statistics, machine learning, biology, physics, and networked systems.

The field of information theory has been going strong for nearly 70 years now, and year after year we see an impressive range of problems and solutions appearing in our conferences and journals. Shannon theory itself offers a seemingly endless number of interesting open problems—of course varying greatly in difficulty—but research in new directions overlapping with other fields now seems to be becoming more popular than ever.

Having worked on "core" topics in Shannon theory during my PhD, I am now looking into some of these interdisciplinary connections. I find it quite fascinating how many statistical problems can be interpreted as source or channel coding problems—albeit oftentimes unconventional ones! To name just a few, I have seen and worked on the development of these connections in compressive sensing, group testing, graphical model selection, and community detection. With so many previous solutions in these areas relying on ad-hoc methods and incremental improvements, one should expect the fundamental limits provided by information theory to be of great importance.

On the other hand, as Shannon himself warned, we must be cautious and aware of the limitations of information theory in fields that it was not originally intended for. It is, of course, crucial to check the underlying assumptions and practical issues. One key difference in many interdisciplinary areas is that the analytically tractable models are often simply too crude, making it rather questionable as to how worthwhile it is to nail down the precise thresholds, as information theorists love to do. Despite this, in my view, there is often good reason to do this beyond the theoretical interest alone – the quest to achieve the thresholds often provides great insight and leads to new algorithms, as well as shining new light on existing ones.

Perhaps information theory will not impact these areas quite as much as it has impacted communication, nor provide results that are quite as elegant as Shannon's source and channel coding theorems. Nevertheless, I believe that with the right amount of care and the correct assumptions and interpretations, information theory can still play a fundamental role in ways that we never would have imagined previously.

From the Field

Yuval Kochman

In 2014, the IEEE Israel Section Information Theory Chapter received the IEEE Information Theory Society Chapter of the Year Award. Back then, we felt that the award reflects advance recognition of current efforts and anticipated activities no less than recognition for past accomplishments. Indeed, 2015 was a very busy year for us.

In April, we held the Information Theory Workshop in Jerusalem. It was the first major IT conference in Israel since ITW 1996 which was held in Haifa. The organizing Committee, led by the general co-chairs Yossi Sternberg, Rami Zamir and Jacob Ziv, consisted of members from all Israeli universities, as well as Alexander Barg from Maryland. The workshop was attended by about 160 participants, three quarters of them coming from abroad. We aimed at creating a rich experience, both from a technical perspective and from a cultural one. The technical program featured over 120 talks, including 5 plenary sessions. A special history session, led by our society historian Anthony Ephremides, featured David Forney, Robert Gallager, Sergio Verdu, Andrew Viterbi and Jacob Ziv. The fascinating stories concentrated on Claude Shannon, including the renowned first Shannon lecture that he himself gave in Ashkelon, Israel, in 1973. A video of the session can be found in the workshop website.

We made every effort to enable the workshop participants to have a taste of the country as well. That started from the location of the conference—Mishkenot Shaánanim, a historical building with a beautiful panorama of the Old City walls. Then, we had a walking tour inside the walls, introducing the diverse cultural heritage of the city. A full-day tour (in excruciating heat that was beyond our control) took the participants to the Judean Desert and the Dead Sea, culminating in a Beduin feast. The banquet was held in the Israel Museum, where dinner was preceded by a tour of the museum and followed by a musical show, with guest star Rami Zamir playing the keyboard.

After ITW, some of the participants continued directly to the Technion workshop on coding for emerging memories and storage technologies, which featured as speakers world-leading coding and information theorists, as well as Technion graduate students. The talks dealt with many of the fundamental subjects and useful techniques for storage reliability and efficiency: LDPC codes (Costello), constrained coding (Siegel), new channel models (Jiang, Jaggi), new data representations (Schwartz), fundamental limits (Verdu), efficient data recovery (Hollanti, Tamo). The workshop attracted a lot of interest inside and outside the Information-Theory community. In particular, the audience included a large representation from the thriving storage industry in Israel, as well as from related high-tech fields in computer and electrical engineering.

We at the Israel Chapter were delighted with the presence of so many visitors, and will be happy to host more in the future. We would like to take this opportunity and remind the community that every two years, IEEE Israel is holding a conference in the beautiful Red-Sea resort of Eilat. The conference always includes some lively IT sessions, as well as ones in neighboring fields such as signal processing and communications. The next IEEEI conference will be held around November-December 2016, and all of you are welcome!

President's Column *continued from page 1*

information theory. The committee has enlisted the creative talent of our student community to create posters telling students what they love about information theory. The committee is also compiling and editing archival material to create a 3-minute video about Shannon for use in exhibits and events.

We are proud to announce that the current list of institutions that have signed on to hold Shannon Day events includes 20 universities, museums, and labs located in over a dozen countries. Please see the website https://en.wikipedia.org/wiki/Claude_Shannon where details are being posted as they emerge. And there is still time to get involved! For more information about adding your institution to our list, please contact Lav Varshney at varshney@illinois.edu.

The Broader Outreach committee has enlisted the help of Greg Wornell and Emre Telatar to petition IEEE to get Shannon's 1948 paper added to the IEEE honor roll of historical Milestones. The Shannon stamp petition (http://www.itsoc.org/about/shannons-centenary-us-postal-stamp) to the United States Postal Service (USPS) now has over 1000 signatures, and Ninoslav Marina has proposed a Shannon stamp for Macedonia. Christina Fragouli and Anna Scaglione are exploring the possiblity of creating a cartoon about Shannon's work for children. Eric Graves, Joerg Kliewer, Anand Sarwate, and Aaron Wagner are organizing a series of information theory discussions on Reddit AMA (http://www.reddit.com/r/science/wiki/scienceamaseries). The website enables science presentations and conversations for participants around the world, regularly drawing audiences of 1000 or more. Eren Sasoglu and Ardan Arac have proposed to Google a special Google homepage theme to mark Shannon's 100th birthday. Let's keep our fingers crossed that Google will make such a page and that billions of people will get to see it. Plans are also underway to contact local museums, radio shows, and newspapers with the hopes of garnering publicity for this important day.

The Conference Committee, under the leadership of Elza Erkip, has been working to support workshops and conferences that explore connections between information theory and other fields and to develop guidelines and procedures for soliciting and handling requests for support in the future. The Online Committee, chaired by Anand Sarwate, has been working to update our website, making it possible to support more multimedia content such as talks and educational materials and to support a more dynamic and interactive online community. The Newsletter editor, Michael Langberg, has established a series of Newsletter articles on topics that connect information theory to other fields. The first article in that sequence, written by Mark Braverman, Rotem Oshman, and Omri Weinstein and published in the September Newsletter, explores the connections between information theory and communication complexity.

Looking forward, the officers, led by our future President Alon Orlitsky and future 1st Vice President Ruediger Urbanke, are already discussing potential broad outreach efforts for future years. These activities are likely to involve educational initiatives that take advantage of the web's ability to harness the talents of individuals to reach anyone anywhere. But I will leave the fun of telling you about those ideas to our future president. Perhaps he will talk about them in an upcoming President's Column.

And so, I sign off on my final President's Column here. It has been an honor and a pleasure to serve as the 2015 President of the IEEE Information Theory Society. As you can see by the long list of names included in this and prior columns and can surmise from the even longer list of people who make this Society what it is, I have benefited from the help, generosity, guidance, and wisdom of a huge number of our Society's members. I am so grateful to all of them and all of you for your support.

As always, I am happy to hear from you. Please send me your thoughts at effros@caltech.edu.

ISIT 2015: Experiments in a Time of Change

Suhas Diggavi, Vijay Kumar, Pierre Moulin, David Tse, and Raymond Yeung

The field of information theory is undergoing a period of change, as existing areas become mature and efforts are made to explore new directions in new domains. In keeping with the spirit of the times, we the ISIT 2015 organizers decided to do a few experiments with our field's flagship conference, whose organization and format have remained largely unchanged for the past two decades. During the conference, we conducted a survey on what the participants thought of the experiments, and we got over 400 responses. The experiments include:

- introduction of semi-plenary sessions to highlight results that are of broad interest, particularly those that explore new directions;
  - 57% of the respondees indicated that they would definitely like to see the semi-plenary experiment continue, 29% said maybe.

- introduction of a new mobile app (on both iOS and Android platforms) to help participants navigate around the conference;
  - 89% of the respondees indicated that they would like to see a similar mobile app in future ISITs;

- introduction of an online (confidential) forum for TPC members to discuss papers during the review period;
  - Approximately 450 papers were part of this online discussion process, where multiple TPC members gave input after the review process. This was found to be very helpful in arriving at the final decisions; in many instances, the comments elicited were so detailed and valuable that with the permission of the relevant TPC members, these comments were passed on to the authors.

The semi-plenaries are the most significant change to the conference, and we will describe this experiment in more detail below.

Semi-plenary Sessions

Semi-plenary sessions were introduced for the first time in ISIT 2015. They were comprised of time slots where only two parallel sessions (instead of the nine as was the case in ISIT 2014) were scheduled. Each paper in these sessions was selected from the submissions pool and received the same 20 minutes for oral presentation and 5 pages in the Proceedings as all other papers in the conference. We were able to schedule the semi-plenary talks without an increase in the number of parallel sessions (this was kept at 9 as was the case in ISIT 2014), and without affecting the overall acceptance rate (which was similar to ISIT 2014).

The main motivation to try this experiment was to highlight work that explores new directions, presents new ideas and makes interesting connections. By highlighting such work, we hoped to encourage the information theory community to further broaden the boundaries of the field. The semi-plenary sessions were envisaged to promote such research. The semi-plenary sessions were also used to showcase papers that contain results that deserve exposure to a broader audience. The finalists for the 2015 IEEE Jack Keil Wolf ISIT Student Paper Award were also presented in the semi-plenary sessions.

Question	Yes	Somewhat	No
Did you feel that the semi-plenary sessions enhanced your ISIT 2015 experience?	46%	33%	20%
Question	Yes	Maybe	No
Would you like to see this experiment continue at a future ISIT?	57%	29%	15%

Overall, we believe that this was a successful experiment, as evidenced by the feedback we received from a pre-banquet survey that was conducted and which is presented in the table above. Of the roughly 750 registered participants for ISIT 2015, 411 took part in the survey.

A three-tier process was employed to select semi-plenary presentations: In the first stage, each reviewer was asked whether a paper should be considered for a semi-plenary (SP) session. The reviewer was permitted to choose from among 3 options, 'Yes', 'Maybe' and 'No' and this input was made available to the TPC member handling the paper. In the second stage, the TPC member was asked whether the paper should be considered for a SP session and was only permitted to provide a hard decision, 'Yes' or 'No'. In the third stage, the TPC Co-Chairs forwarded the papers that were recommended by the TPC members for inclusion in a SP session to a special 12-person sub-committee of the TPC, termed the SP Sub-Committee (SP-SC), and chaired by Rob Calderbank. Of the 63 papers that were recommended by the TPC members, we eliminated 3 that were co-authored by one of the TPC Co-Chairs or General Co-Chairs (to avoid any conflict of interest). Thus in all, 60 papers were forwarded to the SP-SC. This committee, composed of senior members of the IT society, carefully examined the papers and after thorough deliberations, recommended a set of 17 papers to be presented in the SP sessions.

In addition to these 17 papers, another 5 papers, which were 2015 IEEE Jack Keil Wolf ISIT Student Paper Award finalists, were presented in the SP sessions. The selection of the candidates for the finalists was handled by a separate sub-committee of the TPC (the JWA-SC) headed by Andi Loeliger. The committee, composed again of 11 senior members of the IT society, sent a rank-ordered recommendation of 10 papers to IT-Society Awards Committee Chair Alon Orlitsky. The IT-Society committee then selected the 5 finalists which were included in the semi-plenary sessions and which were clearly identified as Jack Wolf student paper award finalists. In this way, a total of 17 + 5 = 22 papers were presented in the SP sessions.

As with any other experiment, it takes a few iterations to get everything perfectly right. We believe that we got off to a good start but also that there is scope for improvement through refinement and iteration. Overall, our feeling is that the feedback on the semi-plenary experiment was largely positive. We very much hope that the community will continue to experiment at future ISITs. Our flagship conference must evolve as the field evolves.

Report on the Munich Workshop on Coding and Modulation (MCM 2015)

Organizers

Georg Böcherer, Gianluigi Liva, and Gerhard Kramer

The Institute for Communications Engineering (LNT) at the Technische Universität München (TUM) and the German Aerospace Center (DLR) organized a Munich Workshop on Coding and Modulation (MCM 2015) on July 30-31, 2015. The technical program comprised 19 talks by global leaders on the topic. On Thursday, July 30, the speakers were Giuseppe Caire, Rüdiger Urbanke, Rick Wesel, Michael Lentmaier, Rami Zamir, Jean-Claude Belfiore, Gottfried Ungerböck, Guido Montorsi, Albert Guillén i Fàbregas, and Georg Böcherer. On Friday, the speakers were Robert Fischer, Sebastian Cammerer, Marco Baldi, Johannes Huber, Erdal Arıkan, Enrico Paolini, Mark Flanagan, Stephan Pfletschinger, and Jossy Sayir. The talk topics included spatial coupling, polar codes, lattice codes, short codes, Reed-Solomon codes, high-order modulation, probabilistic shaping, and codes for MIMO, relaying, and synchronization. Doctoral students and postdocs from LNT and DLR presented posters. Over 70 researchers from academia and industry attended the event.

The social program on Thursday included a Bavarian dinner at the Spatenhaus, which is a traditional restaurant facing the Bavarian State Opera. On Friday after the workshop there was a guided tour of the Lenbachhaus that has a large collection of paintings from the Munich artist group Der Blaue Reiter. The program ended with a relaxed get-together at the Park Café beergarden in the botanic garden close to the main train station.

Funding for the workshop was provided by LNT, DLR, and the Alexander von Humboldt Foundation. The program, presentations, posters, and photos are available at the web address http://www.lnt.ei.tum.de/en/events/munich-workshop-on-coding-and-modulation-2015

Michael Lentmaier, Erdal Arikan, and Georg Böcherer enjoying a Maß.

Active discussion on coding and modulation.

Group photo of participants of MCM 2015.

Report on the Munich Workshop on Massive MIMO (MMM 2015)

Organizers

Stefan Dierks, Markus Jäger, Gerhard Kramer, Roy Timo

The Institute for Communications Engineering (LNT) at the Technische Universität München (TUM) organized a Munich Workshop on Massive MIMO (MMM 2015) on October 7, 2015. The technical program consisted of two talks by Massive MIMO pioneers Tom Marzetta (Bell Labs) and Erik Larsson (Linköping University), a 5G real-time demo by Berthold Panzner (Nokia), and a poster session with presentations by Stefan Dierks (TUM), Andrei Nedelcu (TUM), and Muhammad Bilal Amin (Nokia). Over 50 researchers from academia and industry attended the event.

The social program included a lunch with the speakers at Il Mulino, a local Italian restaurant.

Funding for the workshop was provided by LNT and the Alexander von Humboldt Foundation. The program, presentations, and photos are available at the web address http://www.lnt.ei.tum.de/en/events/munich-workshop-on-massive-mimo-2015/

Erik Larsson, Tom Marzetta, and Giuseppe Caire discuss the limits of Massive MIMO.

Berthold Panzner presenting Nokia's 5G demo to participants from academia and industry.

Report on the Mathematical Tools of Information-Theoretic Security Workshop, September 23–25, 2015

Huawei Mathematical and Algorithmic Sciences Lab, Paris, France

Organizers

(i) Vincent Tan (NUS, Singapore),
(ii) Matthieu Bloch (Georgia Tech-CNRS UMI 2958),
(iii) Merouane Debbah (Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies)

With the advent of inexpensive hardware for transmission and storage of information, information technologies have become an integral part of our modern society. Individuals not only rely on the services provided by these technologies for their daily communications, but also store and exchange an increasing amount of sensitive information, including medical records data, financial records, etc. Concurrently, the widespread use of social networks encourages end-users to share private information with often misperceived privacy guarantees. As the general public has become increasingly aware of the importance of privacy and confidentiality issues inherently associated to modern information systems, there have been significant research efforts in various scientific communities to understand the fundamental mechanisms required to secure communications, to quantify the amount of privacy and confidentiality that could be guaranteed, and to characterize the tradeoffs between privacy and utility in various settings. In particular, while the cryptography community had long spearheaded the design of cryptosystems, there have been renewed research efforts in this direction in the information theory, signal processing, theoretical computer science, and quantum information theory communities. In particular, these communities have developed abstractions of fundamental information processing tasks, which are amenable to *quantitative and fundamental analysis*. Specific successful examples of such approaches include: secure quantum communications and quantum key distribution; information-theoretic security and physical-layer security; differential privacy in theoretical computer science; the use of randomness extractors in theoretical computer science as a building block for cryptographic primitives.

However, many of the mathematical tools and techniques have remained confined to the scientific communities in which they have been developed. The purpose of this 3-day workshop was to provide a common venue for researchers in different communities to get together, interact, exchange views, and share ideas with one another. Approximately 80 participants, including faculty, industry researchers, postdocs, and students attended the workshop in beautiful Paris in September 22–25 2015.

The main themes of the workshop included information-theoretic security, coding techniques for security applications, differential privacy and the tradeoff between utility and privacy, and quantum information and its applications to security. We were pleased to have 25 invited speakers including 5 plenary speakers: Yingbin Liang, Frédérique Oggier, Zhenjie Zhang, Andreas Winter and Shlomo Shamai (Shitz). The participants got the opportunity to appreciate talks on diverse topics, which provided the ground for numerous interactions and exchanges. In particular, participants discussed and debated the similarities and differences between privacy and security. The participants also benefited from several coding-theoretic talks showing how to construct and evaluate the performance of practical codes. The audience was also greatly appreciative of the tutorial-style lecture by Andreas Winter on the basics of quantum Shannon theory and leading all the way to more advanced topics such as quantum key distribution.

We acknowledge the kind and generous support from the Merlion Programme, a joint Franco Singaporean collaboration, I'Agence Nationale de Ia Recherche (ANR), and Huawei Labs, without which the worshop would not have been possible.

In Memoriam: Oscar Moreno de Ayala (1946–2015)

Heeralal Janwa, P. Vijay Kumar and Andrew Z. Tirkel

We pay tribute here to an eminent coding theorist and mathematician, Oscar Moreno de Ayala, who sadly passed away on July 14, 2015. Oscar was born on January 5, 1946 in Camagüey, Cuba, to parents Eva Garcia and Oscar Moreno de Ayala. He moved from Cuba to Colombia in 1962 and then to Puerto Rico the following year. He obtained his Bachelor's degree in Mathematics at the Río Piedras campus of the University of Puerto Rico (UPR) in 1967 and his M.A. and Ph.D. degrees in Mathematics in 1968 and 1973, respectively, from the University of California at Berkeley (UCB). At UCB, Oscar was the first Ph.D. student in Mathematics of Elwyn Berlekamp. He returned to UPR in 1974 where he had previously served as a teacher, and became a Professor at that institution until his retirement in 2007. He was initially with the Department of Mathematics, but later moved to the Department of Computer Science, a department that he helped found.

Oscar's varied research interests included exponential sums, coding theory, graph theory and pseudorandom sequence design with applications to sonar, optical and wireless communication. His early work related to binary Goppa codes where he along with co-author Elwyn Berlekamp showed that extended double-error-correcting Goppa codes are cyclic. He had a strong interest in, and made significant contributions to, the theory of exponential sums. By making a connection with Goppa codes, Oscar and co-author Carlos Moreno established the true minimum distance of a large class of Goppa codes, thereby contributing, in part, to the solution of an open research problem appearing in the book on error-correcting codes by McWilliams and Sloane. Other major results by the same two co-authors included improvements of the Chevalley-Warning and the Ax-Katz theorems. Oscar had a similarly strong interest in the construction of low-correlation pseudorandom sequences for various applications. In connection with a class of arrays with application to sonar known as Costas arrays, Solomon W. Golomb conjectured the existence of a primitive quadratic over any finite field whose roots had trace 1, a conjecture subsequently shown to be true by Oscar in 1989 (JCT-A). Oscar was elected a Fellow of the IEEE in 1999, "for contributions to the theory of error-correcting codes and to the design of sequences".

All through his academic life, Oscar maintained a strong interest in promoting and mentoring undergraduate, masters, and doctoral students attending the UPR. He created the Gauss Research Lab that served as the venue for the mentoring of many future Puerto Rican Mathematicians and Computer Scientists. Many of today's Puerto Rican mathematicians including those working in prestigious research centers, owe much of their academic progress to the guidance and encouragement received from Oscar Moreno as well as the exposure to the rigorous scientific standards he exemplified. Professor Richard Tapia from Rice University, nationally known for his effort to promote participation of Hispanics in Applied Mathematics, once said that Oscar had done outstanding work in mentoring a new generation of Puerto Ricans to obtain a Ph.D. in Mathematics. Oscar was also a key member in the creation of the Intercampus Computer and Information Science and Engineering (CISE) Ph.D. program. Several members of the Department of Computer Science received their Ph.D. in CISE thanks to his direct supervision and/or mentorship of their thesis research.

In 1985, the Resource Center for Science and Engineering (RCSE) of the University of Puerto Rico System, with help from Oscar and other UPR scientists, received funding from the Experimental Program to Stimulate Competitive Research (EPSCoR) for the Jurisdiction of Puerto Rico; the EPSCoR Programs had been created by NSF in 1980. This funding permitted professors throughout the university system of the UPR, as well as other universities of the island, to finance and carry out research projects, with the cooperation and guidance of qualified scientists from RCSE. Oscar oversaw many of the projects that related to the field of mathematics. He was also the first to bring to the island parallel computing, through the acquisition and use of Alliant (1985), Paragon (1990) and Cray XD1 (2000) computers, the best in their respective eras.

In 1988, Oscar, in collaboration with the EPSCoR project, connected Puerto Rico to the NSFNET. During this period many scientists and researchers from across the world visited the island to collaborate on various math projects. One such well-known researcher, the late Professor Leon Henkin, summed it up very well during one of his visits to the island when he was asked "What do you think of the spirit of mathematics in Puerto Rico ?" He replied "Oscar is the spirit of mathematics in Puerto Rico!" In 2009, Puerto Rico's mathematics and scientific computation community dedicated its annual meeting, SIDIM XXIV in Río Piedras, to Oscar.

Oscar was also one of the pioneers of Puerto Rican Internet history. He founded and led the administration of the .pr domain name registry, a responsibility that he had carried on without interruption from 1989. Oscar touched many people in his life, comments from some of them are reproduced below. The authors would like to thank the many people who contributed to the writing of this article.

- "Oscar was my first PhD student in mathematics at UC Berkeley. He was highly motivated and very independent. He wrote a fine thesis on coding theory, and he co-authored a paper with me on a portion of his thesis in which I was also heavily involved. He then spent his entire career in Puerto Rico. I visited him there on at least three occasions, where it became clear to me that he had become a leader, both in education and in implementation of technologies and organizations to expand internet access.

  He established research connections with several coding theorists in southern California, including Prof. Solomon Golomb. Oscar made many visits there over the years." Elwyn R. Berlekamp.

- "Oscar Moreno was very inventive, and had a creative approach to research. He was always a pleasure to work with." Solomon W. Golomb.

- "Oscar's creativity in research and in leadership impressed me. He contributed strong results in Costas arrays, exponential sums, and several other areas." H. F. Mattson, Jr.

- "Oscar was a dynamo of research activity, and I will miss his visits to USC." Robert A. Scholtz.

- "Oscar had a great mathematical intuition combined with a wonderful and unique sense of humor. I always felt well in his company and will surely miss him." Tor Helleseth.

- "Oscar's empathy and kindness have been very important for me." Tom Hoeholdt.

- "I am saddened by his passing and disappointed that I missed him at the ISIT conference last month." Alexander Barg.

- "We will miss his dedication to the development of Mathematics and Computer Science at UPR. It is now up to all of us to honor his memory by continuing his pioneering work." Carlos I. González Vargas, (Dean, College of Natural Sciences, UPR, Río Piedras).

- "I met him in 1975, when he was my professor in the new course on Combinatorial Algorithms, then Math 350. Since then, he has been my professor, mentor, colleague and friend. He will be sorely missed, I will miss him calling us "mijito". His accomplishments will live on." Pedro Juan Rodríguez Esquerdo (Dean of Graduate Study and Research, UPR, Río Piedras).

- "Oscar Moreno is the father of mathematics research in Puerto Rico. He did not only produce high-quality research himself, he inspired many others to get involved and pursue research careers. We have lost an example of hard work and determination." Ivelisse Rubio Canabal.

- "He inspired other people to give their best." Domingo Gómez-Pérez.

- "Oscar was a friend, a research collaborator and a colleague at UPR-RP (since 1997). He became a part of our family, and we miss him dearly." Heeralal Janwa.

- "I learnt a great deal from Oscar and was looking forward to his planned visit to Bengaluru. He will be missed!" P. Vijay Kumar.

- His enthusiasm and passion for work was contagious, and if not for Oscar, I would have retired. I will miss him!" Andrew Z. Tirkel.

Bibliography

1. Carlos I. González Vargas, "Dean's message for Professor Oscar Moreno," July 17, 2015, http://http://natsci.uprrp.edu/.

2. Karina Cortes, "In Memory of Dr. Oscar Moreno de Ayala," https://ccnso.icann.org/about/oscar-moreno-21jul15-en.pdf.

3. Manuel Gomez, "Eulogy for Professor Oscar Moreno," (personal communication).

4. Carlos Carbonera, "The Summer of 1990," (as communicated to Dorothy Bollman).

5. Alberto Cáceres, "Oscar Moreno: el espíritu de las matemáticas en Puerto Rico," UPR-Dialogo, October 17, 2015.

In Memoriam: Victor K. Wei

Lolita Chuang, University of Illinois Urban-Champaign

Yu Hen Hu, University of Wisconsin—Madison

Yih-Fang Huang, University of Notre Dam

Ming-Ting Sun, University of Washington

Victor K. Wei (S'77–M'80–SM'93–F'95) passed away on October 17th 2015 in Hong Kong. He was 61. Victor was born and raised in Taipei, Taiwan. He received the B.S. degree in electrical engineering from National Taiwan University in 1976 and the Ph.D. degree in electrical engineering from the University of Hawaii, Manoa, in 1980.

Dr. Wei was a Member of the Board of Governors (1991–1994) of the IEEE Information Theory Society, an Associate Editor for Coding Theory (1989–1992) for the *IEEE Transactions on Information Theory,* and a Guest Editor for the Special Issue on Algebraic-Geometric Codes (Vol. IT-41, No. 6, November 1995). He was elected a Fellow of the IEEE in 1995 "*for his contributions to coding theory and its applications*".

From 1980 to 1983, he was with the Mathematical Research Center of Bell Laboratories, Murray Hill, NJ. From 1984 to 1994, he was with Bellcore, Morristown, NJ, and became Director of Communication and Computation Principles Research in 1987. As a director, Victor demanded high quality standards for his team's research and created an environment conducive to world-class research. In 1994, he joined the Department of Information Engineering at The Chinese University of Hong Kong at the rank of Professor until his retirement at the end of 2011. During this period, Victor focused his research on cryptography and cultivated a young generation of cryptographers for Hong Kong.

Victor was a brilliant researcher with a keen intellect, broad interests, and creative ideas. He has made note-worthy contributions to several fields, particularly coding theory and data compression. Among his notable contributions, his single-author paper on "generalized Hamming weights for linear codes" (*IEEE Transactions on Information Theory*, Vol. 37, No. 5, pp. 1412–1418, 1991) was considered a breakthrough in coding theory. He introduced this innovative concept, totally new at the time, in response to a challenging problem in information-theoretic security dealing with transmission over a wire-tap channel. Victor's idea on the generalized Hamming weights has had significant impacts to cryptography and data security. He also co-authored with Fan R.K. Chung and Jawad Salehi the well-cited paper, "Optical orthogonal codes—design, analysis and applications," (*IEEE Transactions on Information Theory*, Vol. 35, No. 3, pp. 595–604, 1989).

Victor's work in the 1980's was influential in proving that universal source coding algorithms could compress at the entropy rate of the source, a fundamental barrier. The paper that he co-authored with Bentley, Sleator, and Tarjan ("A locally adaptive data compression scheme," *Communications of the ACM*, Vol. 29, No. 4, pp. 320–330, 1986), for example, is directly related to the move-to-front algorithm (aka *book stack algorithm*). According to the SCI-JCR, the Science Citation Index iV Journal Citation Report (2003), Victor was among the top 250 most cited computer scientists in the 20 years prior to that. He continued to do research after retirement, working on topics that included the minimum rank distance of Gabidulin codes, as well as error control coding for random network coding. Prior to his passing, he had prepared two manuscripts entitled, "Revisiting the minimum rank distance of Gabidulin codes," and "An intractability approach to error control in random network coding."

Victor was a person with many talents and interests. Since his youth, he was ingenious, with a particular facility in mathematics. In the 1972 Joint College Entrance Examination in Taiwan, he attained the highest score, among more than 30,000 entrants. Victor was an unconventional and innovative thinker, a skillful player of both Go and Bridge, and he once made Chinese riddles from the names of his college classmates. Victor was also fun-loving and adventurous. He cherished outdoor activities and ran marathons during his graduate school years in Hawaii. In his later years, he became passionate about history. Victor was dearly loved and will be greatly missed by his colleagues, friends, and family.

Victor is survived by his wife, Betty, and two daughters, Francine and Chloe. A memorial service, *Celebrating Victor Wei's Life*, was held on November 6, 2015 at The Chinese University of Hong Kong.

Call for Nominations

(ordered by deadline date)

Thomas M. Cover Dissertation Award

The IEEE Information Theory Society Thomas M. Cover Dissertation Award, established in 2013, is awarded annually to the author of an outstanding doctoral dissertation.

NOMINATION PROCEDURE: Nominations and letters of endorsement must be submitted by **January 15, 2016**. All nominations should be submitted using the online nomination forms. Please see http://www.itsoc.org/cover-award for details.

IEEE Joint ComSoc/ITSoc Paper Award

The Communications Society/Information Theory Society Joint Paper Award recognizes outstanding papers that lie at the intersection of communications and information theory. Any paper appearing in a ComSoc or ITSoc publication during the preceding three calendar years is eligible for the award.

NOMINATION PROCEDURE: Nominations and letters of endorsement must be submitted by **February 15, 2016**. All nominations should be submitted using the online nomination forms. Please see http://www.itsoc.org/honors/comsoc-information-theory-joint-paper-award/comsoc-itsoc-paper-award-nomination-form for details. Please include a statement outlining the paper's contributions.

IEEE Information Theory Society Claude E. Shannon Award

The IEEE Information Theory Society Claude E. Shannon Award is given annually to honor consistent and profound contributions to the field of information theory.

NOMINATION PROCEDURE: Nominations and letters of endorsement must be submitted by **March 1, 2016**. All nominations should be submitted using the online nomination forms. Please see http://www.itsoc.org/shannon-award for details.

IEEE Information Theory Society Aaron D. Wyner Distinguished Service Award

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

NOMINATION PROCEDURE: Nominations and letters of endorsement must be submitted by **March 1, 2016**. All nominations should be submitted using the online nomination forms. Please see http://www.itsoc.org/wyner-award for details.

IEEE Fellow Program

Do you have a colleague who is a senior member of IEEE and is deserving of election to IEEE Fellow status? If so, please submit a nomination on his or her behalf to the IEEE Fellow Committee. The deadline for nominations is **March 1 2016**.

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 Fellow recommendations 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 Information Theory Society Paper Award

The Information Theory Society Paper Award is given annually for an outstanding publication in the fields of interest to the Society appearing anywhere during the preceding two calendar years. The purpose of this Award is to recognize exceptional publications in the field and to stimulate interest in and encourage contributions to fields of interest of the Society.

NOMINATION PROCEDURE: Nominations and letters of endorsement must be submitted by **March 15, 2016**. All nominations should be submitted using the online nomination forms. Please see http://www.itsoc.org/honors/information-theory-paper-award/itsoc-paper-award-nomination-form for details. Please include a statement outlining the paper's contributions.

IEEE Information Theory Society James L. Massey Research & Teaching Award for Young Scholars

The purpose of this award is to recognize outstanding achievement in research and teaching by young scholars in the Information Theory community. The award winner must be 40 years old or younger and a member of the IEEE Information Theory Society on January 1st of the year nominated.

NOMINATION PROCEDURE: Nominations and supporting materials must be submitted by **April 30, 2016**. All nominations should be submitted using the online nomination forms. Please see http://www.itsoc.org/honors/massey-award/nomination-form for details.

IEEE Awards

The IEEE Awards program pays tribute to technical professionals whose exceptional achievements and outstanding contributions have made a lasting impact on technology, society and the engineering profession. For information on the Awards program, and for nomination procedures, please refer to http://www.ieee.org/portal/pages/about/awards/index.html

CALL FOR PAPERS

2016 Australian Communications Theory Workshop (AusCTW'16)

Melbourne, Victoria
20 - 22 January 2016

General Co-Chairs

Jamie Evans
Monash University

Emanuele Viterbo
Monash University

Technical Program Committee

Phee Lep Yeoh (Chair)
University of Melbourne

Wibowo Hardjawana
University of Sydney

Yi Hong
Monash University

Min Li
Macquarie University

Robby McKilliam
University of South Australia

Lawrence Ong
The University of Newcastle

Parastoo Sadeghi
The Australian National University

Nan Yang
The Australian National University

Jinhong Yuan
University of New South Wales

Local Arrangements Chairs

Shuiyin Liu & Lakshmi Natarajan
Monash University

Finance & Registration Chairs

Katrina He & Rajitha Senanayake
Monash University

Website & Publicity Chair

Bhathiya Pilanawithana
Monash University

Steering Committee

Iain Collings
Macquarie University

Linda Davis
University of South Australia

Jamie Evans
Monash University

Alex Grant
Cohda Wireless

Rod Kennedy
The Australian National University

Lars Rasmussen
KTH Royal Institute of Technology

Graeme Woodward
University of Canterbury

Workshop Announcement

Monash University is pleased to host the 16th Australian Communications Theory Workshop. The workshop will bring together researchers and post-graduate students in physical layer communications and information theory for two and a half days of technical presentations, tutorials and networking. Past workshops have provided formal and informal environments to successfully foster collaborative research.

Invited Talks

Invited talks will be given by leading researchers and outstanding graduate students.

Contributed Papers

Papers presenting original and unpublished contributions are solicited (maximum length is 6 pages). All contributed papers will be subject to peer review. Topics of interest include, but are not limited to:

- coded modulation
- coding theory and practice
- communication systems
- channel modelling
- detection and estimation
- ultra-wide band communications
- OFDM & DMT processing techniques
- blind signal separation techniques
- information theory and statistics
- network coding
- compressed sensing
- iterative decoding algorithms
- multiuser detection
- cross-layer PHY-MAC-NET optimisation
- DSP for communications
- molecular, biological and multi-scale communications

We are pleased to announce technical co-sponsorship by the IEEE Information Theory Society ACT Section Chapter. All accepted papers are to be presented as posters during the conference. Accepted and appropriately presented papers will appear in full in the conference proceedings and will be submitted to IEEE*Xplore* for archival. Please see conference website (www.ausctw.org.au) for paper submission details.

Non-Peer Reviewed Contributions

To facilitate maximum participation, all attendees are invited to present a poster at the workshop for which only an abstract need be submitted. Abstracts are *not* subject to peer review and appear in the workshop *book of abstracts*. Please see conference website (www.ausctw.org.au) for abstract submission details.

2016 Australian School of Information Theory

The 2016 Australian School of Information Theory will be held at the same venue on 17-19 January 2015. Please see conference website for registration details (www.ecse.monash.edu.au/staff/ejamie/AusSIT16).

Key Dates

Paper submission deadline:
Friday, October 16, 2015

Notification of decisions:
Friday, November 20, 2015

Camera-ready papers due:
Friday, December 18, 2015

Poster abstracts due:
Friday, December 18, 2015

Early registration closes:
Friday, January 8, 2015

MONASH University IEEE

Call for Papers

2016 International Zurich Seminar on Communications

March 2 – 4, 2016

The 2016 International Zurich Seminar on Communications will be held at the Hotel Zürichberg in Zurich, Switzerland, from Wednesday, March 2, through Friday, March 4, 2016.

High-quality original contributions of both applied and theoretical nature are solicited in the areas of:

- Wireless Communications
- Information Theory
- Coding Theory and its Applications
- Detection and Estimation
- MIMO Communications
- Optical Communications
- Fundamental Hardware Issues
- Network Algorithms and Protocols
- Network Information Theory and Coding
- Cryptography and Data Security

Invited speakers will account for roughly half the talks. In order to afford the opportunity to learn from and communicate with leading experts in areas beyond one's own specialty, no parallel sessions are anticipated. All papers should be presented with a wide audience in mind.

Papers will be reviewed on the basis of a manuscript (A4, not exceeding 5 pages) of sufficient detail to permit reasonable evaluation. Authors of accepted papers will be asked to produce a manuscript not exceeding 5 pages in A4 double column format that will be published in the Proceedings. Authors will be allowed twenty minutes for presentation.

The deadline for submission is **September 27, 2015**.

Additional information will be posted at

`http://www.izs.ethz.ch/`

We look forward to seeing you at IZS.

Amos Lapidoth and Stefan M. Moser, Co-Chairs.

PRINCETON

Call for Papers CISS 2016

50th Annual Conference on Information Sciences and Systems

March 16, 17, & 18, 2016

Princeton University - Department of Electrical Engineering

and Technical Co-sponsorship with

IEEE Information Theory Society

Authors are invited to submit previously unpublished papers describing theoretical advances, applications, and ideas in the fields of: information theory, coding theory, communication, networking, signal processing, image processing, systems and control, security and privacy, machine learning and statistical inference.

Electronic submissions of up to 6 pages (in Adobe PDF format) including 3-4 keywords must be submitted by **December 15, 2015**. Submissions should be of sufficient detail and length to permit careful reviewing. Authors will be notified of acceptance no later than **January 11, 2016**. Final manuscripts of accepted papers are to be submitted in PDF format no later than **January 25, 2016**. These are firm deadlines that will permit the distribution of Electronic Proceedings at the Conference. Accepted Papers will be allotted 20 minutes for presentation, and will be reproduced in full (up to six pages) in the conference proceedings. IEEE reserves the right to exclude a paper from post-conference distribution (e.g., removal from IEEE Xplore) if the paper is not presented at the conference.

For more information visit us at: http://ee-ciss.princeton.edu/

CONFERENCE COORDINATOR

Lisa Lewis
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544
Phone: (609) 258-6227
Email: CISS@princeton.edu

PROGRAM DIRECTORS

Prof. Mung Chiang
Prof. Peter Ramadge
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544

IMPORTANT DATES

Submission deadline:
December 15, 2015

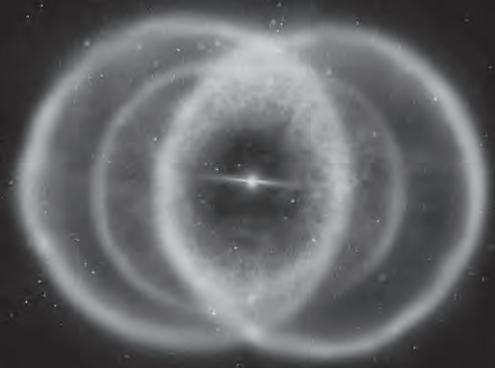
Notification of acceptance:
January 11, 2016

Final manuscript due:
January 25, 2016

Nexus of Information and Computation Theories

Institut Henri Poincaré
Spring 2016 Thematic Program
http://csnexus.info

January 25 - April 1, 2016
Paris, France



About the Program

Recently, a number of advances in the theory of computation have been made by using information-theoretic arguments. Conversely, some of the most exciting ongoing work in information theory has focused on problems with a computational component. The primary goal of this three-month IHP thematic program is to explore the rich interplay between information theory and the theory of computation, and ultimately create new connections and collaborations between both scientific communities.

- **Core of the Program:** eight weeks, split across four major themes (see below for details).
- **Central Workshop (February 29 - March 4):** broadly spanning the interface between CS and IT.
- **Tutorial Week (January 25 - 29) at CIRM (Marseille):** designed for students, but all are welcome.

Registration

Researchers and students who are considering attending any part of the program **must register on the website as soon as possible.** Registration is free but mandatory given the limited number of places. During the registration process, one can choose amongst the thematic weeks and/or the central workshop.

Program Organizers

Mark Braverman (Princeton)
Bobak Nazer (Boston University)
Anup Rao (University of Washington)
Aslan Tchamkerten (Telecom Paristech)

About IHP

The Henri Poincare Institute (IHP) is a research institute dedicated to mathematics and theoretical physics. Each quarter, the institute hosts a thematic program that brings together researchers from a particular discipline to foster the exchange of ideas.

ihp Institut Henri Poincaré

Theme Organizers

Distributed Computation (February 1 - 12)
Péter Gàcs (Boston Univeristy)
János Körner (Sapienza University of Rome)
Leonard Schulman (Caltech)

Fundamental Inequalities (February 15 - 26)
Kasper Green Larsen (Aarhus University)
Babak Hassibi (Caltech)
Iordanis Kerenidis (University Paris Diderot 7)
Raymond Yeung (Chinese University of Hong Kong)

Inference Problems (March 7 - 18)
Amit Chakrabarty (Dartmouth College)
Andrew McGregor (UMass Amherst)
Henry Pfister (Duke University)
Devavrat Shah (MIT)
David Woodruff (IBM)

Secrecy and Privacy (March 21 - April 1)
Prakash Narayan (University of Maryland)
Aaron Roth (University of Pennsylvania)
Anand Sarwate (Rutgers University)
Vinod Vaikuntanathan (MIT)
Salil Vadhan (Harvard University)

CALL FOR PAPERS

2016 IEEE Radar Conference

Enabling Technologies for Advances in Radar

www.radarconf16.org

Key Dates

Paper Summaries Due: 14 November 2015
Notification of Acceptance: 04 January 2016
Paper Submission Due: 05 February 2016
2016 Radar Conference: May 2 – 6, 2016
Loews Philadelphia Hotel, 1200 Market Street, Philadelphia, Pennsylvania, USA

Guide to Paper Submissions

Authors are required to submit a three to four page (inclusive of figures) summary. Electronic submission is required in Adobe pdf format. The cover page must include the title, names of authors (with the contact author identified), organizational affiliation, address, telephone and fax numbers, and email addresses. Authors are permitted to indicate paper suitability for a poster format presentation. Student papers (two to four pages) are also strongly encouraged to be submitted.

All papers must be electronically submitted to the Technical Program Chairman at the *radarconf16.org* web site (available to upload not later than 90 days before the deadline). The deadline for submission of summaries is 14 November 2015. Authors will be notified of acceptance by 4 January 2016, and will receive instructions and forms for publication at that time. Authors will be limited to orally presenting at most two papers at the conference. Your electronically submitted papers in final form will be required by 5 February 2016. They are limited to six pages inclusive of text, figures, and tables. If applicable, government approval for publication as an *unclassified, public-release* paper will also be required with the final paper submission.

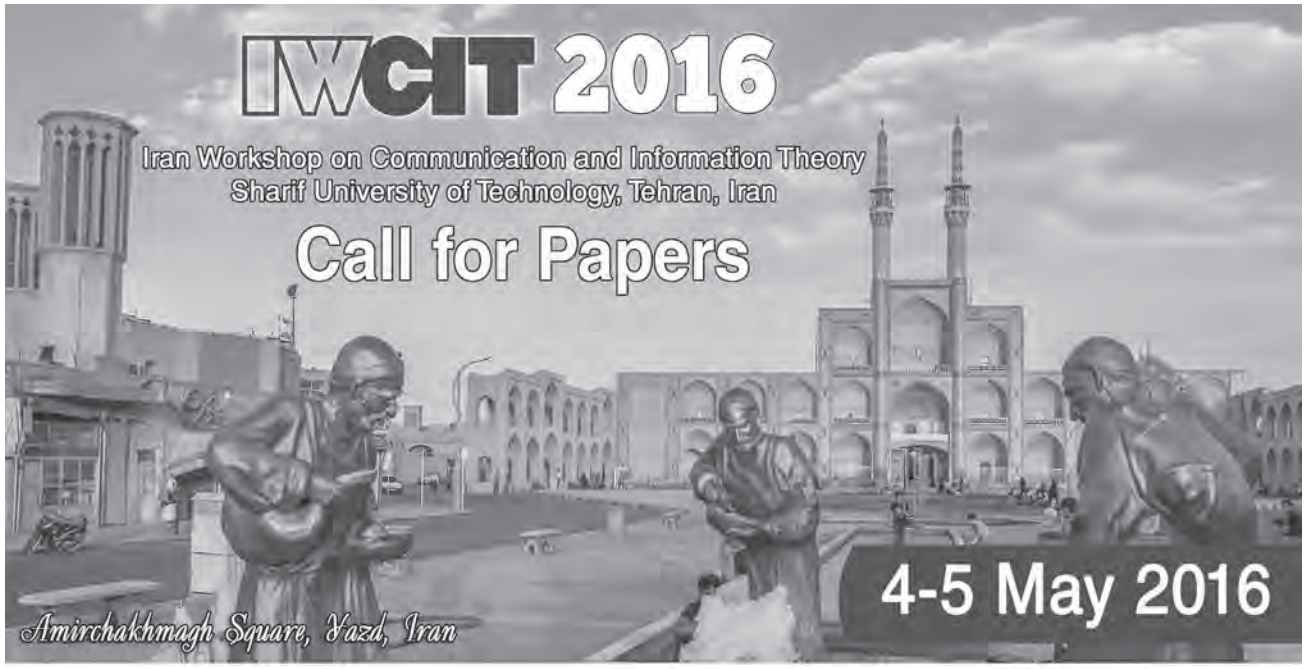
Main tracks

A list of topics within these tracks is on the web site <*www.radarconf16.org*>.

Authors can indicae preference for a track.

Component & Subsystem Development
Radar Signal & Data Processing
Antenna Technolog
Phenomenology
Radar Systems
Emerging Technologies

Technical Program Chair
David J. Farina
Lockheed Martin MST
djfarina@radarconf16.org



The fourth Iran Workshop on Communication and Information Theory will take place at Sharif University of Technology, on May 4th and May 5th 2016, Tehran, Iran. Interested authors are encouraged to submit their original and previously unpublished contributions to the following fields. This conference highly appreciates interdisciplinary related research not necessarily included below.

Shannon Theory
- Complexity theory
- Information theoretic security
- Multi-terminal information theory
- Quantum information theory

Communication Theory
- Cognitive radio systems
- Cooperative communications
- Network resource sharing and scheduling
- Molecular and Nano communications
- Optical and Quantum communication theory

Coding Theory
- Compressed sensing
- Data compression
- Network coding

Applications of Information Theory
- Information theoretic learning
- Information theory and data mining
- Information theory and signal processing
- Information theory and statistics
- Information theory in biology
- Information theory in networks
- Information theory in practice

Important Dates:
- **Paper Submission: January 11th, 2016**
- **Notification of Acceptance: March 15th, 2016**
- **Camera Ready Submission: April 15th, 2016**

General Chairs:
- Aref, M. R.
Sharif University of Technology
- Sharafat, A. R.
Tarbiat Modares University

Technical Program Chair:
- Salehi, J. A.
Sharif University of Technology

Executive Chairs:
- Gohari, A.
Sharif University of Technology
- Seyfe, B.
Shahed University

Contact Us :
- Email:
info@iwcit.org
iwcit@sharif.ir
- Address:
Secretariat of IWCIT 2016 Room 503 Dept. of Electrical Engineering Sharif University of Techology Tehran, Iran
Tel : +98 21 66165910

IEEE IRAN SECTION

WWW . IWCIT . ORG

2016 IEEE International Symposium on Information Theory

Barcelona, Spain | July 10-15, 2016

Call for papers

The 2016 IEEE International Symposium on Information Theory will take place in Barcelona, Spain, from July 10 to 15, 2016. A lively city, known for its style, architecture, culture, gastronomy and nightlife, Barcelona is one of the top tourist destinations in Europe. Interested authors are encouraged to submit previously unpublished contributions from a broad range of topics related to information theory, including but not limited to the following areas:

Topics

Big Data Analytics
Coding for Communication and Storage
Coding Theory
Communication Theory
Complexity and Computation Theory
Compressed Sensing and Sparsity
Cryptography and Security
Detection and Estimation
Emerging Applications of IT
Information Theory and Statistics
Information Theory in Biology
Network Coding and Applications
Network Information Theory
Pattern Recognition and Learning
Physical Layer Security
Quantum Information and Coding Theory
Sequences
Shannon Theory
Signal Processing
Source Coding and Data Compression
Wireless Communication and Networks

Researchers working in emerging fields of information theory or on novel applications of information theory are especially encouraged to submit original findings.

The submitted work and the published version are limited to 5 pages in the standard IEEE conference format. Submitted papers should be of sufficient detail to allow for review by experts in the field. Authors should refrain from submitting multiple papers on the same topic.

Information about when and where papers can be submitted will be posted on the conference web page. The paper submission deadline is January 24, 2016, at 11:59 PM, Eastern Time (New York, USA). Acceptance notifications will be sent out by April 3, 2016.

We look forward to your participation in ISIT in the centennial year of Claude Shannon's birth.

General Co-Chairs
Albert Guillén i Fàbregas
Alfonso Martinez
Sergio Verdú

TPC Co-Chairs
Venkat Anantharam
Ioannis Kontoyiannis
Yossef Steinberg
Pascal Vontobel

Finance
Stefan Moser

Publications
Tobias Koch

http://www.isit2016.org/

Conference Calendar

DATE	CONFERENCE	LOCATION	WEB PAGE	DUE DATE
December 6–10, 2015	**IEEE GLOBECOM**	San Diego, California, USA	http://globecom2015.ieee-globecom.org	Passed
December 7–8, 2015	**2015 Munich Workshop on Information Theory of Optical Fiber (MIO 2015)**	Munich, Germany.	http://www.lnt.ei.tum.de/en/events/munich-workshop-on-information-theory-of-optical-fiber-2015/	—
December 14–16, 2015	**IEEE Global Conference on Signal and Information Processing (GlobalSIP)**	Orlando, Florida, USA	http://2015.ieeeglobalsip.org	Passed
December 15–17, 2015	**DIMACS Workshop on Network Coding: the Next 15 Years**	Rutgers University, New Jersey, USA	http://dimacs.rutgers.edu/Workshops/Next15/	—
January 10–12, 2016	**ACM-SIAM Symposium on Discrete Algorithms**	Arlington, Virginia, USA	http://www.siam.org/meetings/da16/index.php	Passed
January 20–22, 2016	**Australian Communications Theory Workshop (AusCTW)**	Melbourne, Australia	http://www.ausctw.org.au	October 16, 2015
January 25–April 1, 2016	**IHP Thematic Program on the Nexus of Information and Computation Theories**	Paris, France	http://csnexus.info	—
March 2–4, 2016	**2016 International Zurich Seminar on Communications**	Zurich, Switzerland	http://www.izs.ethz.ch	September 27, 2015
March 16–18, 2016	**50th Annual Conference on Information Sciences and Systems**	Princeton University	http://ee-ciss.princeton.edu	December 15, 2015
May 2–6, 2016	**IEEE Radar Conference: Enabling Technologies for Advances in Radar**	Philadelphia, Pennsylvania, USA	http://radarconf16.org/#/	Passed
May 4–5, 2016	**4rd Iran Workshop on Communication and Information Theory (IWCIT)**	Sharif University of Technology, Tehran, Iran.	http://www.iwcit.org	January 11, 2016
May 9–13, 2016	**14th International Symposium on Modeling and Optimization in Mobile, Ad-Hoc, and Wireless Networks (WiOpt)**	Arizona State University Tempe, Arizona, USA	http://www. wi-opt.org	December 18, 2015
July 10–15, 2016	**2016 IEEE International Symposium on Information Theory**	Barcelona, Spain	http://www.isit2016.org	January 24, 2016,

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