” Transferring information in networks ”

EMMANUEL ABBE

SWISS FEDERAL INSTITUTE OF TECHNOLOGY, EPFL, LAUSANNE

Transferring information in networks

Emmanuel Abbe

Introduction

"The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point." This is how Claude Shannon describes the problem of communication in his paper "A Mathematical Theory of Communication", in 1948, (10). In his landmark paper, Shannon develops what will be later called "information theory", providing mathematical models for communication channels and establishing laws governing reliable transmission of information, much like Newton's theory establishes laws governing the motions of mass. Shannon's paper proposes communication schemes that entirely based on binary digits (*bits*), and is a building block to the development of the digital world in which we currently live.

In Shannon's definition, communication is described from one point to another. This is a proper model, as exemplified by voice, telecommunications or emails. In that model, the difficulty is to transfer information reliably despite the external noise, which corrupts the message. For example, in cellular telecommunication, a message may be carried by electromagnetic waveforms in the sky, may cross oceans through optical fiber cables, and no matter how much noise is added to it, must be reproduced to its destination in less than 0.2 second. To meet this challenge Shannon introduced the idea of coding information, to protect information against the noise, and showed that for a given amount of noise, reliable communication is achievable up to a maximal data rate. It took a lot of effort to researchers and engineers to construct codes that are efficient and that achieve Shannon's limit.

Today a new challenge is emerging. Communication is still about reproducing at one point a message selected at another point, but the number of such points is becoming very large. More and more people are using communication media and the amount of information flowing around the world is exploding. As a result, multi-user or network communication is becoming a new paradigm for the problem of transferring information. The idea of communication networks is to look at multiple users exchanging information not just as a collection of individual point-to-point links (as described previously), but to consider them collectively. The goal is to distinguish the perturbation coming from the interference created by the users on themselves and the perturbation coming from the external noise. Since interference is more structured than noise and comes from signals designed by the users, it can be exploited and should not be treated as noise. This requires an extension of the basic model described for point-to-point communication (9), and raises a new challenge: how to efficiently communicate over networks?

In this contribution, we propose a new coding technique that is optimal (reaching the largest possible data rate) and low-complexity (practical) for a class of network communication problems. The technique is among the first ones to achieve these goals while also providing mathematical guarantees to its

performance. The approach builds on the recently discovered polarization technique, proposed first by Arikan in 2008, (7), and extends the method to a class of network problems. It is shown that for networks, rather than a polarization to two extremal noise configurations, several extremal noise-interference configurations take place.

Single-user polarization technique

The polarization technique emerged in (7,8) as a technique to construct channel codes. There is however a more general phenomenon behind this technique, which we now informally describe. Consider a society formed by individuals who are initially all equally wealthy. Consider now picking pairs of individuals and transferring small amounts of wealth within each pair (the problem is particularly interesting for limited interactions and bounded wealth capacities). Finally consider iterating these local interactions many times in the society. How will the society look like in time? The polarization phenomenon shows that for certain ways of selecting the pairs and transferring the wealth, the society eventually becomes *polarized*. Namely, two groups emerge, one containing nearly all the wealth (uniformly spread out) and the other none. In addition, the polarization technique provides an efficient way to reach this configuration.

If from a social point of view this may not be a desirable situation, for engineering applications, this represents an ideal outcome.
We now explain this with a mathematical formulation developed in (1). Given a random process (the process must satisfy some assumptions ensuring a certain amount of independence in it, cf. (1) for more details), there is a so-called polar transformation which maps this process into a new random process having two distinct groups of components, one containing nearly all the randomness and the other none (i.e., almost deterministic). Hence, if this process represents a noise process, we have a transformation to create a new process, which has in total the same amount of randomness, but which has now some noise free components, while other components are maximally noisy. The advantage of such a configuration is that it is much easier to handle the second process: one can simply ignore the components containing the noise and focus on the noise free components, which are very reliable. In other words, we have swept away the noise to create clean components.
Moreover, the `magic' of the polarization technique is that the transformation used is indeed *low-complexity*, deterministic and invertible.

Multi-user polarization technique

The networks considered in (3) are multiple access channels (MACs). These are channels where several users would like to convey independent messages to a single receiver. This task is difficult because each user suffers from both external noise and from interference created by themselves. At the moment, treating inference efficiently and understanding wireless networks represent one of the major challenges in communication. Interference is indeed the component which render networks so complex even for fundamental questions such as determining the capacity regions (6). Multiple accessing is used in CDMA, TDMA

and FDMA, and is a key component of many modern wireless communication systems (which are part of today's standards for mobile communication).

The papers (1,2,3) provide coding schemes which are among the first to be both optimal (in data rate) and of manageable complexity for network information theory problems. The technique developed in (3) proposes a transformation which decouples the users by removing both the external noise and the interference in a MAC. The key tool used in this procedure is the polarization technique described previously. The basic polarization technique transforms independent uses of a single-user noisy channel into successive uses of single-user channels that are of two polarized types: noise-free or pure noise. However, as mentioned previously, in a network problem the user are also subject the to interference in addition to suffering from external noise. Following the polarization spirit, it is shown in (3) that independent uses of a MAC can be transformed into successive uses of, not *polarized* because they are no longer only two, but *extremal* MACs. The key point is that these extremal MACs are all trivial in terms of *both* noise and interference.

This opens the door to a new vision of networks, suggesting their decomposition into extremal networks, for which we can easily handle information transfer.
Indeed, this approach has shown to generalize successfully to other types of networks than the MAC. In (1), a similar technique is developed on the multiple-user source compression problem, constructing low-complexity schemes that allow to optimally compress correlated sources. This is also referred to as the Slepian-Wolf coding problem. From a mathematical point of view, in the latter setting, the transformation allows to extract not only the randomness out of random vectors, but also the dependencies between multiple correlated vectors, leaving us again with trivial distributions for the transformed sources. The approach has later been used for the broadcast channel, but as opposed to the MAC, we only understand special cases of broadcast channels at the moment with the technique. Finally, in (2), the technique is used for the generation of secret keys, in the setting where several users which to generate a secret key while an eavesdropper listens to their communication. There are of course many more network problems to explore with the proposed technique.

Mathematical development

We now present the idea behind the polarization technique, from which we will build our technique for network problems.

Let X_1 and X_2 be two i.i.d. Bernoulli(p) random variables, i.e., binary random variables which take value 1 with probability p. The total entropy of (X_1, X_2) is $2h(p)$, where $h(p)$ denotes the entropy of a Bernoulli(p) random variable. Consider now taking a linear transformation of (X_1, X_2) over the binary field, to obtain $(Y_1, Y_2) = (X_1, X_2)\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.
Since the matrix above is invertible, the total entropy of (Y_1, Y_2) equals that of (X_1, X_2), namely $2h(p)$. However, the entropy of (Y_1, Y_2) is now given by

$H(Y_1)+H(Y_2|Y_1)$, since Y_1 and Y_2 are not independent (this is the chain rule formula).
Hence, $2h(p) = H(Y_1)+H(Y_2|Y_1)$. Note that $H(Y_1) \geq h(p)$, since Y_1 is obtained by processing X_1, which has entropy $h(p)$. As a consequence, $H(Y_2| Y_1) \leq h(p)$. This means that Y_1 is now more entropic than X_1, and if Y_1 is observed, Y_2 is less entropic than X_2 (in average over the distribution of Y_1). In other words, out of two equally random variables, we have created one random variable which has more entropy and one random variable which has less entropy upon observing the other variable. The idea behind polarization is to then iterate again and again this procedure to create events which become eventually all very entropic or very deterministic.

Mathematically, if n is a power of two, X^n is an i.i.d. Bernoulli(p) sequence, and if Y^n is given by $Y^n = X^n G_n$, where

$$G_n = \left(\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right)^{\otimes \log_2(n)},$$

and where $\otimes$ denotes the Kronecker product, the chain rule yields $nh(p) = H(X^n) = \sum_{i=1}^{n} H(Y_i \mid Y^{i-1})$ and most of the terms $H(Y_i \mid Y^{i-1})$ become either close to either 0 or 1. Formally, the following result proved in (7) holds.

For any $\varepsilon>0$, $|\{ i =1,\ldots,n : \varepsilon \leq H(Y_i \mid Y^{i-1}) \leq 1-\varepsilon\}| = o(n)$.

Figure 19 illustrates this result for a fixed dimension, showing that the conditional entropies concentrate around 0 and 1 except for a small fraction of indices.

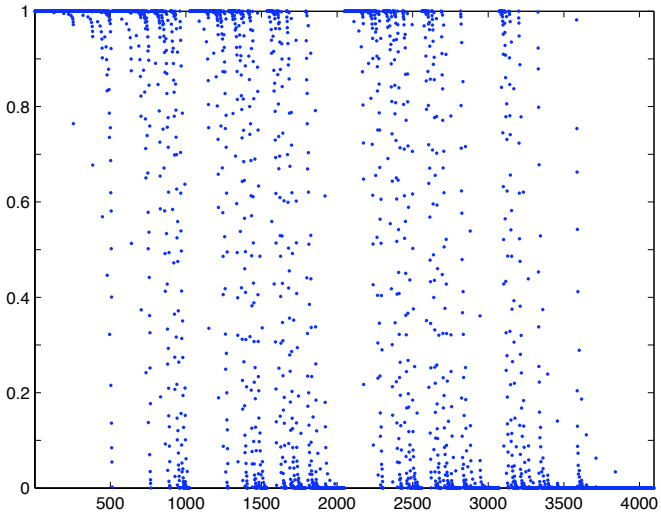

Figure 1: plots of $H(Y_i \mid Y^{i-1})$ for $i=1,\ldots,n=4096$, when $H(p)=1/2$.

This result can be used to communicate over a binary symmetric channel. For a BSC channel, the output Y^n equals the input x^n plus a random sequence Z^n which is i.i.d. Bernoulli(p) (and does not depend on x^n). By applying G_n to Y^n, one obtains $G_n\, Y^n = G_n\, x^n + G_n\, Z^n$. Note that $W^n = G_n\, Z^n$ is precisely the transformed sequence studied previously in the source polarization and the components of W^n can be decomposed into two sets: the high and low conditional entropy components. By arranging $G_n\, x^n$ to have a fixed value known to the decoder, say zero, on the components where W^n has non-negligible conditional entropy, one can learn these components and hence decode the entire vector W^n. This allows to recover Z^n, hence the code word x^n. Note that the number of components which are not fixed in x^n is roughly $n(1-h(p))$, i.e., this scheme achieves the Shannon capacity of the BSC channel. Moreover, the decoding complexity is only $O(n\log(n))$. This follows from the Kronecker structure of G_n, which allows the use of a divide and conquer algorithm for the decoding.

Now, if we consider not one source but several sources which are correlated, the problem becomes a network information theoretic problem. One case use the method described previously for each source individually, but then we are not exploiting the dependency between the sources, i.e., we are not considering the "network" aspect of the problem. In (1), the following theorem is shown, which takes into account the dependencies among multiple sources.

Theorem. Let n be a power of two, X be a $m \times n$ random matrix with i.i.d. columns having distribution Q, and Y be given by $Y = X\, G_n$, where

$$G_n = \left(\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right)^{\otimes \log_2(n)},$$

and where $\otimes$ denotes the Kronecker product. For any $\varepsilon > 0$ and for any i between 1 and n, there exists a matrix A_i such that

(1) $H(A_i Y_i \mid Y^{i-1}) < \varepsilon$

(2) $\left\{ x_1, \ldots, x_m : 0 \le \sum_{j \in S} x_j \le H(Y_i[S] \mid Y^{i-1}) \right\}$

This theorem ensure that the conditional entropies $H(Y_i \mid Y^{i-1})$ are "special", in the sense that they are (almost) integer-valued and behave like the entropy of a linear transformation of pure bits. This implies that all the randomness in X can be distilled in a subset of components of Y which behave like i.i.d. Bernoulli(1/2) components. In particular, for any subset S of integers between 1 and m, the conditional entropy $H(Y_i[S] \mid Y^{i-1})$ defines in the limit an *extremal entropic polytope*

$$\left\{ x_1, \ldots, x_m : 0 \le \sum_{j \in S} x_j \le H(Y_i[S] \mid Y^{i-1}) \right\}$$

which is special by having corner points on the hyper-cube, as illustrated for $m=3$ in Figure 2. Hence, there exists always a subset of components that can be made deterministic, provided that the other components are revealed, and this leads to

a method for compressing correlated sources or for removing both noise and interference in a multiple access channel with additive correlated noise. The general case of multiple access channels is treated in (3).

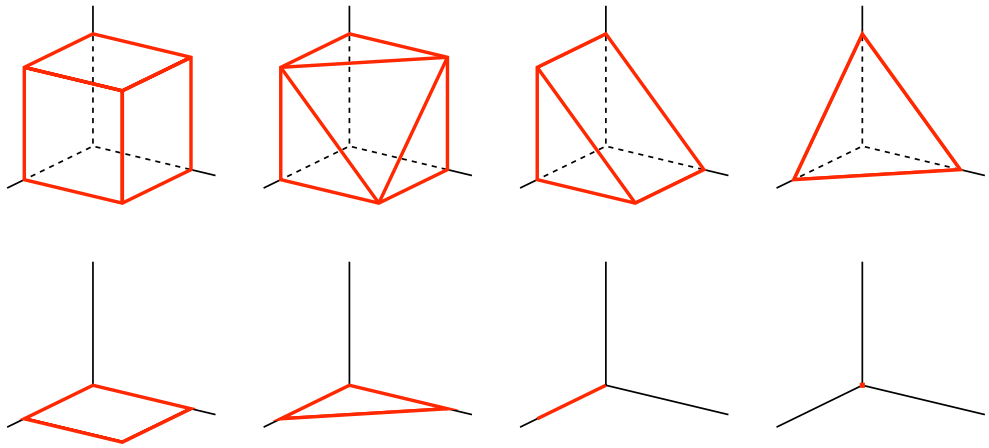

Figure 2: Extremal entropic polytopes for *m=3*

Other contributions

(i) How probability and physics can explain hard combinatorial optimization problems.

In 1971, Stephen Cook found that certain Boolean decision problems, where the goal is to determine if there exists a satisfying assignment to a logical function, are very hard to solve. Cook showed this on a specific decision problem called k-SAT (k-satisfiability) and defined a class of problems called the NP-complete class. The time required to solve any NP-complete problem with currently known algorithms may take millions of years using any amount of computer power available today. On the other hand, these problems arise very frequently in many applications, such as in computer science, computational biology or coding theory. Even worst, one is often interested in knowing not just if a solution exists but in knowing how many solutions exist; this is already a hard problem for 2-SAT, harder than any NP-complete problems. Note also that these problems have a priori no relationship to probability or physics.

It is important to point out that an NP-complete problem may not be hard for any given instance. Over the last twenty years, a considerable effort has been devoted to understanding the typical properties of k-SAT instances, drawing the models at random. There is indeed a fascinating phenomenon arising in this setting: just like water freezes when it crosses the zero temperature threshold, random k-SAT jumps from being solvable to unsolvable (with high probability) abruptly when the ratio between the number of constraints and variables crosses a critical threshold. This is called the phase transition phenomenon. It

has then been conjectured that in the phase where random k-SAT is solvable, the number of solutions can actually be predicted accurately around a deterministic number. This is interesting since we have just argued that in the worst case, this problem could be extremely hard to solve. The conjecture claims that for typical instances, this number can be accurately predicted. Partial progress was made on this conjecture in the 90's, but it remained open.

In (5), this conjecture is proved for random 2-SAT. The technique used in the proof borrows ideas from random graph theory and from statistical physics. Indeed, the analogy between water freezing and random k-SAT becoming unsolvable turns out to be more than just illustrative. There is a fascinating correspondence between the analysis of disordered systems in physics and the computer science problems. In particular, we use in (5) an adaptation of the so-called interpolation method from statistical mechanics to the random k-SAT problem, bringing a new tool to better understand this problem.

(ii) Proving a conjecture for optimal power allocations in MISO systems

In 1995, Telatar introduced in his paper on multiple-input multiple-output channels a conjecture which states (in its simplest form) the following. We are given *n* watts to power *n* devices (e.g., antennas) whose capacities are modeled by independent exponential random variables. In order to minimize the probability that the overall powered system goes in outage (i.e., the overall system's capacity is below a desired threshold), what is the best way to allocate the power? The conjecture states that for certain outage thresholds, full diversification is not the best solution.

In (4) this conjecture is proved. Because of the symmetry in the problem, due to the fact that each device has the same capacity, it is tempting to expect that splitting the power uniformly over all devices is optimal. Yet, the proved conjecture shows that this is a misleading intuition. Even in such a symmetric setting, the system can be more reliable by switching off certain devices and splitting the power uniformly over only a subset of the devices.

Acknowledgements

I am grateful to the Fondation Latsis Internationale and the research commission at EPFL for awarding me the 2011 Fondation Latsis Internationale at EPFL. I would like to thank also Professor Emre Telatar for his support and guidance during my postdoctoral research at EPFL.

References

(1) E. Abbe, *Extracting randomness and dependencies via a matrix polarization*, Information theory and applications workshop (ITA), San Diego, February 2011.

(2) E. Abbe, *Polar coding in networks: known results and new directions*, The Eighth International Symposium on Wireless Communication Systems, Aachen, November 2011.

(3) E. Abbe and E. Telatar, *Polar codes for the m-user MAC*, IEEE Transactions on Information Theory, 2011. Available at arXiv:1002.0777v.

(4) E. Abbe and S.-L. Huang and E. Telatar, *Proof of the outage probability conjecture for MISO channels*, IEEE Transactions on Information Theory, 2011. Available on arXiv:1103.5478v1.

(5) E. Abbe and A. Montanari, *On the concentration of the number of solutions of random satisfiability formulas*, International Congress of Mathematics Satellite Conference on Probability and Stochastic Processes, Bangalore, August 2010. Available on arXiv:1006.3786v1.

(6) R. Ahlswede, *Multi-way communication channels*, Proceedings of 2nd International Symposium on Information Theory, Thakadsor, Armenian SSR, pp. 23-52, Akademiai Kiado, Budapest, September, 1971.

(7) E. Arikan, *Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,* IEEE Trans. Inform. Theory, vol. IT-55, pp. 3051-3073, July 2009.

(8) E. Arikan and E. Telatar, *On the rate of channel polarization,* in Proc. 2009 IEEE Int. Symp. Inform. Theory, Seoul, pp. 1493-1495, 2009.

(9) A. El Gamal and T.M. Cover, *Multiple User Information Theory,* Multiple Access Communications Foundations for Emerging Technologies, edited by Norman Abramson, IEEE Information Theory Society, pp. 26-43, 1992, IEEE Press.

(10) C.E. Shannon, *A mathematical theory of communication,* Bell Syst. Tech. J., vol. 27, pp. 379-423, 623-656, 1948.