Attaining the capacity with Reed-Solomon codes through the $(U|U+V)$ construction and Koetter-Vardy soft decoding

Irene Márquez-Corbella, J.-P. Tillich
Inria de Paris, projet SECRET

July 5, 2016

"Les codes mènent à tout"

AT THE FOREFRONT OF COMMUNICATIONS RESEARCH

Claude Gueguen *Director of Research*

- Born 1941, Rennes (France)
- Graduated from Ecole Nationale Supérieure des Télécommunications 1965.
  PhD in engineering, University of Toulouse 1970.
- Research engineer, Ecole Nationale Supérieure d'Aéronautique 1966-70.
- At Telecom Paris since 1970 :
  Chairman of Systems and Communications Department,
  then Director of Research and Deputy Director of the Institute.
- IEEE Fellow.
- Personal areas of research :
  automatic control, systems theory, signal and speech processing.

THE SCIENTIFIC CONTEXT

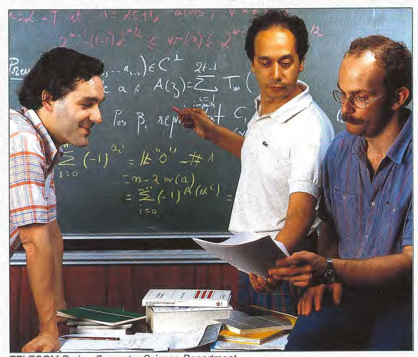
■ TELECOM Paris occupies a strategic position in today's major areas of scientific inquiry. The impact of the mathematician Claude E. Shannon's work is gradually causing conventional concepts of force and energy to give way to those of information, code and message in the interpretation of complex systems. These concepts are shaping the new currents of scientific thought and have found applications in areas as diverse as communications, linguistics, biology and economics. The concept of «information network» has acquired a central structural role in the corporation and in society as large, now rightly referred to at the "information society".

■ We are subjected to an ever-increasing flow of information and data, whose production, consumption and transmission pose a challenge to the engineer. This trend is accelerating under the impulse of technological advances in areas such as electronics, optics and space, which are expanding the capacity of communication channels and memory storage media. But the abundance of raw, unprocessed data is nothing but "noise". To give users effective access to relevant information - that is, to produce knowledge from data - transmission is no longer enough. Processing and understanding are also required. Networks must inevitably become "intelligent".

■ This change is having a deep impact on our lifestyles and is leading to a more specialized division of labor, higher qualifications for professionals, and a "delocalization" of resources. Man is present at all points of the network, and transmission channels must be adapted to human requirements. Today, technical achievement means building the transparent network - a network that offers new services while preserving all the features of communication. We therefore need intelligent, open and user-friendly networks.

■ The research programmes at TELECOM Paris are directed towards these goals.

Sophia Antipolis



TELECOM Paris - Computer Science Department

"Les codes mènent à tout"

Here ?

You meet codes everywhere even where you do not expect it

Reading this ?

GAFA, Geom. funct. anal.
Vol. 7 (1997) 438 – 461
1016-443X/97/030438-24 $ 1.50+0.20/0

GAFA Geometric And Functional Analysis

INFLUENCES OF VARIABLES AND THRESHOLD INTERVALS UNDER GROUP SYMMETRIES

J. Bourgain and G. Kalai

0 Introduction

A subset A of $\{0,1\}^n$ is called monotone provided if $x \in A, x' \in \{0,1\}^n$, $x_i \leq x'_i$ for $i = 1, \dots, n$ then $x' \in A$. For $0 \leq p \leq 1$, define μ_p the product measure on $\{0,1\}^n$ with weights $1-p$ at 0 and p at 1. Thus

$$(0.1) \qquad \mu_p\big(\{x\}\big) = (1-p)^{n-j}p^j \quad \text{where} \quad j = \#\{i = 1, \dots, n \mid x_i = 1\}\,.$$

If A is monotone, then $\mu_p(A)$ is clearly an increasing function of p. Considering A as a "property", one observes in many cases a threshold phenomenon, in the sense that $\mu_p(A)$ jumps from near 0 to near 1 in a short interval when $n \to \infty$. Well known examples of these phase transitions appear for instance in the theory of random graphs. A general understanding of such threshold effects has been pursued by various authors (see for instance Margulis [M] and Russo [R]). It turns out that this phenomenon occurs as soon as A depends little on each individual coordinate (Russo's zero-one law). A precise statement was given by Talagrand [T] in the form of the following inequality.

Define for $i = 1, \dots, n$

$$(0.2) \qquad A_i = \big\{x \in \{0,1\}^n \mid x \in A,\ U_i x \notin A\big\}$$

where $U_i(x)$ is obtained by replacement of the i^{th} coordinate x_i by $1 - x_i$ and leaving the other coordinates unchanged. Let

$$(0.3) \qquad \gamma = \sup_{i=1,\dots,n} \mu_p(A_i)\,.$$

Then

$$(0.4) \qquad \frac{d\mu_p(A)}{dp} \geq c\frac{\log(1/\gamma)}{p(1-p)\log[2/p(1-p)]}\,\mu_p(A)\ [1 - \mu_p(A)]$$

where $c > 0$ is some constant.

Defining for $i = 1, \dots, n$ the functions

$$(0.5) \qquad \varepsilon_i(x) = 2x_i - 1$$

one gets

What I should have read

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 41, NO. 2, MARCH 1995 469

The Threshold Probability of a Code

Gilles Zémor, *Member, IEEE*, and Gérard D. Cohen, *Senior Member, IEEE*

***Abstract*—We define and estimate the threshold probability θ of a linear code, using a theorem of Margulis originally conceived for the study of the probability of disconnecting a graph. We then apply this concept to the study of the erasure and Z-channels, for which we propose linear coding schemes that admit simple decoding. We show that θ is particularly relevant to the erasure channel since linear codes achieve a vanishing error probability as long as $p \leq \theta$, where p is the probability of erasure. In effect, θ can be thought of as a *capacity* notion designed for codes rather than for channels. Binomial codes have highest possible θ (and achieve capacity). As for the Z-channel, a subcapacity is derived with respect to the linear coding scheme. For a transition probability in the range $]\log(3/2); 1[$, we show how to achieve this subcapacity. As a by-product we obtain improved constructions and existential results for intersecting codes (linear Sperner families) which are used in our coding schemes.**

***Index Terms*— Threshold probability, erasure, Z-channel, intersecting code, binomial code.**

I. Introduction

SHANNON theory often tells us that residual error probabilities after decoding behave in a "threshold" manner, jumping suddenly from almost zero to almost one as a function of the channel error probability. However, those are results that describe the *average* behavior of large sets of codes. In this paper we investigate such a threshold phenomenon for *every* binary linear code.

More precisely, let us consider a binary linear code C, of parameters $[n, k, d]$, and let us choose randomly a vector v of length n such that every coordinate is given independently the value "1" with probability p and the value "0" with probability $1 - p$, $0 \leq p \leq 1$. Call $f_C(p)$ the probability with which v "covers" some nonzero codeword of C (i.e., is such that the support $\mathrm{supp}(v)$ of v contains the support of some codeword c). In other words

$$f_C(p) = \sum_{v \in W(C)} p^{|v|}(1-p)^{n-|v|}$$

where $|v|$ denotes the weight of v and

$$W(C) = \{v \mid \mathrm{supp}(v) \supset \mathrm{supp}(c),\ c \in C,\ c \neq 0\}.$$

The behavior we focus on is that whenever C has a large enough minimal distance, the (nondecreasing) function $p \mapsto f_C(p)$ jumps suddenly from almost zero to almost one, around a "threshold" probability θ. We will show how this fact stems from a theorem of Margulis, originally designed to prove a threshold phenomenon for the probability $f(p)$ of disconnecting a graph, when every edge is severed with probability p.

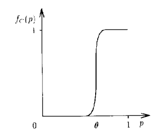


Fig. 1. A threshold phenomenon.

Threshold phenomena have been extensively studied in the context of random graphs (see, e.g., [3]). We have tried to apply those techniques to the coding context, and draw some consequences.

We will first place ourselves in the context of the erasure channel, and show that the threshold probability is a particularly relevant parameter for measuring the efficiency of a linear code. Indeed, $f_C(p)$ is exactly the probability of meeting with a decoding ambiguity, so that θ really measures the largest channel error probability the code can sustain. In effect, θ can be thought of as a *capacity* notion designed for codes rather than for channels.

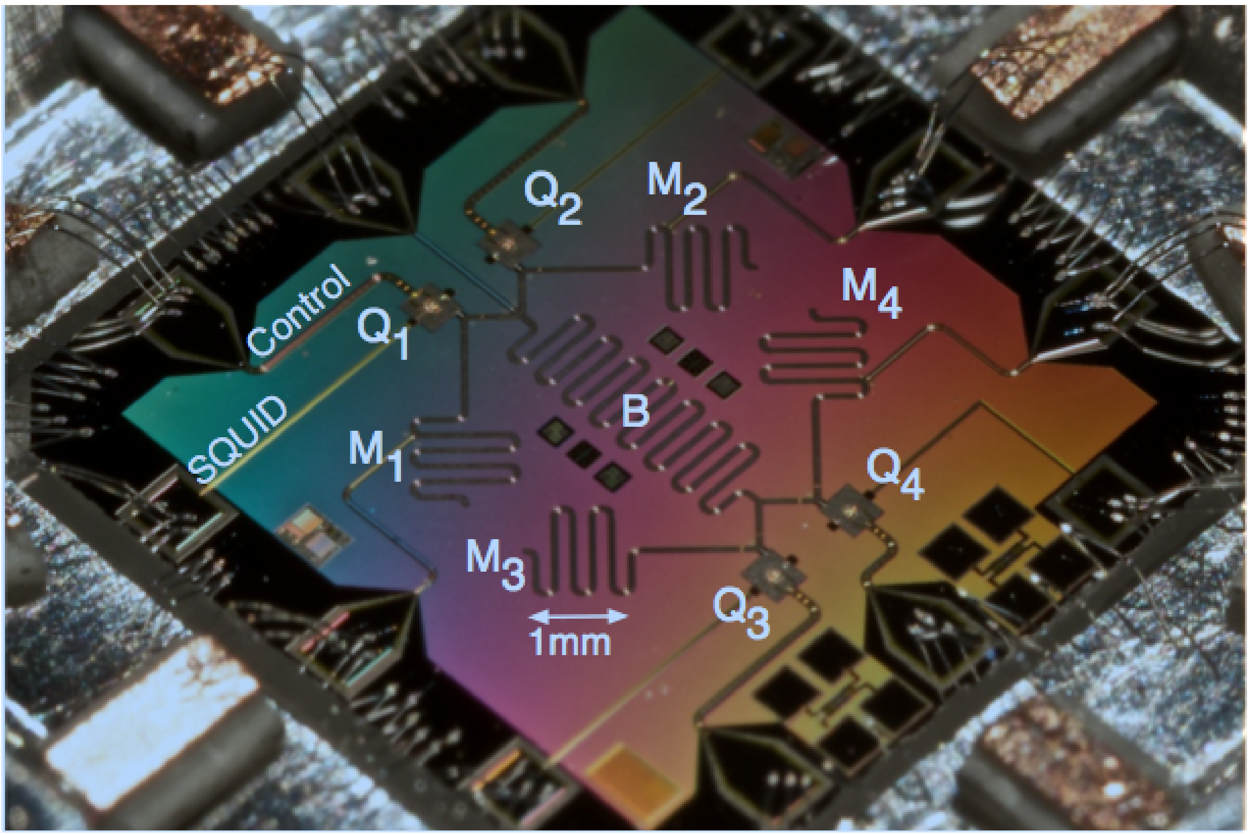
We will also discuss at some length an application of the threshold phenomenon to the problem of devising efficient codes for the asymmetrical channel (the so-called Z-channel) where every 0 can be transformed into a 1 with a given probability p, while 1's are always correctly received. In this setting, decoding of a received vector is unambiguous whenever the latter covers no codeword apart from the one that was initially sent. The idea, broadly speaking, is to use linear codes with properly chosen threshold properties: the point is, the probability that the received vector covers some parasite codeword should be very small whenever the proportion of $0 \to 1$ faulty transitions stays under a threshold value.

We will show why highly *intersecting* codes are a good choice. A linear code is said to be s-intersecting if the supports of any two nonzero codewords intersect on at least s coordinate positions. By highly intersecting we mean codes for which s is "large." We shall provide some constructions, and discuss the behavior of intersecting codes relative to the capacity of the Z-channel. It will turn out that for high error probabilities (e.g., $0.586 \leq p \leq 1$) our schemes perform quite acceptably. In passing, we improve known results on intersecting codes.

Outline of the Paper and Results: The paper is organized as follows. Section II describes a result of Margulis and its

Manuscript received September 15, 1992; revised July 20, 1994.
The authors are with Ecole Nationale Supérieure des Télécommunications, 75 634 Paris Cedex 13, France.
IEEE Log Number 9408650.

Doing this ?



Introduction : Codes nothing but codes

- Polar codes (Arıkan 2009): attain the symmetric capacity of any memoryless channel.
- Probability of error after decoding $O\left(2^{-N^{1/2-\epsilon}}\right)$.
- Improving this probability at a reasonable algorithmic cost.
- Changing the construction a little bit and using algebraic codes with a soft decoder: Reed-Solomon codes with the Koetter-Vardy decoder.

Polar codes "à la Dumer"

Definition 1. [$(U|U+V)$ code construction] *Let U be an $[n, k_u, d_u]_q$ code and V be an $[n, k_v, d_v]_q$ code. We define the $(U \mid U+V)$-construction of U and V as the linear code:*

$$\mathcal{C} = \{(\mathbf{u} \mid \mathbf{u}+\mathbf{v}) \mid \mathbf{u} \in U \textit{ and } \mathbf{v} \in V\}.$$

The code $\mathcal{C}$ has parameters $[2n, k_u + k_v, \min\{2d_u, d_v\}]_q$. A generator matrix of $\mathcal{C}$ is:

$$\left(\begin{array}{c|c} G_u & G_u \\ \hline \mathbf{0} & G_v \end{array}\right) \in \mathbb{F}_q^{(k_u+k_v)\times 2n}$$

where G_u and G_v are generator matrices of U and V respectively.

(Soft) Decoding $(U|U+V)$ codes

- We send $(\boldsymbol{u}|\boldsymbol{u}+\boldsymbol{v})$, we receive $(\boldsymbol{y}_1, \boldsymbol{y}_2)$
- Step 1, decoding the V-code : decode $\boldsymbol{y}_2 - \boldsymbol{y}_1 \rightarrow v$, probabilistic model :

$$\text{Prob}(v(i) = \alpha | y_1(i), y_2(i)) = \sum_{\beta \in \mathbb{F}_q} \text{Prob}(u(i) = \beta | y_1(i)) \text{Prob}(u(i) + v(i) = \alpha + \beta | y_2(i) \tag{1}$$

- Step 2: $\boldsymbol{y}_2 \rightarrow \boldsymbol{y}_2 - \boldsymbol{v}$
- Step 3, decoding the U-code : probabilistic model for decoding the U-code (two noisy versions of $\boldsymbol{u}$: $\boldsymbol{y}_1$ and $y'_2 \stackrel{\text{def}}{=} \boldsymbol{y}_2 - \boldsymbol{v}$)

$$\text{Prob}(u(i) = \alpha \mid y_1(i), y'_2(i)) = \frac{\text{Prob}(u(i) = \alpha \mid y_1(i)) \text{Prob}(u(i) = \alpha \mid y'_2(i))}{\sum_{\beta \in \mathbb{F}_q} \text{Prob}(u(i) = \beta \mid y_1(i)) \text{Prob}(u(i) = \beta \mid y'_2(i))} \tag{2}$$

Ingredient 1 : Arıkan's conservation law

- Model : symbol is transmitted correctly with probability $1-p$ and erased with probability p. Channel capacity $1-p$
- Noise model for the V-decoder: erasure channel of probability $2p-p^2$
- Noise model for the U-decoder: erasure channel of probability p^2

Nothing is lost in terms of capacity with this strategy

$$1-p=\frac{(1-2p+p^2)+1-p^2}{2}$$

- For other channels this also holds (Arıkan conservation law of mutual information)

Channel polarization

W memoryless channel with input alphabet $\mathbb{F}_q$ and output alphabet $\mathcal{Y}$

- W^0 channel viewed by the V-decoder;
- W^1 channel viewed by the U-decoder.

$$W^{a_0 a_1 \dots a_n} \stackrel{\text{def}}{=} \left(W^{a_0 a_1 \dots a_{n-1}}\right)^{a_n}$$

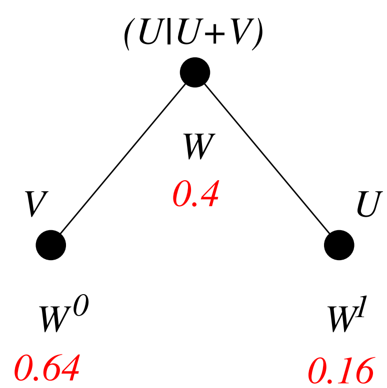
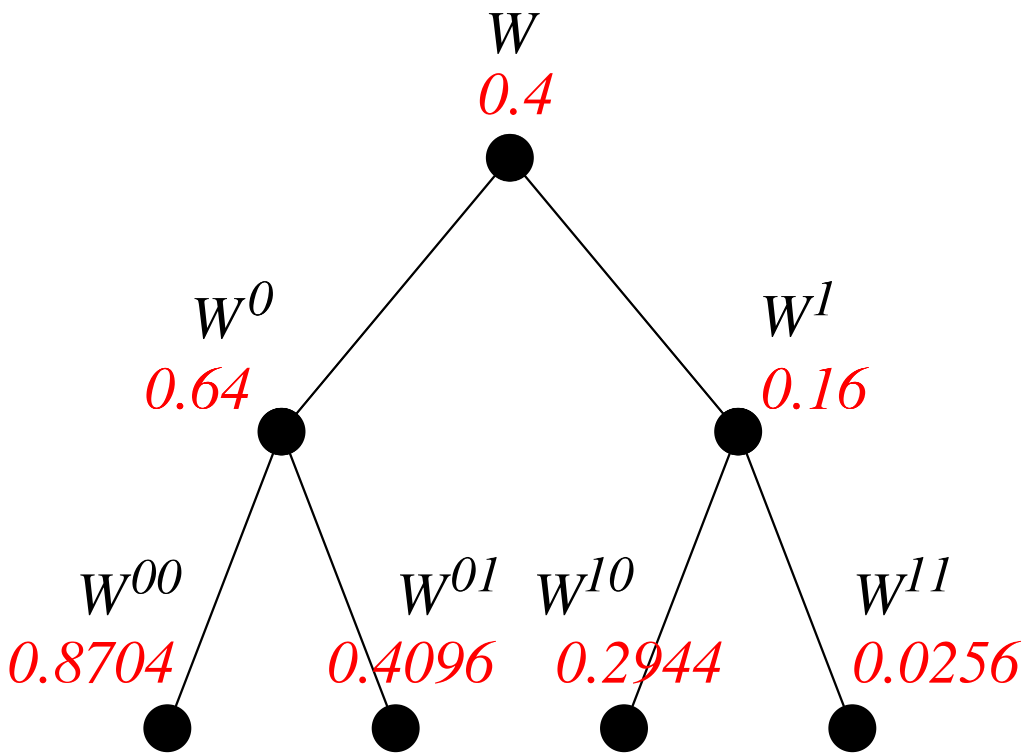
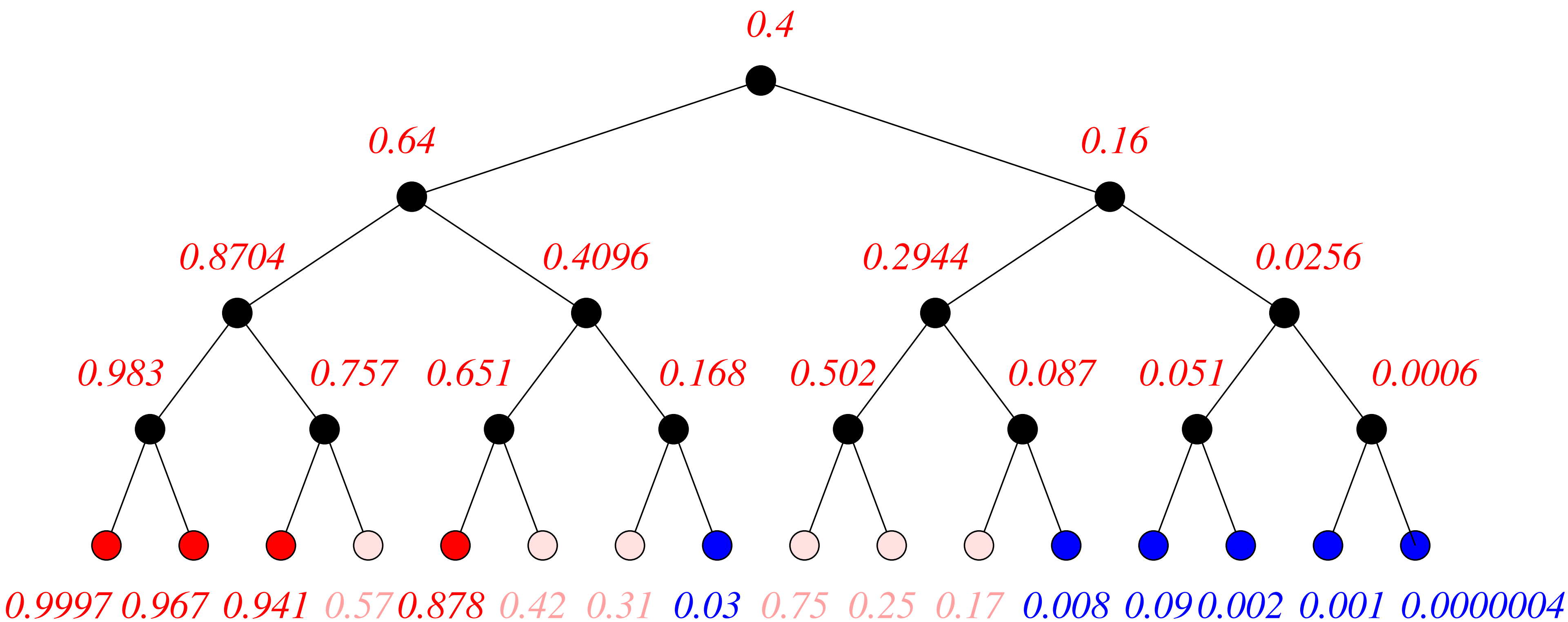


Figure 1: Example of an erasure channel with $p = 0.4$

Example



Example (II)



0.4
0.64
0.16
0.8704
0.4096
0.2944
0.0256
0.983
0.757
0.651
0.168
0.502
0.087
0.051
0.0006
0.9997 0.967 0.941 0.57 0.878 0.42 0.31 0.03 0.75 0.25 0.17 0.008 0.09 0.002 0.001 0.0000004

Polar codes

- Standard polar codes $=$ recursive $(u|u+v)$ construction where all the leaves are codes of length 1 (=symbols) and of rate 1 for the good channels and 0 for the bad channels.

Polarization

Bhattacharyya parameter $\mathcal{Z}(W)$

$$\mathcal{Z}(W) \stackrel{\text{def}}{=} \frac{1}{q(q-1)} \sum_{x,x' \in \mathbb{F}_q, x' \neq x} \sum_{y \in \mathcal{Y}} \sqrt{W(y|x)W(y|x')}$$

Theorem 1. [Şaşoğlu-Telatar-Arıkan] *For a symmetric channel of capacity C with q-ary inputs (q prime) and for all $0 < \beta < \frac{1}{2}$*

$$\lim_{\ell \to \infty} \frac{1}{N} \left| \{ i \in \{0,1\}^\ell : \mathcal{Z}\left(W^i\right) \leqslant 2^{-N^\beta} \} \right| = C,$$

where $N \stackrel{\text{def}}{=} 2^\ell$

$\Rightarrow$ probability of error of a standard polar code $2^{-\mathcal{O}\left(N^{1/2-\epsilon}\right)}$ where $N =$ length of the polar code. Follows from

$$P_e \leqslant (q-1)\mathcal{Z}(W)$$

Changing a little bit the structure

- Polar codes = recursive $(\boldsymbol{u}|\boldsymbol{u}+\boldsymbol{v})$ construction where all leaves are symbols.
- Our codes = recursive $(\boldsymbol{u}|\boldsymbol{u}+\boldsymbol{v})$ construction where all leaves are codes that admit an efficient soft decoder.
- Our choice : Reed-Solomon codes with the Koetter Vardy decoder.

Reed-Solomon codes

Definition 2. [Reed-Solomon code] *Let* $x_1, \ldots, x_n$ *be* n *distinct elements in* $\mathbb{F}_q$*. The Reed-Solomon code* $\mathcal{C}$ *associated to* $x_1, \ldots, x_n$ *of dimension* k *is the* $[n, k, d = n - k + 1]_q$ *code defined by*

$$\mathcal{C} = \{(P(x_i)_{1 \leqslant i \leqslant n} : \deg P \leqslant k, P \in \mathbb{F}_q[X]\}$$

Reed-Solomon codes

IG Codes

Irving and Gustave codes

The Koetter-Vardy decoder

- Soft (list) decoder of a Reed-Solomon code based on the reliability matrix Π associated to the received word $\boldsymbol{y} = (y_1, \dots, y_n)$ after $\boldsymbol{x} = (x_1, \dots, x_n)$ has been sent:

$$\Pi \stackrel{\text{def}}{=} \left(\text{Prob}(x_j = \alpha | y_j)\right)_{\substack{\alpha \in \mathbb{F}_q \\ 1 \leqslant j \leqslant n}}$$

- decoding algorithm that outputs a list that contains the codeword $\mathbf{c} \in C$ if

$$\frac{\langle \Pi, \lfloor \mathbf{c} \rfloor \rangle}{\sqrt{\langle \Pi, \Pi \rangle}} \geqslant \sqrt{k-1} + o(1)$$

where $\lfloor \mathbf{c} \rfloor$ represents a $q \times n$ matrix with entries $c_{i,\alpha} = 1$ if $c_i = \alpha$, and 0 otherwise; and $\langle A, B \rangle$ denotes the inner product of the two $q \times n$ matrices A and B, i.e.

$$\langle A, B \rangle \stackrel{\text{def}}{=} \sum_{i=1}^{q} \sum_{j=1}^{n} a_{i,j} b_{i,j}.$$

Symmetric channels

$x =$ symbol sent through the channel
$y =$ the received symbol

$$\pi = (\text{Prob}(x = \alpha | y))_{\alpha \in \mathbb{F}_q}$$

Definition 3. [discrete symmetric channel with q-ary inputs] *A DMC with q-ary inputs is said to be symmetric if and only if for any α in $\mathbb{F}_q$ we have*

$$p(\alpha)\mathit{Prob}(\pi = \boldsymbol{p}) = p(0)\mathit{Prob}(\pi = \boldsymbol{p}^{+\alpha}). \quad (3)$$

where $\boldsymbol{p}^{+\alpha} = (p(\beta + \alpha))_{\beta \in \mathbb{F}_q}$.

Analysis of the Koetter-Vardy decoder over symmetric channels

We clearly have

$$\mathbb{E}\left(\langle \Pi, \Pi \rangle\right) = n\mathbb{E}\left\|\pi\right\|_2^2$$

Lemma 1. *Over a symmetric channel*

$$\mathbb{E}\left(\langle \Pi, \lfloor \mathbf{0} \rfloor \rangle\right) = n\mathbb{E}\left\|\pi\right\|_2^2$$

$$\frac{\langle \Pi, \lfloor \mathbf{c} \rfloor \rangle}{\sqrt{\langle \Pi, \Pi \rangle}} \approx \sqrt{n\mathbb{E}\left\|\pi\right\|_2^2} \geqslant \sqrt{k-1} + o(1)$$

Analysis of the Koetter-Vardy decoder

Capacity of the Koetter-Vardy decoder for a certain symmetric channel

$$C_{\mathsf{KV}} = \mathbb{E}\,\|\pi\|_2^2$$

For instance consider the q-ary symmetric channel of crossover probability p

$$C_{\mathsf{KV}} = (1-p)^2 + (q-1)\frac{p^2}{(q-1)^2} = (1-p)^2 + \frac{p^2}{q-1} = (1-p)^2 + \mathcal{O}\big(\frac{1}{q}\big)$$

Analysis of the Koetter-Vardy decoder

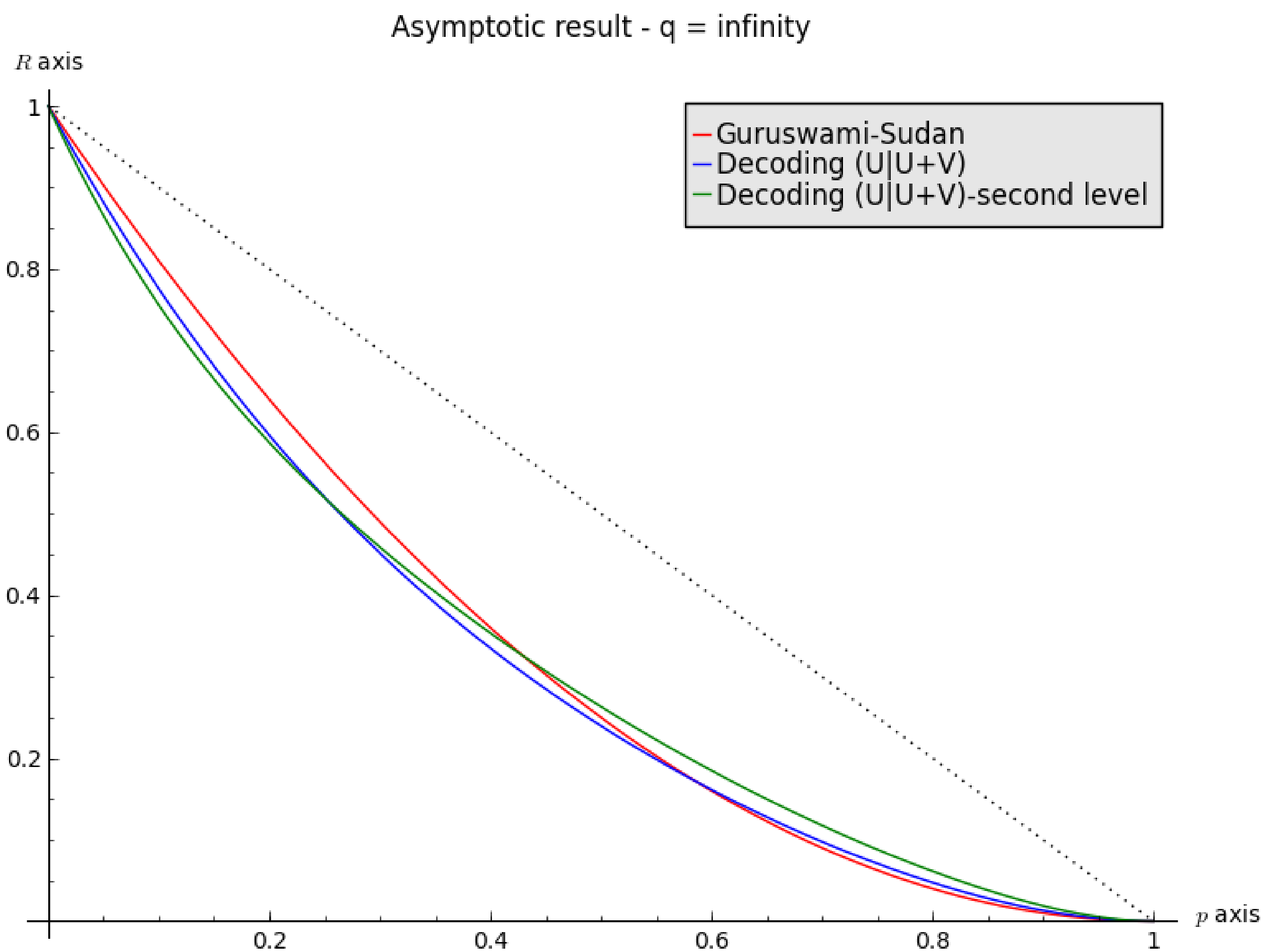
Theorem 2. *Let* $(\mathcal{C}_n)_{n\geqslant 1}$ *be an infinite family of Reed-Solomon codes of rate* $\leqslant R$. *Denote by* q_n *the alphabet size of* $\mathcal{C}_n$ *that is assumed to be a non decreasing sequence that goes to infinity with* n. *Consider an infinite family of* q_n*-ary symmetric channels with associated probability error vectors* π_n *such that* $\mathbb{E}\left(\|\pi_n\|_2^2\right)$ *has a limit as* n *tends to infinity. Let*

$$C_{KV} \stackrel{def}{=} \lim_{n\to\infty} \mathbb{E}\left(\|\pi_n\|_2^2\right).$$

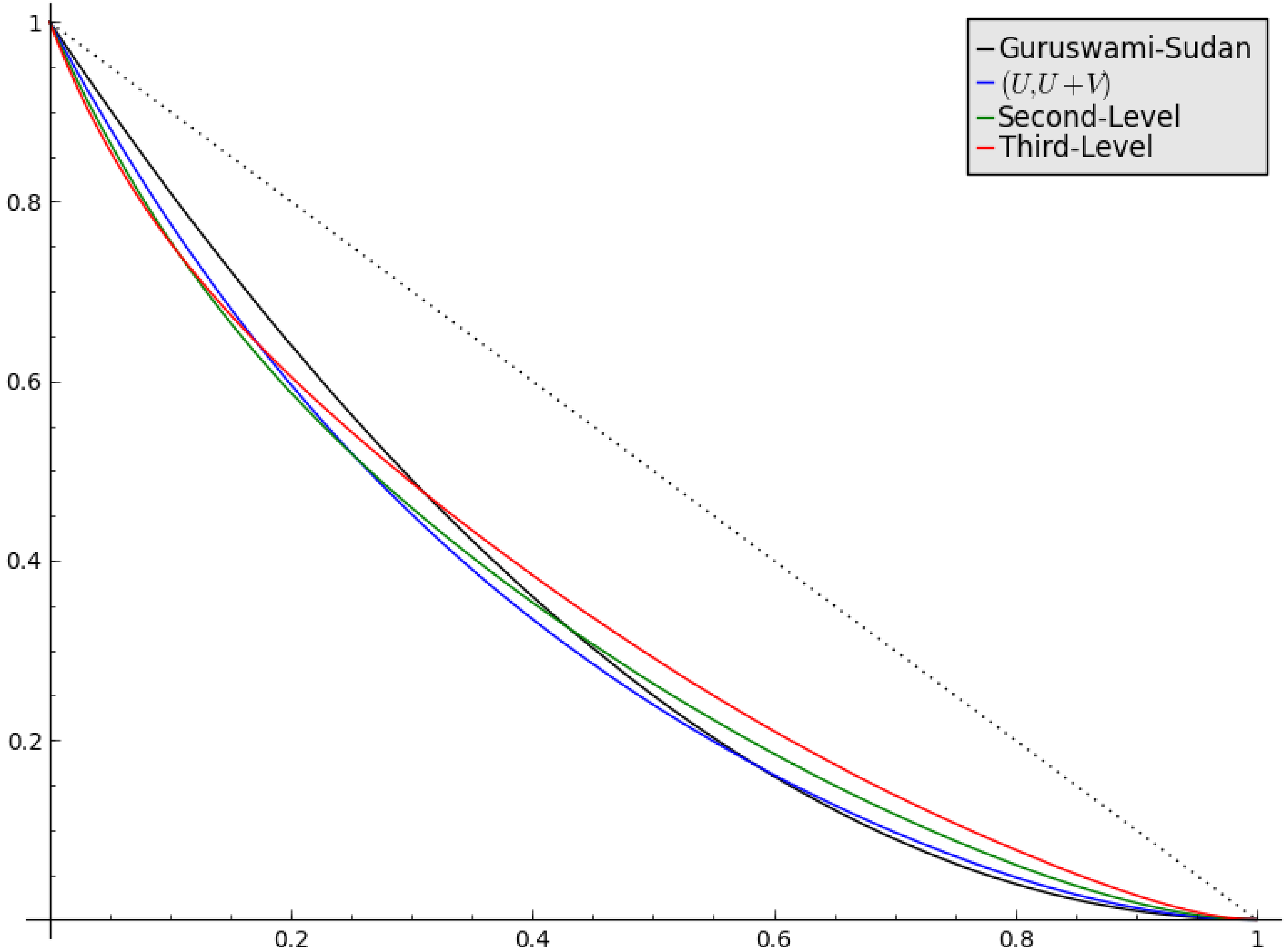
This infinite family of codes can be decoded correctly by the Koetter-Vardy decoding algorithm with probability $1-o(1)$ *as* n *tends to infinity as soon as there exists* $\epsilon > 0$ *such that*

$$R \leqslant C_{KV} - \epsilon.$$

Results up to 2 levels



Results up to 3 levels



Finite length analysis

Theorem 3. *If we decode a Reed-Solomon code of length n and rate $R < \mathbb{E}\left(\|\pi\|_2^2\right)$ over a symmetric channel with the Koetter-Vardy decoder, the probability that it outputs in its list the right codeword is upper-bounded by*

$$\mathcal{O}\left(e^{-K\delta^2 n}\right)$$

for some constant K and where $\delta = \mathbb{E}\left(\|\pi\|_2^2\right) - R$.

Finite length analysis (II)

Proposition 1. *For a symmetric channel*

$$1 - C_{KV} \leqslant (q-1)\mathcal{Z}(W)$$

Follows rather directly from the well known fact that the Rényi entropy

$$H_\alpha(X) \stackrel{\text{def}}{=} \frac{1}{1-\alpha}\log\sum_x p(x)^\alpha$$

is decreasing in α.

Polarization with RS leaves

- ℓ levels of polarization, leaves that are RS codes of maximal length q, $n \overset{\text{def}}{=} 2^\ell$
- Assume that q is prime.

For a symmetric channel of capacity C and for all $0 < \beta < \frac{1}{2}$

$$\begin{aligned} \lim_{\ell \to \infty} \frac{1}{n} \left| \left\{ i \in \{0,1\}^\ell : \mathcal{Z}\left(W^i\right) \leqslant 2^{-n^\beta} \right\} \right| &= C \\ &\Rightarrow \\ \lim_{\ell \to \infty} \frac{1}{n} \left| \left\{ i \in \{0,1\}^\ell : \mathbb{E}\left(\left\| \pi^i \right\|_2^2 \right) \geqslant 1 - (q-1) 2^{-n^\beta} \right\} \right| &= C \end{aligned}$$

- Take RS codes of rate $1 - \epsilon$ for those leaves.
- Take RS codes of rate 0 for the other leaves.

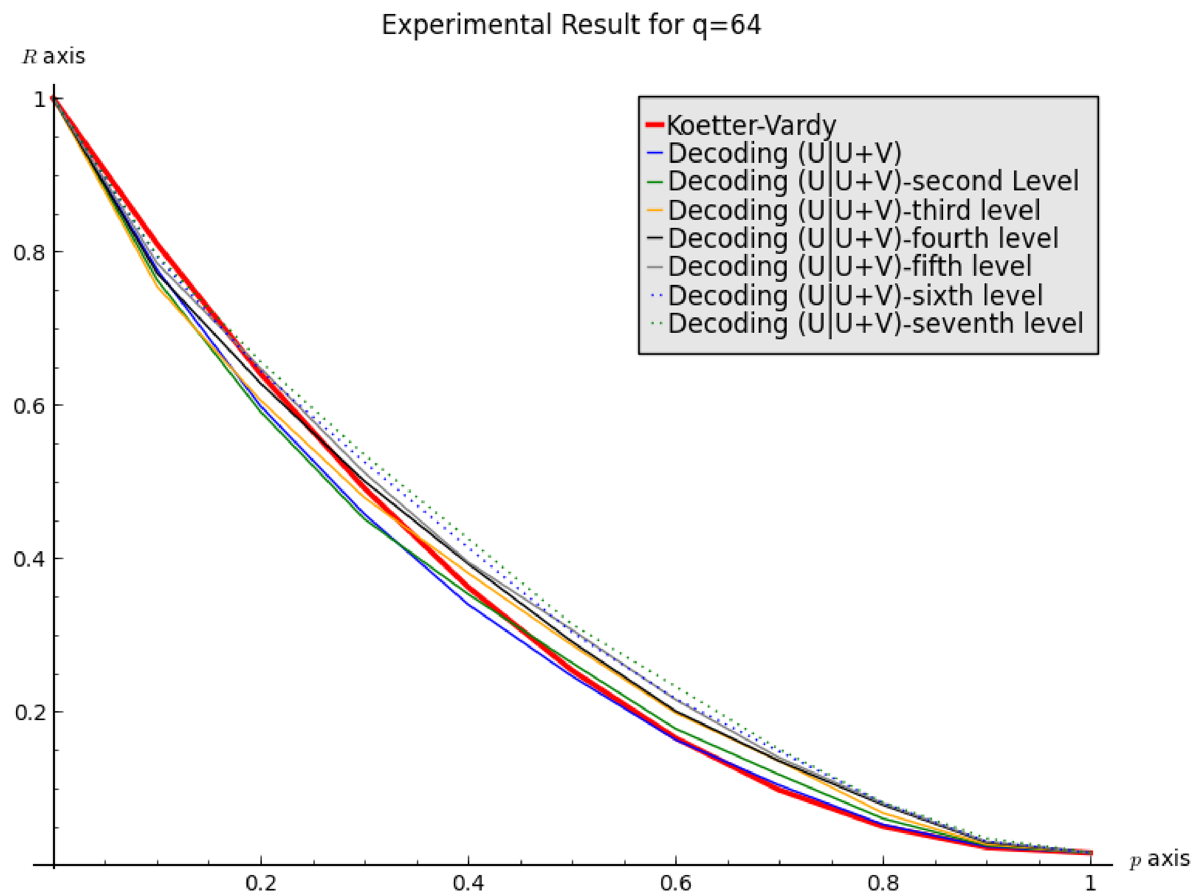
Polarization with RS leaves (II)

- Non zero leaves are decoded wrongly with probability $p = e^{-K\epsilon^2 q}$ when $(q-1)2^{-n^\beta} \leqslant \frac{1}{2}\epsilon$ say.
- Probability of failure for those leaves much better that if we had decoded symbol leaves at level $\ell + \log_q$ (probability of order $2^{-\sqrt{qn}}$).
- Overall rate of the code $\approx C(1-\epsilon)$

Finite length performance

- Get rather close to the channel capacity even with only 4 levels for $q = 64, 128, 256$ over the q-ary symmetric channel.
- Simulations : work in progress.

Finite alphabet Koetter-Vardy capacities



Going further: algebraic geometry codes

- Problem with RS codes : length $\leqslant q$.
- Algebraic geometric codes : more or less the same behaviour as RS codes but with an unbounded length and a fixed alphabet size.
- Allows to replace in the previous strategy q by an arbitrary length N : $p = e^{-K\epsilon^2 N}$ when $(q-1)2^{-n^{\beta}} \leqslant \frac{1}{2}\epsilon$.

Other strategies related to changing the kernel of polarization

- $(U|U+V) \rightarrow (U|U+V|U+V+W)$
- Improves the behaviour at the origin for the decoder.

Complexity of the Koetter Vardy decoder

- Polynomial complexity, but it amounts to solve a linear system with $\leqslant qnm^2$ where m is the number unknowns...
- More precisely of order $\sum_{\alpha \in \mathbb{F}_q, j \in \{1,\dots,n\}} m_{\alpha,j}^2$ where $m_{\alpha j} \approx$ proportional to $\Pi_{\alpha,j}$.
- Clearly better to perform this task over an iterated $(\boldsymbol{u}|\boldsymbol{u}+\boldsymbol{v})$ construction based on RS codes than on a RS code of the same length.
- Polarization process helps a lot to keep low multiplicities for the high rate parts of the iterated $(\boldsymbol{u}|\boldsymbol{u}+\boldsymbol{v})$ construction.

What kind of code is needed ?

The main ingredient: a family of codes of rate $R = 1 - \epsilon$ with an efficient soft decoder for any memoryless channel such that the probability that the decoder fails is

$$\mathcal{O}\left(e^{-K\delta^2 n}\right)$$

for some constant K and where $\delta =$ capacity of the channel $- R$.

Perspectives and conclusion

- Finite length behaviour over various channels by using only a few levels of the iterated $(\boldsymbol{u}|\boldsymbol{u}+\boldsymbol{v})$ construction.
- Studying various multiplicity choices for the Koetter-Vardy decoder.
- This strategy is of course not restricted to prime lengths.
- Gives in a natural way an exponential decay of the probability of error after decoding with a fixed number of levels.
- Scaling of the error probability in terms of gap ϵ to capacity ?
- Non negligible gain of $(U|U+V|U+V+W)$ over $(U|U+V)$?
- Study more precisely the error probabilities for algebraic geometry codes
- This strategy can be followed by using other decoders and/or other codes
- Applications to rate distorstion codes also for instance.