

The capacity of Abelian group codes over symmetric channels

Giacomo Como, Fabio Fagnani *

Abstract

In this paper the capacity achievable by Abelian group codes when employed over symmetric channels is determined. For certain important examples, like the AWGN channel with m -PSK modulation, it follows that this capacity coincides with the corresponding Shannon capacity of these channels. In other words using Abelian group codes in this case there is no loss of capacity (as it happens for binary linear codes for binary symmetric channels). Finally, a three dimensional modulation is presented for which instead, despite its group symmetry, the use of Abelian group codes leads to a loss in capacity.

Keywords: non-binary modulation, geometrically uniform modulation, m -PSK, group codes, Shannon capacity, error exponent, channel coding theorem.

1 Introduction

It is a well known fact that binary linear codes suffice to reach capacity on binary input symmetric channels [19, 36]. Moreover, by averaging over the ensemble of linear codes, we achieve the same mean error exponent as by averaging over the ensemble of all codes. In this paper we investigate the same question for group codes employed over non-binary channels exhibiting certain symmetries. The main example we have in mind is the AWGN channel with input set a geometrically uniform constellation (m -PSK for instance) [15] and with possibly hard or soft decoding rule. In [25] it was conjectured that group codes should suffice in this case to achieve capacity exactly as in the binary case and, up to our knowledge, there has not been any progress towards this direction. On the other hand, interest in group codes has not decreased in these years: indeed they give the possibility to use more spectral efficient modulations while keeping many good qualities of the binary linear codes like the uniform error property and nice structure for the corresponding minimal encoders and minimal trellis representations. See [32, 23, 35, 16, 2, 3, 26, 5, 27, 12, 24, 13, 14, 17] and references therein for an overview of the many research lines on group codes which have been developing during last years.

Recently, group codes have made their appearance also in the context of turbo concatenated schemes [21, 9, 10, 11] and of low density schemes [4, 8, 34]. In the binary case an important issue, for these type of high performance coding schemes, is the evaluation of the gap to Shannon capacity and also the rate of convergence to zero of the word and bit error rate. For regular low density schemes such gaps have been evaluated quite precisely [18, 28, 29] and it is shown that when the density parameters increase such schemes tend to attain the performance of generic binary linear codes which, as already said, are known

*Dipartimento di Matematica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10126 TORINO, Italy
giacomo.como@polito.it and fabio.fagnani@polito.it

to achieve Shannon capacity and the correct mean error exponent. In [4] the authors try to extend such an analysis to codes over the cyclic group \mathbb{Z}_q . We believe however that, without first a complete understanding of our original question, namely if group codes do themselves allow to reach capacity and the correct exponent, this type of analysis is inherently incomplete, since it can not be proved whether the gap to capacity is due to the fact that we are using a low density scheme or rather simply a group code. In a subsequent paper, we will propose a fundamental analysis of low density group codes, which will be based on the general results for group codes we will present in this paper.

Our work focuses on the case when groups are Abelian and consists of two parts. In the first part we introduce the concept of G -symmetric channel where G is an Abelian group and we determine (in a computational effective way) the capacity achievable using group codes over this channel: this capacity is called the G -capacity. The result is contained in Theorem 5 which is a sort of inverse Shannon theorem and in Theorem 12 which exhibits an average result working in the ensemble of group encoders. Also the mean error exponent is determined.

In the second part we prove that for an important class of examples including the AWGN channel with m -PSK modulation (and m the power of a prime), the \mathbb{Z}_m -symmetric capacity and the classical Shannon capacity do coincide so that Abelian group codes allow to achieve capacity in this case. Finally, we present a three dimensional AWGN example where instead the two capacities differ from each other. It remains an open problem if using possibly non-Abelian generating groups we can always achieve the Shannon capacity.

In Section 2 we briefly resume Shannon theory of memoryless channels and basic concepts concerning geometrically uniform constellations and we formally state the main question if group codes can achieve capacity of a symmetric channel.

In Section 3 we prove an inverse coding theorem for Abelian group codes, defining the G -capacity of a symmetric channel and showing that no reliable transmission is possible with G -codes at rates beyond this threshold value. The theorem is proved first for cyclic group codes, and the result is then extended to arbitrary Abelian groups.

Section 4 contains the main result consisting in a channel coding theorem for Abelian group codes over symmetric channels, stating that reliable transmission is possible at any rate below the G -capacity. As usual in information theory, the result is obtained by using a probabilistic method: we introduce an ensemble of random group encoders and prove that its average word error probability goes to 0 as the blocklength is increased. More precisely we show that the average error probability goes to 0 exponentially fast in the blocklength and that the exponential rate of convergence is at least equal to a certain function $E_G(R)$ which we call the G -random coding exponent. Although we have no complete tightness result for $E_G(R)$ we show that even when there is no loss of capacity there is a loss in the error exponent at low rates. We also state a similar result holding for a different ensemble of group codes using the kernel representation instead of the encoder image one.

Section 5 is devoted to the proof that for the AWGN channel with m -PSK constellation as input (and m the power of a prime) \mathbb{Z}_m -capacity and Shannon one do coincide, implying thus that \mathbb{Z}_m -codes employed over this channel achieve capacity.

Finally, in Section 6 we provide an explicit counterexample consisting in a three-dimensional geometrically uniform constellation admitting \mathbb{Z}_m as generating group: the AWGN channel with input restricted over this constellation the \mathbb{Z}_m -capacity is strictly less than Shannon capacity, implying thus that there is an algebraic obstruction to the use of \mathbb{Z}_m -codes in this case. It seems to be possible, but remains a completely open question, whether using non-Abelian group codes it allows to achieve capacity on this channel.

Some of the material of this paper has been presented at the ISIT 2005 [6].

2 Shannon theory for memoryless symmetric channels

In this section we introduce all relevant notation and definition and we formally state the problem.

2.1 Notation

Throughout the paper $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ will denote the ring of integers, $\mathbb{Z}^+ = \{0, 1, \dots\}$ the set of nonnegative integers, $\mathbb{N} = \{1, 2, \dots\}$ the set of natural numbers. For two naturals n and m , $\gcd(n, m) \in \mathbb{N}$ will denote the greater common divisor of n and m . For any m in \mathbb{N} , we shall denote by \mathbb{Z}_m the ring of integers modulo m , i.e. $\mathbb{Z}_m = \mathbb{Z}/(m\mathbb{Z})$, while, for any prime p and natural r , we shall denote by \mathbb{F}_{p^r} the Galois field with p^r elements, which as a group is isomorphic to \mathbb{Z}_p^r . For two groups G and H we will write $G \simeq H$ to mean they are isomorphic, while for a subset A of G $A \leq G$ will mean that A is a subgroup of G . As usual \mathbb{R} will be the real field, $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$ the sets respectively of nonnegative and of positive reals. \mathbb{C} will be the complex field which as a group is isomorphic to the linear space \mathbb{R}^2 through the bijection $(x, y) \mapsto z = x + iy$ where i is a root of $x^2 + 1 = 0$. The functions $\exp : \mathbb{R} \rightarrow \mathbb{R}_+$ and $\log : \mathbb{R}_+ \rightarrow \mathbb{R}$ have to be considered with respect to the same, arbitrary chosen, base $a \in \mathbb{R}_+$, unless explicit mention to the contrary.

2.2 Shannon theory for memoryless channels

We first review some notation and recall some results from Shannon classical theory of memoryless channels.

Let $\Omega = (A, \mathcal{A}, \mu)$ be a σ -finite measure space (see [30]). As usual $L^1(\Omega)$ will denote the space of (equivalence classes of) absolutely integrable functions $f : A \rightarrow \mathbb{R}$, and $\mathcal{P}(\Omega) \subseteq L^1(\Omega)$ the subset of probability densities, namely functions $f \in L^1(\Omega)$ such that $f(x) \geq 0$ for every $x \in \mathbb{R}$ and such that $\int_A f(x) d\mu(x) = 1$.

In the applications we have in mind there will basically be two possible situations. One case is when A is finite, \mathcal{A} consists of all the subsets of A and μ is the counting measure on A . In this case $L^1(\Omega) = \mathbb{R}^A$, the space of all the possible functions from A to \mathbb{R} and

$$\int_A f(x) d\mu(x) = \sum_{x \in A} f(x).$$

$\mathcal{P}(\Omega)$ thus consists of the usual probability distributions over the finite set A , namely functions $f : A \rightarrow \mathbb{R}^+$ such that $\sum_{x \in A} f(x) = 1$. With slight abuse of notation we will also write in this case, $\mathcal{P}(A)$ for $\mathcal{P}(\Omega)$.

The other case we will consider is when $A = \mathbb{R}^n$, \mathcal{A} is the Borel σ -algebra and μ is the Lebesgue measure. In this case $\mathcal{P}(\Omega)$ consists of the usual probability densities on \mathbb{R}^n . The readers preferring concrete formalism may think of these two examples. We prefer to keep the abstract formalism in our derivations: in this way we will be able to cover discrete and continuous examples at once in a rigorous way.

Given $f \in \mathcal{P}(\Omega)$ we define the entropy of f as

$$H(f) = - \int_A f(x) \log f(x) d\mu(x).$$

Notice that the definition of entropy is thus dependent on the specific chosen measure space and in particular it is carried on with respect to the specific measure μ . In the finite case

it is the usual discrete entropy taking values in $[0, \log |A|]$, while in the continuous case it coincides with the so called differential entropy taking values in $[-\infty, +\infty]$.

A memoryless channel (MC) consists of

- a finite input set \mathcal{X} ,
- an output set consisting of a σ -finite measure space $\mathcal{Y} = (Y, \mathcal{B}, \mu)$,
- a family of transition probability densities $W(\cdot|x) \in \mathcal{P}(\mathcal{Y})$ indexed by the elements $x \in \mathcal{X}$.

Such a channel will be denoted $(\mathcal{X}, \mathcal{Y}, W)$. We will say that two MCs $(\mathcal{X}, \mathcal{Y}, W)$ and $(\mathcal{X}', \mathcal{Y}', W')$ are equivalent if there exist a bijection $\varphi : \mathcal{X} \rightarrow \mathcal{X}'$ and a measurable map $\theta : \mathcal{Y} \rightarrow \mathcal{Y}'$ admitting measurable inverse, such that $L \in \mathbb{R}_+$ exists such that $\mu'(\theta(A)) = L\mu(A)$ for all $A \in \mathcal{B}$ and

$$W'(\theta(y)|\varphi(x)) = \frac{1}{L}W(y|x) .$$

From a MC as above we can define the N -th extension having input set \mathcal{X}^N and output set $\mathcal{Y}^N = (Y^N, \mathcal{B}^N, \mu^N)$ where \mathcal{B}^N is the product σ -algebra and μ^N is the product measure. The corresponding transition probability densities are given by $W_N(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^N W(y_j|x_j)$ and this motivates the name memoryless, the various transmissions being probabilistically independent once the input signals have been fixed.

A block encoder for the MC $(\mathcal{X}, \mathcal{Y}, W)$ consists of a finite set \mathcal{U} and of a map $\phi : \mathcal{U} \rightarrow \mathcal{X}^N$. N is said to be the encoder length and $R = \log |\mathcal{U}|/N$ its rate. A decoder is any measurable mapping $\mathcal{D} : \mathcal{Y}^N \rightarrow \mathcal{U}$. A coding scheme consists of a pair of an encoder and a decoder. Once a coding scheme has been fixed we can speak of word error probability as follows. Assume \mathbf{U} is a r.v. uniformly distributed on \mathcal{U} and let $\mathbf{X} = \phi \circ \mathbf{U}$. Let moreover \mathbf{Y} be the r.v. on \mathcal{Y}^N whose probabilistic description is given by the conditional density $W_N(\mathbf{y}|\mathbf{x})$ and whose marginal density is thus given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} W_N(\mathbf{y}|\phi(u))$$

(in doing this we are automatically enforcing independence between \mathbf{U} and the channel). Finally, put $\hat{\mathbf{U}} = \mathcal{D} \circ \mathbf{Y}$. The error event is defined as $e = \{\hat{\mathbf{U}} \neq \mathbf{U}\}$ and the probability of error as the probability of such event

$$P(e) = P(\hat{\mathbf{U}} \neq \mathbf{U}) .$$

It is useful also to consider the probability of error conditioned to the transmission of the information word u , which we denote by $P(e|u)$. Clearly,

$$P(e) := \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} P(e|u) .$$

For a finite set A and $f : A \rightarrow \mathbb{R}$ we will denote by $\operatorname{argmax}_{a \in A} f(a)$ a random variable taking values in A with uniform distribution over the subset $A_f = \{a' \in A : \max_{a \in A} f(a) = f(a')\}$, i.e. in the case of non uniqueness, we pick one of the maxima at random with uniform probability. It is well known that, given an encoder, the decoding scheme minimizing the error probability is the so called maximum likelihood (ML) decoding

$$\mathcal{D}_{\text{ML}}(\mathbf{y}) = \operatorname{argmax}_{\mathbf{u} \in \mathcal{U}} P(\mathbf{U} = \mathbf{u} | \mathbf{Y} = \mathbf{y}) = \operatorname{argmax}_{\mathbf{u} \in \mathcal{U}} W_N(\mathbf{y}|\mathbf{u}) .$$

Actually the ML-decoder defined above is not a deterministic measurable map from \mathcal{Y}^N to \mathcal{X}^N . Nevertheless it is possible to generalize the definition of decoding scheme considering measurable functions \mathcal{D} from \mathcal{Y}^N to the set $\mathcal{P}(\mathcal{X}^N)$ of probability measures over \mathcal{X}^N : \mathcal{D}_{ML} can be shown to minimize the error probability over this larger set of decoders.

From now on we will always assume that ML decoding is used. We will also use the notation $P(e|\phi)$ and $P(e|\phi, u)$ whenever we want to emphasize the dependence on the particular chosen encoder ϕ .

We recall a few simple consequences of ML decoding that will be used in the paper. We assume we have fixed an MC $(\mathcal{X}, \mathcal{Y}, W)$, an encoder $\phi : \mathcal{U} \rightarrow \mathcal{X}^N$ and an element $u \in \mathcal{U}$.

(1) Let $\Psi : \mathcal{U}' \rightarrow \mathcal{U}$ be a bijection; then,

$$P(e|\phi, u) = P(e|\phi \circ \Psi, \Psi^{-1}(u)) \quad (1)$$

(2) Consider a partition $\mathcal{U} \setminus \{u\} = \mathcal{U}_1 \cup \dots \cup \mathcal{U}_r$ and define $\phi_i = \phi|_{\mathcal{U}_i \cup \{u\}}$. Then

$$\max_{1 \leq i \leq r} P(e|\phi_i, u) \leq P(e|\phi, u) \leq \sum_{i=1}^r P(e|\phi_i, u) \quad (2)$$

(3) If $|\phi^{-1}(\phi(u))| > 1$, then

$$P(e|\phi, u) \geq \frac{1}{2}. \quad (3)$$

(4) Let $(\mathcal{X}', \mathcal{Y}', W')$ an MC which is equivalent to $(\mathcal{X}, \mathcal{Y}, W)$ through the measurable bijections $\varphi : \mathcal{X} \rightarrow \mathcal{X}'$ and $\theta : \mathcal{Y} \rightarrow \mathcal{Y}'$. Let $\varphi_N : \mathcal{X}^N \rightarrow (\mathcal{X}')^N$ be the componentwise extension of φ and define the encoder $\phi' : \mathcal{U} \rightarrow \mathcal{X}'$ as $\phi'(u) = \varphi_N \circ \phi$. The error probability $P'(e)$ of ϕ' over the channel $(\mathcal{X}', \mathcal{Y}', W')$ satisfies

$$P'(e|\phi', u) = P(e|\phi, u). \quad (4)$$

It follows from (1) that, if ϕ is injective, $P(e|\phi)$ only depends on the encoder ϕ through its image, the code $\phi(\mathcal{U})$. In this paper we will prefer working with encoders instead of codes since they admit simpler parameterizations which are suitable for probabilistic averaging arguments. For the same reason we will be forced to consider also non-injective encoders.

A further step in Shannon construction consists in considering, for given $R \in [0, \log |\mathcal{X}|]$ and $N \in \mathbb{N}$, a r.v. Φ uniformly distributed over all possible maps from \mathcal{U} to \mathcal{X}^N , where $|\mathcal{U}| = \lceil \exp(RN) \rceil$. $\overline{P(e)}^R$ will denote the average error probability with respect to such probability distribution over the set of all possible encoders having rate equal to R .

In order to state the classical Shannon result we are only left with defining capacity and error exponents. The capacity of the MC $(\mathcal{X}, \mathcal{Y}, W)$ is defined as

$$C := \max_{p \in \mathcal{P}(\mathcal{X})} \left\{ \sum_{x \in \mathcal{X}} p(x) \left(\int_{\mathcal{Y}} W(y|x) \log \left(\frac{W(y|x)}{\sum_{z \in \mathcal{X}} p(z)W(y|z)} \right) d\mu(y) \right) \right\}. \quad (5)$$

Its random coding exponent is instead defined as follows. We put, for any $\rho \in [0, 1]$ and $p \in \mathcal{P}(\mathcal{X})$,

$$E_0(\rho, p) := -\log \left(\int_{\mathcal{Y}} \left(\sum_{x \in \mathcal{X}} p(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu(y) \right) \quad (6)$$

and we define

$$E(R) := \max_{0 \leq \rho \leq 1} \max_{p \in \mathcal{P}(\mathcal{X})} (E_0(\rho, p) - \rho R), \quad R \in [0, \log |\mathcal{X}|]. \quad (7)$$

A well known fact (see [19], [36]) is that

$$E(R) > 0 \Leftrightarrow R < C. \quad (8)$$

Moreover $E(R)$ is continuous, monotonically decreasing and convex in the interval $[0, C)$, while the dependence of both C and $E(R)$ from the transition probabilities of the channel is continuous (with respect to the $L^1(\mathcal{Y})$ norm). Also notice that, if $(\mathcal{X}, \mathcal{Y}, W)$ and $(\mathcal{X}', \mathcal{Y}', W')$ are equivalent MCs, then their capacities and error exponents do coincide.

We can now state Shannon classical result:

Theorem 1 *Assume we have fixed a MC $(\mathcal{X}, \mathcal{Y}, W)$ having capacity C and random coding exponent $E(R)$. It holds*

(a)

$$\overline{P(e)}^R \leq \exp(-NE(R)).$$

In particular this implies that the average error probability tends to 0 exponentially fast for $N \rightarrow +\infty$, provided that the rate of the encoders is kept below C .

(b) *For every $R > C$ there exists a constant $A_R > 0$ independent of N such that for any coding scheme having rate not smaller than R , we have that $P(e) \geq A_R$.*

Given an MC $(\mathcal{X}, \mathcal{Y}, W)$, we can consider subchannels obtained by simply restricting the inputs to a subset $\mathcal{X}_0 \subseteq \mathcal{X}$: they will be denoted by $(\mathcal{X}_0, \mathcal{Y}, W)$. In the sequel we will assume that every MC $(\mathcal{X}, \mathcal{Y}, W)$ considered satisfies the following non-degenerate assumption: for any $\mathcal{X}_0 \subseteq \mathcal{X}$ with $|\mathcal{X}_0| \geq 2$, the subchannel $(\mathcal{X}_0, \mathcal{Y}, W)$ has strictly positive capacity.

2.3 GU constellations and symmetric channels

In this paper we will focus on channels with symmetries. We start by stating a few concepts about group actions. Given a finite group G , with identity 1_G , and a set A we say that G acts on A if, for every $g \in G$, it is defined a bijection of A denoted by $a \mapsto ga$, such that

$$h(ga) = (hg)a \quad \forall h, g \in G, \forall a \in A.$$

In particular we have that the identity map corresponds to 1_G and the maps corresponding to an element g and its inverse g^{-1} are the inverse of each other. The action is said to be transitive if for every $a, b \in A$ there exists $g \in G$ such that $ga = b$. Finally, the action is said to be simply transitive if the element g above is always unique in G . If G acts simply transitively on a set A , it is necessarily in bijection with A , a possible bijection being given by $g \mapsto ga_0$ for any fixed $a_0 \in A$.

Given a σ -finite measure space $\mathcal{Y} = (Y, \mathcal{B}, \mu)$ we say that the group G acts isometrically on \mathcal{Y} if it is defined an action of G on Y consisting of measurable bijections such that

$$\mu(gA) = \mu(A) \quad \forall A \in \mathcal{B}, \forall g \in G. \quad (9)$$

Notice that in the case when Y is a finite set, (9) is trivially always verified so that in this case all actions are isometric. Instead in the case when $Y = \mathbb{R}^n$, (9) is a real restriction and is verified if the maps $y \mapsto gy$ are isometries of \mathbb{R}^n , i.e. maps preserving the Euclidean distance.

Definition 2 Let G be a group. A MC $(\mathcal{X}, \mathcal{Y}, W)$ is said to be G -symmetric if

- (a) G acts simply transitively on \mathcal{X} ,
- (b) G acts isometrically on \mathcal{Y} ,
- (c) $W(y|x) = W(gy|gx)$ for every $g \in G$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

An important property of G -symmetric channels is that, for both their Shannon capacity C and their random coding exponent $E(R)$, the maximizing probability distribution $p \in \mathcal{P}(\mathcal{X})$ in the variational definitions (5) and (7) respectively can be chosen to be the uniform distribution over the input set \mathcal{X} .

Given the n -dimensional Euclidean space \mathbb{R}^n , an n -dimensional *constellation* in this context is a finite subset $S \subset \mathbb{R}^n$ spanning \mathbb{R}^n ; i.e. every $\mathbf{x} \in \mathbb{R}^n$ can be written as $\mathbf{x} = \sum_{s \in S} \alpha_s s$ with $\alpha_s \in \mathbb{R}$. Given an n -dimensional constellation S and any isometry $\tau : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\tau(S)$ also is an n -dimensional constellation. For this reason we can restrict ourselves to the study of constellations $S \subset \mathbb{R}^n$ with barycenter $\mathbf{0}$, i.e. such that $\sum_{s \in S} s = \mathbf{0}$: they are the ones minimizing the average per symbol energy over the class of those constellations obtained one from the other by applying isometries.

We denote by $\Gamma(S)$ its symmetry group, namely the set of all isometric permutations of S with the group structure endowed by the composition operation. Clearly $\Gamma(S)$ acts on S . S is said to be *geometrically uniform (GU)* if this action is transitive; a subgroup $G \leq \Gamma(S)$ is a *generating group* for S if for every $s, r \in S$ a unique $g \in G$ exists such that $gs = r$, namely if G acts simply transitively on S . It is well known that not every GU constellation admits a generating group (see [33] for a counterexample). However in what follows we will always assume that the constellations we are dealing with, do admit generating groups, and, actually, Abelian ones.

Let S be an n -dimensional GU constellation equipped with a generating group G . Define the S -AWGN channel as the n -dimensional unquantized AWGN channel with input set S , output \mathbb{R}^n with the usual Lebesgue measure structure, and transition densities given by

$$W(y|x) = N(y - x),$$

where $N(x)$ is the density of an n -dimensional diagonal Gaussian r.v.:

$$N(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|x\|^2}{2\sigma^2}}.$$

Now let S' be another GU constellation such that $S \subseteq S'$ and $G \leq \Gamma(S')$. Let us introduce the quantization map over the Voronoi regions of S'

$$Q : \mathbb{R}^n \rightarrow S' \quad Q(x) = \operatorname{argmin}_{s \in S'} \|x - s\|,$$

as previously resolving non uniqueness cases by assigning to $Q(x)$ a value picked at random with uniform probability over the set of minima. We define the (S, S') -AWGN channel as the MC obtained by applying Q to the output of the S -AWGN channel. Note that the special case $S = S'$ coincides with the so called hard decoding rule.

Proposition 3 The S -AWGN channel and the (S, S') -AWGN channel are both G -symmetric.

Notice that the above construction of G -symmetric channels with a GU constellation as input can be extended to a much wider class of channels than the AWGN case. Indeed, let S an n -dimensional GU constellation admitting generating group G . Let $f \in \mathcal{P}(\mathbb{R}^n)$ be any probability density over \mathbb{R}^n depending only on the Euclidean norm of the argument, i.e. such that there exists $\tilde{f} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $f(x) = \tilde{f}(\|x\|)$. An S additive isotropic noise (S -AIN) channel is a memoryless channel (S, \mathbb{R}^n, W) such that a function $f \in \mathcal{P}(\mathbb{R}^n)$ as above exists with $W(y|x) = f(y - x)$ for all $y \in \mathbb{R}^n, x \in S$.

Example 1 The unquantized isotropic Laplacian channel with input constrained on S is a K_m -AIN channel. Here $\mathcal{Y} = \mathbb{R}^n$ with the Lebesgue measure μ , while transition laws are given by

$$W(y|x) = \frac{\lambda^n \Gamma(n/2)}{2\pi^{n/2} \Gamma(n)} e^{-\lambda \|x-y\|},$$

where $\lambda > 0$ is a fixed parameter and

$$\Gamma(t) := \int_0^{+\infty} x^{t+1} e^{-x} dx$$

is the well known Euler's Γ function. □

Now let S' be another GU constellation such that $S \subseteq S'$ and $G \leq \Gamma(S')$. We define an (S, S') -AIN channel as the MC obtained by applying a quantization over Voronoi regions of S' to the output of an S -AIN channel. It is easy to see that the following generalization of Proposition 3 holds true.

Proposition 4 Any S -AIN channel and any (S, S') -AIN channel are both G -symmetric.

A well known fact (see [32]) is that every GU constellation S lies on a sphere, clearly centered in the origin $\mathbf{0}$ since we restricted ourselves to the study of constellations with barycenter $\mathbf{0}$. If we denote by L the ray of such a sphere, its square L^2 will be equal to the per symbol energy of our transmission. Moreover, for every $a \in \mathbb{R}_+$ the homothety $\theta_a : \mathbb{R}^n \rightarrow \mathbb{R}^n, x \xrightarrow{\theta_a} ax$ is such that $\theta_a(S)$ is a GU constellation with the same isometry group and the same generating groups of S . Consider an S -AWGN channel (S, \mathbb{R}^n, W') with energy per symbol L^2 and noise variance σ^2 . Then the $\theta_{1/L}(S)$ -AWGN channel $(\theta_{1/L}(S), \mathbb{R}^n, W')$ with standard deviation $\sigma' = \frac{1}{L}\sigma$ is equivalent to the first one since $\mu(\theta_{1/L}(A)) = \frac{1}{L^N} \mu(A)$ and

$$W'(\theta_{1/L}(y)|\theta_{1/L}(x)) = \frac{1}{\sqrt{(2\pi(\sigma')^2)^n}} e^{-\frac{\|\frac{1}{L}y - \frac{1}{L}x\|^2}{\sigma'^2}} = L^N \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{\|y-x\|^2}{(\sigma)^2}} = L^N W(y|x).$$

This shows that S -AWGN channels with the same signal to noise ratio $L^2/(2\sigma^2)$ are equivalent. The signal to noise ratio and the dimension n of the constellation, which is a measure of the bandwidth required for each transmission, are the two engineering parameters of interest for S -AWGN channels.

In the following we present some examples of GU constellations admitting Abelian generating group. We present only the cases with unitary energy per symbol, as the other cases can be obtained by homotheties.

Example 2 The simplest, one-dimensional, GU constellation is the 2-PAM, defined by

$$K_2 := \{1, -1\}.$$

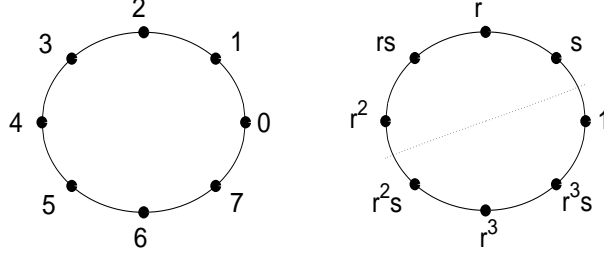


Figure 1: K_8 -constellation with the two labelings \mathbb{Z}_8 and D_4

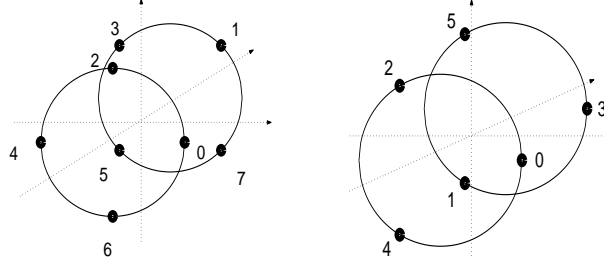


Figure 2: \mathbb{Z}_8 -labelled K_8^β and \mathbb{Z}_6 -labelled $K_{3 \times 2}^\beta$

It is trivial to see that $\Gamma(K_2) \simeq \mathbb{Z}_2$ is a generating group for K_2 . It is also possible to show that K_2 is the only one-dimensional GU constellation. \square

Example 3 For any integer $m \geq 2$, define $\xi_m := e^{i\frac{2\pi}{m}}$. Define the m -PSK constellation as

$$K_m := \left\{ \xi_m^k, k = 0, \dots, m-1 \right\} \subset \mathbb{C} \simeq \mathbb{R}^2 .$$

Clearly S is two-dimensional for $m \geq 3$. It can be shown that $\Gamma(K_m) \simeq D_m$, where D_m is the dihedral group with $2m$ elements. K_m admits \mathbb{Z}_m , i.e. the Abelian group of integers modulo m , as generating group. When m is even there is another generating group (see [15], [25]): the dihedral group $D_{m/2}$, which is noncommutative for $m \geq 6$. Now, let $m' = am$ be an arbitrary multiple of m and define the quantization map over Voronoi regions of the m' -PSK constellation. The m -PSK-AWGN channel and the $(m$ -PSK, m' -PSK)-AWGN channel are both \mathbb{Z}_m -symmetric and (for even m) $D_{m/2}$ -symmetric. Constellation K_8 with the two possible labelings \mathbb{Z}_8 and D_4 is reported in Fig.1 \square

Next example shows how higher dimensional GU constellations can be obtained as Cartesian product of lower dimensional ones.

Example 4 For any integer $m > 2$ consider the family of 3D GU constellations parametrized by $\beta \in (0, +\infty)$

$$K_{m \times 2}^\beta := \left\{ \left(\sqrt{\frac{1}{1+\beta^2}} \xi_m^k, \sqrt{\frac{\beta^2}{1+\beta^2}} (-1)^l \right), k = 0, 1, 2, l = 0, 1 \right\} \subset \mathbb{C} \times \mathbb{R} \simeq \mathbb{R}^3 .$$

Fig.2 shows the special case $m = 3$. It's easy to show that $\mathbb{Z}_m \times \mathbb{Z}_2$ is a generating group for $K_{m \times 2}^\beta$; notice that, for odd m , $\mathbb{Z}_m \times \mathbb{Z}_2 \simeq \mathbb{Z}_{2m}$. Thus, for odd m , unquantized and quantized AWGN channels with input m -PSK \times 2-PAM are \mathbb{Z}_{2m} -symmetric. \square

Finally we provide an example of an 'effectively' three-dimensional constellation.

Example 5 For even $m > 2$ we introduce the family of 3-dimensional GU constellations, parametrized by $\beta \in (0, +\infty)$

$$K_m^\beta = \left\{ \left(\sqrt{\frac{1}{1+\beta^2}} \xi_m^k, \sqrt{\frac{\beta^2}{1+\beta^2}} (-1)^k \right), k = 1, \dots, m \right\} \subset \mathbb{C} \times \mathbb{R} \simeq \mathbb{R}^3 .$$

An example with $m = 8$ is shown in Fig.2. It can be shown that, similarly to the constellations K_m , the constellations K_m^β have two different generating groups, \mathbb{Z}_m and $D_{m/2}$; so, in the standard way, we obtain channels that are both \mathbb{Z}_m -symmetric and $D_{m/2}$ -symmetric. \square

Since the input of a G -symmetric channel can be identified with the group G itself, block encoders for such channels are (eventually non injective) maps $\phi : \mathcal{U} \rightarrow G^N$, where N is the *block length* and \mathcal{U} a finite set. We will focus our attention on the class of G -encoders: namely we assume that \mathcal{U} is a group with identity $1_{\mathcal{U}}$ and ϕ a group homomorphism.

One reason for considering G -encoders is that for them the uniform error property (UEP) under ML decoding holds true, when they are employed on a G -symmetric channel. This means that the word error probability using the encoder ϕ conditioned to the transmission of the information word u , $P(e|\phi, u)$, does not depend on u . In particular we have that $P(e|\phi, u) = P(e|\phi, 1_{\mathcal{U}})$. This also implies that

$$P(e|\phi) = P(e|\phi, 1_{\mathcal{U}}) .$$

Restricting the class of encoders for a G -symmetric channel to that of G -encoders gives raise to the following fundamental question: *which is the capacity achievable by G -encoders over a G -symmetric channel?* In this paper we will give an answer for the special case when G is Abelian.

3 The converse to the channel coding theorem for Abelian G -encoders on G -symmetric channels

In this section we define a new concept of capacity for G -symmetric channels when G is an arbitrary Abelian group and then we exhibit a sort of inverse Shannon theorem: we prove that the probability of error for any G -encoder having rate above such capacity is bounded away from 0, independently of its blocklength.

Whenever dealing with Abelian groups, we will use the additive notation to denote group operation, while 0 will always denote the identity element. We will use the symbol \oplus to denote both external direct sum of groups, as well internal sum of subgroups when their intersection reduce to $\{0\}$. We will use the symbol $+$ and \sum instead to denote general summation of subgroups. Some facts about the theory of Abelian groups will be recalled when needed, while we refer to [22] for further details.

3.1 The cyclic case

We start our analysis with the special case when $G = \mathbb{Z}_{p^r}$ for some prime p and positive integer r . Note that \mathbb{Z}_{p^r} also has ring structure with the product induced by that of \mathbb{Z} .

Suppose you want to communicate over a \mathbb{Z}_{p^r} -symmetric MC $(\mathcal{X}, \mathcal{Y}, W)$, using \mathbb{Z}_{p^r} -encoders. Our aim is to find out the range of rates at which reliable communication is possible under these conditions.

From now on we will identify \mathcal{X} with \mathbb{Z}_{p^r} . For $l = 1, \dots, r$, consider the channel obtained by restricting the input set from \mathbb{Z}_{p^r} to its subgroup $p^{r-l}\mathbb{Z}_{p^r}$: call it the l -th subchannel and denote its capacity by C_l . The l -subchannel is easily seen to be $p^{r-l}\mathbb{Z}_{p^r}$ -symmetric, so that C_l can be obtained, in the variational definition (5), with uniform distribution over the input set $p^{r-l}\mathbb{Z}_{p^r}$. As we will see soon, subchannels will play a fundamental role in our analysis. Let \mathcal{U} be a finite Abelian group and $\phi : \mathcal{U} \rightarrow \mathbb{Z}_{p^r}^N$ a homomorphic encoder. It is not restrictive to assume that

$$\mathcal{U} = \mathbb{Z}_p^{k_1} \oplus \mathbb{Z}_{p^2}^{k_2} \oplus \dots \oplus \mathbb{Z}_{p^r}^{k_r} . \quad (10)$$

for suitable positive integers k_1, \dots, k_r . Indeed, in next subsection it will be shown that if \mathcal{U} has not such a structure, than ϕ is surely noninjective so that $P(e|\phi) \geq \frac{1}{2}$ by property (3) of ML decoding. As a consequence of (10), there exist homomorphisms $\phi^j : \mathbb{Z}_{p^j}^{k_j} \rightarrow \mathbb{Z}_{p^r}^N$ such that, if we consider $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_r)$ with $\mathbf{u}_j \in \mathbb{Z}_{p^j}^{k_j}$ for every j , we have that $\phi(\mathbf{u}) = \sum_{j=1}^r \phi^j(\mathbf{u}_j)$. ϕ 's rate is given by

$$R := \frac{\log |\mathcal{U}|}{N} = \frac{1}{N} \sum_{j=1}^r j k_j \log p .$$

For every $l = 1, \dots, r$, consider

$$\mathcal{U}_{(l)} = \mathbb{Z}_p^{k_1} \oplus \dots \oplus \mathbb{Z}_{p^l}^{k_l} \oplus p\mathbb{Z}_{p^{l+1}}^{k_{l+1}} \oplus \dots \oplus p^{(r-l)}\mathbb{Z}_{p^r}^{k_r} .$$

Note that

$$\phi(\mathcal{U}_{(l)}) \leq p^{r-l}\mathbb{Z}_{p^r}^N .$$

Define ϕ_l as the restriction of ϕ to $\mathcal{U}_{(l)}$ and denote by $R^{(l)}$ its rate.

The converse to the channel coding theorem (item (b) of Theorem 1) states that necessary condition for $P(e|\phi_l)$ to be made arbitrarily small is that

$$R^{(l)} \leq C_l . \quad (11)$$

Notice that,

$$\begin{aligned} R^{(l)} &= \frac{\log |\mathcal{U}_{(l)}|}{N} \\ &= \frac{\log p}{N} \left(\sum_{j=1}^l j k_j + l \sum_{j=l+1}^r k_j \right) \\ &\geq \frac{\log p}{N} \left(\frac{l}{r} \sum_{j=1}^l j k_j + l \sum_{j=l+1}^r \frac{j}{r} k_j \right) \\ &= \frac{\log p}{N} \frac{l}{r} \left(\sum_{j=1}^l k_j \right) \\ &= \frac{l}{r} R , \end{aligned} \quad (12)$$

with equality if and only if $k_j = 0$ for $j = 1, \dots, r-1$, i.e. if and only if $\mathcal{U} = \mathbb{Z}_{p^r}^K$ with $K = \frac{RN}{r \log p}$.

By the property (2) of ML decoding,

$$P(e|\phi_l) \leq P(e|\phi) , \quad l = 1, \dots, r . \quad (13)$$

From (11), (12) and (13) it follows that necessary condition for $P(e|\phi)$ to be made arbitrarily small is that

$$R \leq \min_{l=1, \dots, r} \frac{r}{l} C_l , \quad (14)$$

and that the only way to eventually achieve this bound is by using encoders whose domain is a free \mathbb{Z}_{p^r} module, i.e. $\phi : \mathbb{Z}_{p^r}^K \rightarrow \mathbb{Z}_{p^r}^N$.

In the rest of this chapter we will generalize these considerations to generic Abelian groups G . In Section 4 we will then prove the converse result which, for the particular cyclic case, will amount to say that at any rate below $\min_{l=1,\dots,r} \frac{r}{l} C_l$ we can reliably transmit using \mathbb{Z}_{p^r} -encoders.

3.2 Arbitrary Abelian group

In order to generalize our considerations to arbitrary Abelian groups, we need to set down some more notation and recall some basic facts about finite Abelian groups.

Let M be a finite Abelian group. Given $\mu \in \mathbb{N}$ define the following subgroups of M :

$$\mu M = \{\mu x \mid x \in M\}, \quad M_{(\mu)} = \{x \in M \mid \mu x = 0\}.$$

It is immediate to verify that $\mu M = \{0\}$ if and only if $M_{(\mu)} = M$. Let then

$$\mu_M = \min\{\mu \in \mathbb{N} \mid M_{(\mu)} = M\} = \min\{\mu \in \mathbb{N} \mid \mu M = \{0\}\}.$$

Notice that μ_M is well defined and $\mu_M \leq |M|$, since, as it is easy to see, $M_{(|M|)} = M$ or equivalently $|M|M = \{0\}$.

Decompose $\mu_M = p_1^{r_1} \cdots p_s^{r_s}$ where $p_1 < p_2 < \cdots < p_s$ are distinct primes and r_1, \dots, r_s are non-negative integers, existence and uniqueness of such a decomposition being guaranteed by the fundamental theorem of algebra. It is a standard fact that M admits the direct sum decomposition

$$M = M_{(p_1^{r_1})} \oplus \cdots \oplus M_{(p_s^{r_s})}. \quad (15)$$

Each $M_{(p_i^{r_i})}$ is a $\mathbb{Z}_{p_i^{r_i}}$ -module and, up to isomorphisms, can be further decomposed, in a unique way, as a direct sum of cyclic groups

$$M_{(p_i^{r_i})} = \mathbb{Z}_{p_i^{k_{i,1}}} \oplus \mathbb{Z}_{p_i^{k_{i,2}}} \oplus \cdots \oplus \mathbb{Z}_{p_i^{k_{i,r_i}}}. \quad (16)$$

The sequence $\sigma^M = (p_1, \dots, p_s)$ will be called the spectrum of M , the sequence $\mathbf{r}^M = (r_1^M, \dots, r_s^M)$ the multiplicity and, finally, the double indexed sequence

$$\mathbf{k}^M = (k_{i,j} \mid i = 1, \dots, s; j = 1, \dots, r_i^M),$$

will be called the type of M . It will be convenient often to use the following extension: $k_{i,j} = 0$ for $j > r_i^M$. Given a sequence of primes $\sigma = (p_1, \dots, p_s)$, we will say that M is σ -adapted if σ^M is a subsequence of σ . Notice that, once the sequence of primes σ has been fixed, all σ -adapted Abelian groups are completely determined by their type (which includes the multiplicities r_i^M with the agreement that some of them could be equal to 0). We will denote by $M_{\mathbf{k}}$ the finite Abelian group having type \mathbf{k} .

Notice that if M is a finite Abelian group with type \mathbf{k} and $N \in \mathbb{N}$, the Abelian group M^N has the same spectrum and multiplicity of M and type $N\mathbf{k}$.

If M and L are finite Abelian groups and $\phi \in \text{Hom}(M, L)$, then $\phi(M_{(\mu)}) \subseteq L_{(\mu)}$ and $\phi(\mu M) \subseteq \mu L$ for every $\mu \in \mathbb{N}$. It follows that ϕ is surely non-injective if M is not σ^L -adapted or if any of the multiplicities in M is strictly larger than the corresponding in L .

Suppose now we have fixed, once for all, a finite Abelian group G having spectrum $\sigma^G = (p_1, \dots, p_s)$, multiplicity $\mathbf{r}^G = (r_1^G, \dots, r_s^G)$ and type \mathbf{k}^G . We will consider G -encoders $\phi \in \text{Hom}(\mathcal{U}, G^N)$ with domain consisting of a finite Abelian group \mathcal{U} which is σ^G -adapted and is such that, $\mathbf{r}^{\mathcal{U}} \leq \mathbf{r}^G$ (in the sense that $r_i^{\mathcal{U}} \leq r_i^G$ for each i). In fact if \mathcal{U} does not fulfil these requirements then ϕ is surely noninjective for our previous considerations, and thus its ML

word error probability is bounded from below by the constant 1/2. The group \mathcal{U} admits a decomposition as illustrated above in (15) and (16). Let us fix now a matrix

$$\mathbf{l} = (l_{i,j} \in \mathbb{Z}^+ \mid i = 1, \dots, s, j = 1, \dots, r_i^G)$$

such that $l_{i,j} \leq j$ for every i and j . We will say that \mathbf{l} is an \mathbf{r}^G -compatible matrix. Define

$$\mathcal{U}(\mathbf{l}) = \bigoplus_{i=1}^s \mathcal{U}_{(p_i^{r_i^G})}(\mathbf{l}_i). \quad (17)$$

$$\mathcal{U}_{(p_i^{r_i^G})}(\mathbf{l}_i) = \bigoplus_{j=1}^{r_i^G} p_i^{j-l_{i,j}} \mathbb{Z}_{p_i^j}^{k_{i,j}}. \quad (18)$$

An immediate consequence of previous considerations is that

$$\phi(\mathcal{U}(\mathbf{l})) \subseteq \left(\bigoplus_{i=1}^s \sum_{j=1}^{r_i^G} p_i^{j-l_{i,j}} G_{(p_i^j)} \right)^N.$$

These inclusions automatically give information theoretic constraints to the possibility of reliable transmission using this type of encoders. Denote by $R_{\mathbf{l}}$ the rate of $\phi|_{\mathcal{U}(\mathbf{l})}$ and by $C_{\mathbf{l}}$ the capacity of the subchannel having as input alphabet the subgroup $G_{\mathbf{l}}$ of G defined by:

$$G_{\mathbf{l}} = \bigoplus_{i=1}^s \sum_{j=1}^{r_i^G} p_i^{j-l_{i,j}} G_{(p_i^j)}.$$

Then,

$$R_{\mathbf{l}} \leq C_{\mathbf{l}} \text{ for every } \mathbf{r}^G \text{ - compatible } \mathbf{l} \quad (19)$$

is a necessary condition for reliable transmission. This does not give explicit constraints yet to the rates R at which reliable transmission is possible using G -encoders. For this we need some extra work using the structure of the Abelian groups $\mathcal{U}(\mathbf{l})$. Notice that

$$R_{\mathbf{l}} = \frac{1}{N} \sum_{i=1}^s \sum_{j=1}^{r_i^G} l_{i,j} k_{i,j} \log p_i.$$

It is useful to introduce the following probability distribution on the pairs (i, j) :

$$\alpha_{i,j} = \frac{j k_{i,j} \log p_i}{\log |\mathcal{U}|}.$$

From the above definition, and recalling that $\log |\mathcal{U}| = RN$, we can represent

$$k_{i,j} = \frac{RN \alpha_{i,j}}{j \log p_i}.$$

Hence,

$$R_{\mathbf{l}} = R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}.$$

Consequently, (19) can be equivalently expressed as

$$R \leq \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} \frac{C_1}{\sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}}, \quad (20)$$

where $\mathbf{l} \neq \mathbf{0}$ means that $l_{i,j} \neq 0$ for some i, j .

Denote now by $\mathcal{P}(\mathbf{r}^G)$ the set of probability distributions $\alpha_{i,j}$ on the set of pairs (i, j) such that $i = 1, \dots, s$ and $j = 1, \dots, r_i^G$. We define the G -capacity of a G -symmetric channel as

$$C_G = \max_{\alpha \in \mathcal{P}(\mathbf{r}^G)} \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} \frac{C_1}{\sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}}. \quad (21)$$

Since $\mathcal{P}(\mathbf{r}^G)$ is compact and

$$f : \alpha \mapsto \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} \frac{C_1}{\sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}}$$

is a continuous map from $\mathcal{P}(\mathbf{r}^G)$ to \mathbb{R}^+ , definition (21) is well posed in the sense that f has a maximum point in $\mathcal{P}(\mathbf{r}^G)$. Such a maximum point could be not unique in principle: nevertheless we will call G -optimal splitting and denote by α_G any element of $\mathcal{P}(\mathbf{r}^G)$ such that

$$\min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} \frac{C_1}{\sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}^G} = C_G. \quad (22)$$

It clearly follows from our previous considerations that C_G is a un upper bound to reliable transmission using G -encoders. Precisely, we have the following result which is an immediate consequence of the inverse Shannon coding theorem (item (b) of Theorem 1).

Theorem 5 *Consider a G -symmetric channel and let C_G be its G -capacity. Then, for every $R > C_G$ there exists $A_R > 0$ depending on R but not on N , such that, for every G -encoder ϕ of rate R and length N , with any decoding rule, the corresponding word error probability satisfies*

$$P(e|\phi) \geq A_R.$$

In the next three examples we present some explicit computations of C_G for groups G with particular algebraic structure. First we examine the field case, showing as in this case G -capacity C_G do coincide with Shannon capacity C , as follows from classical linear coding theory.

Example 6 Suppose the group G admits Galois field structure. In this case we necessarily have $G \simeq \mathbb{Z}_p^k$ for some prime p and positive integer k . Thus

$$\sigma^G = (p), \quad \mathbf{r}^G = (1).$$

Consequently, the only \mathbf{r}^G -compatible \mathbf{l} is given by $\mathbf{l} = 1$ and therefore we have that in this case $C_G = C$.

However, GU constellations admitting a generating group which is isomorphic to a Galois field are affected by a constraint on their bandwidth efficiency. In fact, if S is an n -dimensional

GU constellation admitting \mathbb{Z}_p^k as generating group, then standard arguments using group representation theory allow to conclude that

$$n \geq \begin{cases} k, & \text{if } p = 2 ; \\ 2k, & \text{if } p \geq 2 . \end{cases} \quad (23)$$

□

In next example we would like to show that in the special case when $G = \mathbb{Z}_{p^r}$ condition (20) coincides with condition (14) obtained in the previous subsection.

Example 7 Let $G = \mathbb{Z}_{p^r}$. We want to show that

$$C_G = \min_{l=1, \dots, r} \frac{r}{l} C_l .$$

Notice first that in this case $\sigma^G = (p)$ and $\mathbf{r}^G = r$. A vector $\mathbf{l} = (l_1, \dots, l_r)$ is \mathbf{r}^G -compatible if and only if $l_j \leq j$ for every $j = 1, \dots, r$. Notice now that

$$G_{\mathbf{l}} = \sum_{j=1}^r p^{j-l_j} G_{(p^j)} = \sum_{j=1}^r p^{j-l_j} p^{r-j} \mathbb{Z}_{p^r} = \sum_{j=1}^r p^{r-l_j} \mathbb{Z}_{p^r} = p^{r-l^*} \mathbb{Z}_{p^r} ,$$

where

$$l^* = \max_{j=1}^r l_j .$$

Hence, $C_{\mathbf{l}} = C_{l^*}$.

Notice now that $\mathcal{P}(\mathbf{r}^G)$ simply consists of the probability distributions $\alpha = (\alpha_1, \dots, \alpha_r)$. Suppose we have fixed $\alpha \in \mathcal{P}(\mathbf{r}^G)$. We have that

$$\min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} \frac{C_{\mathbf{l}}}{\sum_{j=1}^r \frac{l_j}{j} \alpha_j} = \min_{\rho=1}^r C_{\rho} \frac{1}{\max_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.} \\ l^* = \rho}} \sum_{j=1}^r \frac{l_j}{j} \alpha_j} .$$

Now,

$$\max_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.} \\ l^* = \rho}} \sum_{j=1}^r \frac{l_j}{j} \alpha_j \geq \frac{\rho}{r}$$

and equality holds true if and only if $\alpha_r = 1$ and $\alpha_j = 0$ for every $j \neq r$.

Hence, in this case we have rediscovered what we had already found out in the previous subsection, i.e.

$$C_{\mathbb{Z}_{p^r}} = \min_{\rho=1}^r \frac{r}{\rho} C_{\rho} , \quad \alpha^{\mathbb{Z}_{p^r}} = (0, \dots, 0, 1) .$$

□

Example 8 Now consider the $K_{2 \times 3}^{\beta}$ constellation introduced in Example 4. Consider a $K_{2 \times 3}^{\beta}$ -AWGN channel. It is easy to show that the independence of orthogonal components of the Gaussian noise imply that the capacity $C_6(\beta)$ of such a channel is equal to the sum of the capacities of its two subchannels, $C_2(\beta)$ and $C_3(\beta)$. This fact allows us to explicitly write down the optimal splitting, i.e. the $\alpha \in \mathcal{P}(r^G)$ solution of the variational problem (21) defining $C_{\mathbb{Z}_6}$, as a function of the parameter β .

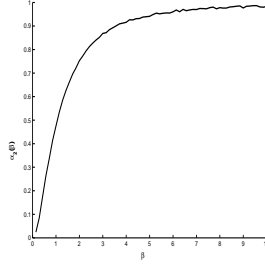


Figure 3: The optimal splitting for $K_{2 \times 3}^\beta$ as a function of β

Since $\mathbb{Z}_6 \simeq \mathbb{Z}_2 \times \mathbb{Z}_3$, we have that $s = 2$, $p_1 = 2$, $p_2 = 3$, and $\mathbf{r}^G = (r_1^G, r_2^G) = (1, 1)$. (21) reduces to

$$C_{\mathbb{Z}_6}(\beta) = \max_{\alpha \in \mathcal{P}(\{2,3\})} \min \left\{ \frac{C_2(\beta)}{\alpha_2}, \frac{C_3(\beta)}{\alpha_3}, C_6(\beta) \right\}.$$

We claim that, for every $\beta \in (0, +\infty)$, $C_{\mathbb{Z}_6}(\beta) = C_6(\beta)$ and the optimal splitting is given by

$$\alpha^{\mathbb{Z}_6}(\beta) = \left(\alpha_2^{\mathbb{Z}_6}(\beta), \alpha_3^{\mathbb{Z}_6}(\beta) \right) = \frac{1}{C_6(\beta)} (C_2(\beta), C_3(\beta)).$$

Indeed we have that

$$\begin{aligned} C_6(\beta) &\geq C_{\mathbb{Z}_6}(\beta) \\ &= \max_{\alpha \in \mathcal{P}(\{2,3\})} \min \left\{ C_6(\beta), \frac{C_2(\beta)}{\alpha_2}, \frac{C_3(\beta)}{\alpha_3} \right\} \\ &\geq \min \left\{ C_6(\beta), \frac{C_2(\beta)}{\alpha_2^G(\beta)}, \frac{C_3(\beta)}{\alpha_3^G(\beta)} \right\} \\ &= C_6(\beta). \end{aligned}$$

In Figure 3 $\alpha_2^{\mathbb{Z}_6}(\beta)$ is plotted: notice how the optimal splitting follows the geometry of the constellation as $\alpha_2(\beta)$ is monotonically increasing in β with $\lim_{\beta \rightarrow 0} \alpha^{\mathbb{Z}_6}(\beta) = (0, 1)$ (as β goes to 0 $K_{2 \times 3}(\beta)$ collapses onto constellation K_3) and $\lim_{\beta \rightarrow +\infty} \alpha^{\mathbb{Z}_6}(\beta) = (1, 0)$ (as β goes to $+\infty$ $K_{2 \times 3}(\beta)$ collapses onto constellation 2-PAM). \square

As we shall see later in Section 5, there are important cases other than the field one when $C_G = C$. In Section 6 we will also exhibit examples where $C_G < C$ and the more general problem of evaluating C_G will be discussed.

Of course up to now it is not at all clear if the G -capacity C_G can actually be achieved by means of G -encoders. In principle there could be other algebraic constraints coming into the picture which we have overlooked in our analysis. In Section 4 we will see that this is not the case: the conditions $R < C_G$ will be proved to be sufficient for reliable transmission using G -encoders over a G -symmetric channel.

4 Classical ensembles of G -codes

We now present a result which completes Theorem 5 by stating that at every rate $R < C_G$ reliable transmission over a G -symmetric channel is possible using G -encoders.

Following the classical technique originally proposed by Shannon we will use a probabilistic method, introducing ensembles of G -encoders and analyzing their average performances. This will then allow us to obtain our result. This technique had already been used to study

performances of linear codes in [19]: this covers the case when $G \simeq \mathbb{Z}_p^k$ for some prime p . For general Abelian group however the derivation is more complicate.

We define an ensemble as a sequence of Abelian groups \mathcal{U}_N and of independent uniformly distributed random variables $\Phi_N \in \text{Hom}(\mathcal{U}_N, G^N)$. We will see later that different choices of ensembles are possible and give similar results.

The above ensemble is completely determined by the sequence \mathcal{U}_N . We now describe the construction of specific examples. Given a design rate $R \in [0, \log |G|]$, and a *splitting* distribution $\alpha \in \mathcal{P}(\mathbf{r}^G)$, for each block length $N \in \mathbb{N}$ define \mathbf{k}_N by

$$(k_N)_{i,j} = \left\lfloor \frac{RN\alpha_{i,j}}{j \log p_i} \right\rfloor. \quad (24)$$

Let $\mathcal{U}_{\mathbf{k}_N}$ be the corresponding Abelian group having type \mathbf{k}_N . The corresponding ensemble will be denoted by $\mathcal{E}_G(R, \alpha)$. Note that, for each N , Φ_N 's rate is a deterministic constant R_N (i.e. it is the same for each realization of Φ_N) with $R_N \leq R$, and $\lim_{N \rightarrow +\infty} R_N = R$.

Let $\overline{P(e|\Phi_N)}^{(R, \alpha)}$ denote the word error probability averaged over the ensemble $\mathcal{E}(R, \alpha)$. Our goal is to estimate this average. To do this we will need to establish a number of preliminary results extending the classical Gallager bound.

4.1 Gallager Bound for codes over groups

In this subsection we state a convenient version of the Gallager bound (see [19]) for the special case of G -symmetric channels; it is based on the techniques presented in [31].

We start by recalling the Gallager bound.

Lemma 6 (Gallager bound) *Given a MC $(\mathcal{X}, \mathcal{Y}, W)$, suppose we have a block encoder*

$$\phi : \mathcal{U} \rightarrow \mathcal{X}^N,$$

and ML decoding is used. Then, for any fixed $u \in \mathcal{U}$ and $\rho \in [0, +\infty)$ the conditioned word error probability satisfies

$$P(e|\phi, u) \leq \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\phi(u))^{\frac{1}{1+\rho}} \left(\sum_{v \neq u} W_N(\mathbf{y}|\phi(v))^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}). \quad (25)$$

We now want to rewrite the Gallager bound in the special case when the channel is G -symmetric for an Abelian group G . It is not restrictive to assume that $\mathcal{X} = G$. We start by introducing some notation of types. Given an arbitrary finite set \mathcal{G} and a vector $\mathbf{x} \in \mathcal{G}^N$, the type of \mathbf{x} is $\boldsymbol{\theta}(\mathbf{x}) \in \mathcal{P}(\mathcal{G})$ defined by letting, for any $x \in \mathcal{G}$, $\theta_x(\mathbf{x})$ be the relative frequency of the symbol x in \mathbf{x} , i.e.

$$\theta_x(\mathbf{x}) = \frac{1}{N} |\{j : 1 \leq j \leq N, x_j = x\}|.$$

The subset of $\mathcal{P}(\mathcal{G})$ containing all types of vectors $\mathbf{x} \in \mathcal{G}^N$ is denoted by $\mathcal{P}_N(\mathcal{G})$. For $\boldsymbol{\theta} \in \mathcal{P}_N(\mathcal{G})$ we define $\mathcal{T}_{\boldsymbol{\theta}}^N$ as the subset of \mathcal{G}^N containing all vectors of type $\boldsymbol{\theta}$.

We now introduce distance spectra of an encoder $\phi : \mathcal{U} \rightarrow G^N$. For each $u \in \mathcal{U}$ and $\boldsymbol{\theta} \in \mathcal{P}_N(G)$ we define $S(\boldsymbol{\theta}|\phi, u)$ as the cardinality of the subset of $\mathcal{U} \setminus \{u\}$ consisting of those v such that the difference $\phi(v) - \phi(u)$ has type $\boldsymbol{\theta}$, i.e.

$$S(\boldsymbol{\theta}|\phi, u) = \sum_{v \in \mathcal{U} \setminus \{u\}} \mathbb{1}_{\mathcal{T}_{\boldsymbol{\theta}}^N}(\phi(v) - \phi(u)). \quad (26)$$

Lemma 7 Given a G -symmetric MC (G, \mathcal{Y}, W) , suppose we have a block encoder

$$\phi : \mathcal{U} \rightarrow G^N,$$

and ML decoding is used. For every $u \in \mathcal{U}$ the conditioned error probability satisfies the following inequality:

$$P(e|\phi, u) \leq \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} \frac{S(\boldsymbol{\theta}|\phi, u)}{\binom{N}{N\boldsymbol{\theta}}} \sum_{\mathbf{x} \in \mathcal{T}_{\boldsymbol{\theta}}^N} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}). \quad (27)$$

Proof: We generate the following random encoder from ϕ :

$$\Phi = \mathbf{G} + \Omega\phi\Pi$$

where:

- Π is a random variable uniformly distributed over the group of permutations of the set \mathcal{U} leaving u fixed;
- Ω is a random variable uniformly distributed over S_N , the group of permutations of $\{1, \dots, N\}$, independent from Π (we intend $\omega \in S_N$ acting on $\mathbf{x} \in G^N$ by permuting its components, i.e. $(\omega\mathbf{x})_i := (\mathbf{x})_{\omega_i}$);
- \mathbf{G} is a random variable uniformly distributed over G^N , independent from Π and Ω .

Throughout the proof we will denote by $\mathbb{E}[\cdot]$ the average operator with respect to such a probabilistic structure. The crucial point here is that the average word error probability of the random encoder Φ is equal to the word error probability of ϕ . In fact for any realization π of Π

$$P(e|\phi\pi, u) = P(e|\phi, \pi u) = P(e|\phi, u).$$

For every $\omega \in S_N$ realization of Ω we have that, due to the memoryless property of the channel, ML decision regions $\Lambda_\phi(v)$ satisfy $\Lambda_{\omega\phi}(v) = \omega\Lambda_\phi(v)$, thus

$$\begin{aligned} P(e|\omega\phi, u) &= 1 - \int_{\Lambda_{\omega\phi}(u)} W_N(\mathbf{y}|\omega\phi u) d\mu^N(\mathbf{y}) \\ &= 1 - \int_{\omega\Lambda_\phi(u)} W_N(\mathbf{y}|\omega\phi u) d\mu^N(\mathbf{y}) \\ &= 1 - \int_{\Lambda_\phi(u)} W_N(\omega\mathbf{y}|\omega\phi u) d\mu^N(\mathbf{y}) \\ &= 1 - \int_{\Lambda_\phi(u)} W_N(\mathbf{y}|\phi u) d\mu^N(\mathbf{y}) \\ &= P(e|\phi, u). \end{aligned}$$

Moreover, due to the G -symmetry of the channel, for any $\mathbf{g} \in G^N$ realization of \mathbf{G} , we have that ML decision regions satisfy $\Lambda_{\mathbf{g}+\phi}(v) = \mathbf{g} + \Lambda_\phi(v)$, so implying

$$P(e|\mathbf{g} + \phi, u) = P(e|\phi, u).$$

Thus we have

$$\mathbb{E}[P(e|\Phi, u)] = P(e|\phi, u).$$

Now fix an arbitrary $\mathbf{x} \in G^N$; we have that

$$\begin{aligned} P(\Phi(u) = \mathbf{x}) &= P(\mathbf{G} + \Omega\phi(\Pi u) = \mathbf{x}) \\ &= \sum_{\mathbf{z} \in G^N} P(\mathbf{G} = \mathbf{x} - \mathbf{z} | \Omega\phi(\Pi u) = \mathbf{z}) P(\Omega\phi(\Pi u) = \mathbf{z}) \\ &= \sum_{\mathbf{z} \in G^N} P(\mathbf{G} = \mathbf{x} - \mathbf{z}) P(\Omega\phi(\Pi u) = \mathbf{z}) \\ &= \sum_{\mathbf{z} \in G^N} \frac{1}{|G|^N} P(\Omega\phi(\Pi u) = \mathbf{z}) = \frac{1}{|G|^N}; \end{aligned} \quad (28)$$

hence $\Phi(u)$ has uniform distribution over G^N . We now want to find out for any fixed $v \in \mathcal{U} \setminus \{u\}$ the conditional distribution of $\Phi(v)$ given $\Phi(u)$. We start by noticing that

$$P(\phi(\Pi v) = \mathbf{x}) = \frac{1}{|\mathcal{U}| - 1} \sum_{w \in \mathcal{U} \setminus \{u\}} \mathbb{1}_{\{\mathbf{x}\}}(\phi(w)). \quad (29)$$

From the independence of Ω , Π and \mathbf{G} and the uniform distribution of \mathbf{G} in G^N , it follows that Ω , Π and $\mathbf{G} + \Omega\phi(u)$ are independent, and so

$$\begin{aligned} P(\Phi(v) = \mathbf{z} + \mathbf{x} | \Phi(u) = \mathbf{x}) &= P(\Phi(v) - \Phi(u) = \mathbf{x} | \Phi(u) = \mathbf{x}) \\ &= P(\Omega\phi(\Pi v) + \mathbf{G} - \Omega\phi(\Pi u) - \mathbf{G} = \mathbf{x} | \mathbf{G} + \Omega\phi(u) = \mathbf{x}) \\ &= P(\Omega\phi(\Pi v) - \Omega\phi(u) = \mathbf{x} | \mathbf{G} + \Omega\phi(u) = \mathbf{x}) \\ &= P(\Omega\phi(\Pi v) - \Omega\phi(u) = \mathbf{x}). \end{aligned} \quad (30)$$

For every $\mathbf{x} \in G^N$ we denote by $Stab(\mathbf{x})$ the stabilizer of \mathbf{x} in S_N , i.e. the subgroup of S_N containing all permutation leaving \mathbf{x} fixed; the cardinality of $Stab(\mathbf{x})$ is

$$(N\boldsymbol{\theta}(\mathbf{x}))! := \prod_{g \in G} (N\theta_g(\mathbf{x}))!. \quad (31)$$

Successively applying (30), (31), (29) and (26) we get

$$\begin{aligned} P(\Phi(v) = \mathbf{z} + \mathbf{x} | \Phi(u) = \mathbf{x}) &= P(\Omega(\phi(\Pi v) - \phi(u)) = \mathbf{x}) \\ &= \sum_{\omega \in S_N} \frac{1}{N!} P(\phi(\Pi v) - \phi(u) = \omega \mathbf{x}) \\ &= \frac{1}{N!} \sum_{\mathbf{y} \in T_{\boldsymbol{\theta}(\mathbf{x})}^N} \sum_{\omega \in Stab(\mathbf{y})} P(\phi(\Pi v) = \phi(u) + \mathbf{y}) \\ &= \frac{1}{N!} \sum_{\mathbf{y} \in T_{\boldsymbol{\theta}(\mathbf{x})}^N} (N\boldsymbol{\theta}(\mathbf{x}))! \frac{1}{|\mathcal{U}| - 1} \sum_{v \in \mathcal{U} \setminus \{u\}} \mathbb{1}_{\{\phi(u) + \mathbf{y}\}}(\phi(v)) \\ &= \binom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1} \sum_{\mathbf{y} \in T_{\boldsymbol{\theta}(\mathbf{x})}^N} \frac{1}{|\mathcal{U}| - 1} \sum_{v \in \mathcal{U} \setminus \{u\}} \mathbb{1}_{\{\mathbf{y}\}}(\phi(v) - \phi(u)) \\ &= \binom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1} \frac{1}{|\mathcal{U}| - 1} \sum_{v \in \mathcal{U} \setminus \{u\}} \mathbb{1}_{T_{\boldsymbol{\theta}(\mathbf{x})}}(\phi(v) - \phi(u)) \\ &= \frac{1}{|\mathcal{U}| - 1} \binom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1} S(\boldsymbol{\theta}(\mathbf{x}) | \phi, u). \end{aligned} \quad (32)$$

We now apply the Gallager bound to each realization of the random encoder Φ . We get

$$\begin{aligned}
P(e|\phi, u) &= \mathbb{E}[P(e|\Phi, u)] \\
&\leq \mathbb{E} \left[\int_{\mathcal{Y}^N} W_N(\mathbf{y}|\Phi(u)) \left(\sum_{v \in \mathcal{U} \setminus \{u\}} \left(\frac{W_N(\mathbf{y}|\Phi(v))}{W_N(\mathbf{y}|\Phi(u))} \right)^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}) \right] \\
&= \mathbb{E} \left[\int_{\mathcal{Y}^N} W_N(\mathbf{y}|\Phi(u))^{\frac{1}{1+\rho}} \left(\sum_{v \in \mathcal{U} \setminus \{u\}} (W_N(\mathbf{y}|\Phi(v)))^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}) \right] \\
&= \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \mathbb{E} \left[\left(\sum_{v \in \mathcal{U} \setminus \{u\}} (W_N(\mathbf{y}|\Phi(v)))^{\frac{1}{1+\rho}} \right)^\rho \middle| \Phi(u) = \mathbf{z} \right] d\mu^N(\mathbf{y}), \tag{33}
\end{aligned}$$

last equality following from (28). The conditional expectation in the last term of (33) can be upperbounded by the Jensen inequality, yielding

$$\begin{aligned}
&\mathbb{E} \left[\left(\sum_{v \in \mathcal{U} \setminus \{u\}} (W_N(\mathbf{y}|\Phi(v)))^{\frac{1}{1+\rho}} \right)^\rho \middle| \Phi(u) = \mathbf{z} \right] \\
&\leq \left(\mathbb{E} \left[\sum_{v \in \mathcal{U} \setminus \{u\}} (W_N(\mathbf{y}|\Phi(v)))^{\frac{1}{1+\rho}} \middle| \Phi(u) = \mathbf{z} \right] \right)^\rho \\
&= \left(\sum_{\mathbf{x} \in G^N} \sum_{v \in \mathcal{U} \setminus \{u\}} P(\Phi(v) = \mathbf{z} + \mathbf{x} | \Phi(u) = \mathbf{z}) (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho \tag{34} \\
&= \left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} \sum_{\mathbf{x} \in \mathcal{T}_\theta^N} S(\boldsymbol{\theta}(\mathbf{x})|\phi, u) \binom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho \\
&= \left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} S(\boldsymbol{\theta}|\phi, u) \binom{N}{N\boldsymbol{\theta}}^{-1} \sum_{\mathbf{x} \in \mathcal{T}_\theta^N} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho
\end{aligned}$$

where the second equality follows from (32) and from $G^N = \bigcup_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} \mathcal{T}_\theta^N$. Substituting (34) into (33) yields (27). \blacksquare

We would like to emphasize the fact that both Definition 26 and Lemma 7 do not need ϕ to be a G -encoder; in what follows we will make use of this generality. When ϕ is a G -encoder it is easy to show that $S(\boldsymbol{\theta}|\phi, u)$ does not depend on u , so in this case we will use the notation $S(\boldsymbol{\theta}|\phi)$. Generalizations of both Definition 26 and Lemma 7 to non Abelian groups are straightforward: the only difference is that one has to define left and right distance spectra (the two notions coincide for group encoders but they generally do not for arbitrary encoders).

4.2 Averaged estimations

The idea is to average estimation (27) on our ensembles. In the field case this would lead us to the classical direct Shannon theorem for linear codes. However, in this context, we

need some more carefulness since averaging distance spectra becomes more delicate. For this we first develop some further considerations on random variables taking values over Abelian groups.

Let M and L be finite Abelian groups and let Φ be a r.v. uniformly distributed on the Abelian group $\text{Hom}(M, L)$. Given $m \in M$, we want to investigate the probability distribution of the r.v.'s $\Phi(m)$. In the case when both M and N are vector space over a finite field \mathbb{F}_{p^r} , it is a standard fact that, if $m \neq 0$, $\Phi(m)$ is a r.v. uniformly distributed over L . In the general case however the analysis is a bit more complicate due to algebraic constraints which show up in the problem. We start with a simple preliminary result.

Suppose we have a finite Abelian group G and a r.v. X uniformly distributed over G . Let H be another Abelian group and $\theta : G \rightarrow H$ a surjective homomorphism.

Lemma 8 $\theta \circ X$ is a r.v. uniformly distributed over H .

Proof: Let $y \in H$. Notice that since θ is surjective, $|\theta^{-1}(y)| = |G|/|H|$ for every y . We now clearly have

$$P(\theta \circ X = y) = P(X \in \theta^{-1}(y)) = \frac{|\theta^{-1}(y)|}{|G|} = \frac{1}{|H|}.$$

■

Let us go back to our setting with the Abelian groups M and L . Given any $m \in M$ we can consider the valuation homomorphism $\psi_{M,L,m} : \text{Hom}(M, L) \rightarrow L$ given by $\psi_{M,L,m}(\phi) = \phi(m)$. Using Lemma 8 we thus obtain that the r.v. $\Phi(m)$ is uniformly distributed on $\text{Im}(\psi_{M,L,m})$. The problem is therefore to characterize the image of $\psi_{M,L,m}$: this depends on the choice of the element m .

We gather a few simple properties of the valuation homomorphism:

Lemma 9 $\psi_{M,L,m}$ satisfies the following properties:

- (1) If $M = \mathbb{Z}_{p^r}$ and $m \in M$ is invertible, we have that $\text{Im}(\psi_{M,L,m}) = L_{(p^r)}$.
- (2) Assume that $M = M_1 \oplus M_2$ and let $m = (m_1, m_2) \in M$. Then, $\text{Im}(\psi_{M,L,m}) = \text{Im}(\psi_{M_1,L,m_1}) + \text{Im}(\psi_{M_2,L,m_2})$.

Assume that M has the structure given by (15) and (16). Each $m \in M$ can be decomposed accordingly

$$m = (m_1, \dots, m_s), \quad m_i = (m_{i,1}, \dots, m_{i,r_i}).$$

For any $i = 1, \dots, s$ and $j = 1, \dots, r_i$, let $l_{i,j} \in \mathbb{Z}^+$ be such that

$$m_{i,j} \in p_i^{j-l_{i,j}} \mathbb{Z}_{p_i^{j}}^{k_{i,j}} \setminus p_i^{j+1-l_{i,j}} \mathbb{Z}_{p_i^j}^{k_{i,j}}.$$

We will use the notation $l_{i,j}(m)$ (and $\mathbf{l}(m)$ in a more compact form) to emphasize the dependence on the chosen m . Clearly, $\mathbf{l}(m)$ is \mathbf{r} -compatible. Finally, given an \mathbf{r} -compatible \mathbf{l} , define

$$H_{\mathbf{l}} = \{m \in M \mid \mathbf{l}(m) = \mathbf{l}\}. \quad (35)$$

Clearly, the various $H_{\mathbf{l}}$ are pairwise disjoint and form a partition of M .

Proposition 10 Let $m \in H_{\mathbf{l}}$. Then,

$$\text{Im}(\psi_{M,L,m}) = \sum_{i=1}^s \sum_{j=1}^{r_i} p_i^{j-l_{i,j}} L_{(p_i^j)}.$$

Proof Immediate consequence of Lemma 9. ■

Corollary 11 *Let $m \in H_1$. Then, the r.v. $\Phi(m)$ is uniformly distributed over the set*

$$\sum_{i=1}^s \sum_{j=1}^{r_i} p_i^{j-l_{i,j}} L_{(p_i^j)}.$$

Notice that the first summation above is direct while the second is not in general. There are relations among the various $p_i^{j-l_{i,j}} L_{(p_i^j)}$ as j varies keeping the index i fixed. Indeed it holds

$$pL_{(p^r)} \subseteq L_{(p^{r-1})} \subseteq L_{(p^r)}.$$

Let us apply these considerations to our context. Recall that we have fixed a G -symmetric MC (G, \mathcal{Y}, W) where G is a finitely generated Abelian group having spectrum $\sigma^G = (p_1, \dots, p_s)$, multiplicity $\mathbf{r}^G = (r_1^G, \dots, r_s^G)$ and type \mathbf{k}^G . Recall moreover that the ensemble $\mathcal{E}_G(R, \boldsymbol{\alpha})$ consists of the sequence of independent random variables Φ_N with Φ_N uniformly distributed over $\text{Hom}(\mathcal{U}_{\mathbf{k}_N}, G^N)$, where \mathbf{k}_N is defined by

$$(k_N)_{i,j} = \left\lfloor \frac{RN\alpha_{i,j}}{j \log p_i} \right\rfloor. \quad (36)$$

For a random variable X will denote by $\overline{X}^{(R, \boldsymbol{\alpha})}$ the average operator with respect to such a probabilistic structure.

We are now ready to prove our first fundamental result:

Theorem 12 *Let (G, \mathcal{Y}, W) be a G -symmetric MC. For every $R \in [0, \log |G|]$, $\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{r}^G)$, the following estimation holds true:*

$$\overline{P(e|\Phi_N)}^{(R, \boldsymbol{\alpha})} \leq \sum_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-compatible}}} \exp(-NE_1(R_{\mathbf{l}})), \quad (37)$$

where E_1 is the error exponent of the subchannel obtained by restricting the input set to $G_{\mathbf{l}}$, and

$$R_{\mathbf{l}} := R \sum_{i=1}^s \sum_{j=1}^{r_i} \frac{l_{i,j}}{j} \alpha_{i,j}.$$

Proof Let $H_{\mathbf{k}_N, \mathbf{l}}$ be the set defined by (35) for the group $\mathcal{U}_{\mathbf{k}_N}$. We can thus decompose

$$\mathcal{U}_{\mathbf{k}_N} = \bigcup_{\substack{\mathbf{l} \\ \mathbf{r}^G\text{-comp.}}} H_{\mathbf{k}_N, \mathbf{l}}. \quad (38)$$

It follows from Corollary 11 that, if $\mathbf{u} \in H_{\mathbf{k}_N, \mathbf{l}}$, $\Phi_N(\mathbf{u})$ is a r.v. uniformly distributed over $G_{\mathbf{l}}^N$.

We now notice that, because of the uniform error property, all estimations of the word error probability can be done assuming that the all-zero information word $\mathbf{u} = \mathbf{0}$ has been transmitted, i.e.

$$P(e|\phi) = P(e|\phi, \mathbf{0})$$

for every $\phi \in \text{Hom}(\mathcal{U}_{\mathbf{k}_N}, G^N)$.

For any \mathbf{r}^G -compatible \mathbf{l} , we define the encoder $\phi_{\mathbf{l}}$ as the restriction of ϕ to the set $\{\mathbf{0}\} \cup H_{\mathbf{k}_N, \mathbf{l}}$. Note that the encoders $\phi_{\mathbf{l}}$ are not G -encoders since their domain is not a group, so that the UEP does not necessarily hold true for them but for ϕ only. Since

$$\{\mathbf{0}\} \cup H_{\mathbf{k}_N, \mathbf{l}} \subseteq \bigoplus_{i=1}^s \bigoplus_{j=1}^{r_i} p_i^{r_i - l_{i,j}} \mathbb{Z}_{p_i}^{(k_N)_{i,j}}$$

$\phi_{\mathbf{l}}$'s rate satisfies

$$\frac{\log(1 + |H_{\mathbf{k}_N, \mathbf{l}}|)}{N} \leq \sum_{i=1}^s \sum_{j=1}^{r_i} \frac{1}{N} \log p_i l_{i,j} (k_N)_{i,j} \leq R_{\mathbf{l}}.$$

A union bound yields

$$P(e|\phi, \mathbf{0}) \leq \sum_{\substack{\mathbf{l} \\ \mathbf{r}^G\text{-comp.}}} P(e|\phi_{\mathbf{l}}, \mathbf{0}) = \sum_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} P(e|\phi_{\mathbf{l}}, \mathbf{0}),$$

(the equality follows from the fact that $H_{\mathbf{k}_N, \mathbf{0}} = \{\mathbf{0}\}$ and thus $P(e|\phi_{\mathbf{0}}, \mathbf{0}) = 0$).

Consider the r.v. $\Phi_{N, \mathbf{l}}$ obtained by restricting Φ_N to the subset $H_{\mathbf{k}_N, \mathbf{l}}$. Now, given an \mathbf{r}^G -compatible \mathbf{l} , apply the bound of Lemma 7 (which, as we already remarked, does not need the encoder to be an homomorphism) to each realization of $P(e|\Phi_{N, \mathbf{l}}, \mathbf{0})$, and then average with respect to Φ_N . For any $\rho \in [0, 1]$ we obtain

$$\begin{aligned} & \overline{P(e|\Phi_{N, \mathbf{l}}, \mathbf{0})}^{(R, \alpha)} \leq \\ & \leq \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \overline{\left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} \frac{S(\boldsymbol{\theta}|\Phi_{N, \mathbf{l}}, 0)}{\binom{N}{N\boldsymbol{\theta}}} \sum_{\mathbf{x} \in \mathcal{T}_{\boldsymbol{\theta}}^N} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^{\rho}}^{(R, \alpha)} d\mu^N(\mathbf{y}) \\ & \leq \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \overline{\left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} \frac{\overline{S(\boldsymbol{\theta}|\Phi_{N, \mathbf{l}}, 0)}^{(R, \alpha)}}{\binom{N}{N\boldsymbol{\theta}}} \sum_{\mathbf{x} \in \mathcal{T}_{\boldsymbol{\theta}}^N} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^{\rho}}^{(R, \alpha)} d\mu^N(\mathbf{y}), \end{aligned} \quad (39)$$

where the last inequality follows from Jensen inequality.

It remains to calculate the average distance spectra of $\Phi_{N, \mathbf{l}}$. Using the fact (see Corollary 11) that for any $\mathbf{u} \in H_{\mathbf{k}_N, \mathbf{l}}$ we have that $\Phi_N(\mathbf{u})$ is uniformly distributed over $G_{\mathbf{l}}^N$, we obtain

$$\begin{aligned} \overline{S(\boldsymbol{\theta}|\Phi_{N, \mathbf{l}}, \mathbf{0})}^{(R, \alpha)} &= \overline{\sum_{\mathbf{u} \in H_{\mathbf{k}_N, \mathbf{l}}} \mathbb{1}_{\mathcal{T}_{\boldsymbol{\theta}}^N}(\Phi_{\mathbf{l}} \mathbf{u})}^{(R, \alpha)} = \sum_{\mathbf{u} \in H_{\mathbf{k}_N, \mathbf{l}}} \overline{\mathbb{1}_{\mathcal{T}_{\boldsymbol{\theta}}^N}(\Phi_{\mathbf{l}} \mathbf{u})}^{(R, \alpha)} \\ &= \sum_{\mathbf{u} \in H_{\mathbf{k}_N, \mathbf{l}}} P(\Phi_{N, \mathbf{l}}(\mathbf{u}) \in \mathcal{T}_{\boldsymbol{\theta}}^N) = |H_{\mathbf{k}_N, \mathbf{l}}| \frac{\binom{N}{N\boldsymbol{\theta}} \mathbb{1}_{\mathcal{P}_N(G_1)}(\boldsymbol{\theta})}{|G_1|^N} \end{aligned} \quad (40)$$

Now fix a set $\Omega_{\mathbf{l}} \subset G^N$ of coset representatives, i.e. a set of cardinality $\frac{|G|^N}{|G_1|^N}$ containing

exactly one element for each coset of G_1^N . By substituting (40) into (39) we obtain

$$\begin{aligned}
& \overline{P(e|\Phi_{N,1}, \mathbf{0})}^{(R,\alpha)} \\
& \leq \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} |H_{\mathbf{k}_N,1}| \frac{1}{|G_1|^N} \mathbb{1}_{\mathcal{P}_N(G_1)}(\boldsymbol{\theta}) \sum_{\mathbf{x} \in \mathcal{T}_{\boldsymbol{\theta}}^N} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}) \\
& = \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} W_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \left(|H_{\mathbf{k}_N,1}| \frac{1}{|G_1|^N} \sum_{\mathbf{x} \in G_1^N} (W_N(\mathbf{y}|\mathbf{z} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}) \\
& = |H_{\mathbf{k}_N,1}|^\rho \int_{\mathcal{Y}^N} \sum_{\mathbf{v} \in \Omega_1} \frac{|G_1|^N}{|G|^N} \sum_{\mathbf{w} \in G_1^N} \frac{1}{|G_1|^N} W_N(\mathbf{y}|\mathbf{v} + \mathbf{w})^{\frac{1}{1+\rho}} \left(\frac{1}{|G_1|^N} \sum_{\mathbf{x} \in G_1^N} (W_N(\mathbf{y}|\mathbf{v} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^\rho d\mu^N(\mathbf{y}) \\
& = |H_{\mathbf{k}_N,1}|^\rho \sum_{\mathbf{v} \in \Omega_1} \frac{|G_1|^N}{|G|^N} \int_{\mathcal{Y}^N} \left(\frac{1}{|G_1|^N} \sum_{\mathbf{x} \in G_1^N} (W_N(\mathbf{y}|\mathbf{v} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu^N(\mathbf{y}) .
\end{aligned} \tag{41}$$

By the G -symmetry of the channel and the memoryless property, we have that, for each $\mathbf{v} \in \Omega_1$,

$$\begin{aligned}
& \int_{\mathcal{Y}^N} \left(\frac{1}{|G_1|^N} \sum_{\mathbf{x} \in G_1^N} (W_N(\mathbf{y}|\mathbf{v} + \mathbf{x}))^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu^N(\mathbf{y}) \\
& = \int_{\mathcal{Y}^N} \left(\frac{1}{|G_1|^N} \sum_{\mathbf{x} \in G_1^N} (W_N((-\mathbf{v})\mathbf{y}|\mathbf{x}))^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu^N(\mathbf{y}) \\
& = \int_{\mathcal{Y}^N} \left(\frac{1}{|G_1|^N} \sum_{\mathbf{x} \in G_1^N} (W_N(\mathbf{y}|\mathbf{x}))^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu^N(\mathbf{y}) \\
& = \left(\int_{\mathcal{Y}} \left(\frac{1}{|G_1|} \sum_{x \in G_1} (W_N(y|x))^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu(y) \right)^N
\end{aligned} \tag{42}$$

where $(-\mathbf{v})\mathbf{y}$ denotes the action of each component of $-\mathbf{v}$ on the corresponding component of \mathbf{y} (recall that by definition of G -symmetric channel, G isometrically acts on \mathcal{Y}). Therefore,

$$\overline{P(e|\Phi_{N,1}, \mathbf{0})}^{(R,\alpha)} \leq |H_{\mathbf{K},1}|^\rho \left(\int_{\mathcal{Y}} \left(\frac{1}{|G_1|} \sum_{x \in G_1} (W_N(y|x))^{\frac{1}{1+\rho}} \right)^{1+\rho} d\mu(y) \right)^N .$$

Recalling that the random coding exponent $E_1(R)$ is obtained with uniform distribution over the input set G_1 , and since the choice of $\rho \in [0, 1]$ is arbitrary, we can rewrite the last inequality as

$$\overline{P(e|\Phi_1, \mathbf{0})}^{(R,\alpha)} \leq \exp(-NE_1(R_1)) .$$

Now (37) follows because E_1 is a non increasing function. ■

Define now two important figures. The G -random coding exponent is

$$E_G(R) = \max_{\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{r}^G)} \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G \text{-comp.}}} E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} \right) , \tag{43}$$

while the G -optimal splitting rate function is defined by letting, for every $R \in [0 \log |G|]$, $\alpha^G(R)$ be one of the elements of $\mathcal{P}(\mathbf{r}^G)$ for which the maximum in (43) is achieved, i.e.

$$\min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}^G(R) \right) = \max_{\alpha \in \mathcal{P}(\mathbf{r}^G)} \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} \right) \quad (44)$$

Since $\mathcal{P}(\mathbf{r}^G)$ is compact and $f_R(\alpha) = \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} \right)$ is continuous from $\mathcal{P}(\mathbf{r}^G)$ to \mathbb{R} for every $R \in [0, \log |G|]$, the above definition of $\alpha^G(R)$ is coherent since f_R has at least (but not necessarily only) one maximum point in $\mathcal{P}(\mathbf{r}^G)$.

We can now state the following result which is an easy consequence of Theorem 12.

Corollary 13 *Consider a G -symmetric memoryless channel of G -capacity C_G , G -random coding exponent $E_G(R)$, G -optimal splitting rate function $\alpha^G(R)$. Then, $E_G(R) > 0$ if and only if $R < C_G$ and*

$$\overline{P(e|\Phi_N)}^{(R, \alpha^G(R))} \leq A_G \exp(-NE_G(R)), \quad (45)$$

where

$$A_G = |\{\mathbf{l} \neq \mathbf{0}, \mathbf{r}^G\text{-compatible}\}| = \sum_{i=1}^s \frac{r_i^G(r_i^G + 3)}{2} - 1. \quad (46)$$

Proof

Notice that, if $\mathbf{l} \neq \mathbf{0}$ and α is any splitting, we have that

$$E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} \right) > 0 \Leftrightarrow R \sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} < C_1 \Leftrightarrow R < \frac{C_1}{\sum_{i=1}^s \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}}.$$

Hence,

$$\min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i} \frac{l_{i,j}}{j} \alpha_{i,j} \right) > 0 \Leftrightarrow R < \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} \frac{C_1}{\sum_{i=1}^s \sum_{j=1}^{r_i} \frac{l_{i,j}}{j} \alpha_{i,j}}. \quad (47)$$

By choosing $\alpha = \alpha^G$ (the G -optimal splitting for which C_G is achieved), we thus obtain

$$\min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-comp.}}} E_1 \left(R \sum_{i=1}^s \sum_{j=1}^{r_i} \frac{l_{i,j}}{j} \alpha_{i,j}^G \right) > 0 \Leftrightarrow R < C_G. \quad (48)$$

This clearly implies that if $R < C_G$ then $E_G(R) > 0$. Actually, Theorem 5 immediately implies that

$$E_G(R) > 0 \Leftrightarrow R < C_G. \quad (49)$$

Using now Theorem 12 we obtain the result. ■

Remark: It follows from the proof of Corollary 13 that using input groups corresponding to the G -optimal splitting α^G , we can reach C_G -capacity. However, in order to obtain best mean rate of convergence one has to use input groups corresponding to the splitting $\alpha_G(R)$

which in general is a function of the rate R . Straightforward continuity arguments allow to show that $\alpha^G(R)$ can always be chosen in such a way that $\alpha^G(C_G) = \alpha^G$. Notice that in the cyclic example $G = \mathbb{Z}_{p^r}$ ($s = 1$) it was already proven that $\alpha^G(R)_r = 1$ and $\alpha^G(R)_j = 0$ if $j < r$. This corresponds to take $\mathcal{U} = \mathbb{Z}_{p^r}^{\lfloor RN \rfloor}$. In other words free input groups over \mathbb{Z}_{p^r} in this case suffice to achieve G -symmetric capacity.

Standard probabilistic considerations allow us to state the following.

Corollary 14 *Consider a G -symmetric memoryless channel of G -capacity C_G , G -random coding exponent $E_G(R)$, and G -optimal splitting rate function $\alpha^G(R)$. The ensemble $\mathcal{E}(R, \alpha^G(R))$ satisfies*

$$P_G \left(\liminf_N \frac{-\log P(e|\Phi_N)}{N} \geq E_G(R) \right) = 1, \quad (50)$$

where P_G denotes the probability on the ensemble.

Proof: For any $\varepsilon \in (0, E_G(R))$, $N \in \mathbb{N}$ define the event A_N^ε as

$$A_N^\varepsilon := \{P(e|\Phi_N) \geq A_G \exp(-N(E_G(R) - \varepsilon))\},$$

where A_G is defined in (46). By applying (45) and the Markov inequality to each r.v. $P(e|\Phi_N)$ we obtain

$$P_G(A_N^\varepsilon) \leq P_G \left(P(e|\Phi_N) \geq \exp(N\varepsilon) \overline{P(e|\Phi_N)}^{(R, \alpha^G(R))} \right) \leq \exp(-N\varepsilon).$$

Then

$$\sum_{N=1}^{+\infty} P(A_N^\varepsilon) \leq \sum_{N=1}^{+\infty} \exp(-N\varepsilon) < +\infty. \quad (51)$$

Let us denote by $\{A_N^\varepsilon \text{ i.o.}\}$ the event ' A_N^ε occurs infinitely many times ', i.e.

$$\{A_N^\varepsilon \text{ i.o.}\} := \bigcap_{k \in \mathbb{N}} \bigcup_{N \geq k} A_N^\varepsilon.$$

By Borel Cantelli theorem, (51) implies that $P_G(A_N^\varepsilon \text{ i.o.}) = 0$ for every $\varepsilon > 0$. But clearly

$$\{A_N^\varepsilon \text{ i.o.}\}^c \subseteq \left\{ \liminf_N \frac{-\log P(e|\Phi_N)}{N} \geq E_G(R) - \varepsilon \right\}.$$

By the σ -additivity of P_G , this implies

$$P_G \left(\liminf_N \frac{-\log P(e|\Phi_N)}{N} \geq E_G(R) \right) = 1. \quad \blacksquare$$

Corollary 15 *Consider a G -symmetric channel whose G -capacity is C_G . Then, for every $R < C_G$ and for every $\varepsilon > 0$, there exists a G -encoder ϕ_G , of rate greater than or equal to R , whose ML decoding word error probability satisfies*

$$P(e|\phi_G) < \varepsilon. \quad (52)$$

Proof: Trivial consequence of Corollary 14. \blacksquare

4.3 On tightness of the error exponent

In the previous subsection an upper bound to the average word error probability of the G-codes ensembles has been derived, consisting in a term which is exponentially decreasing to 0 in the block length for every rate below C_G . We now want to address the question whether this bound is exponentially tight or not.

First we want to specify what we actually mean by 'tight'. Consider a memoryless channel $(\mathcal{X}, \mathcal{Y}, W)$. It is a well known fact (see [19], [36], [1]) that the random coding exponent in (7) is given by

$$E(R) = \begin{cases} R_0 - R, & 0 \leq R \leq R_{cr} \\ E_{sp}(R), & R_{cr} \leq R \leq C. \end{cases} \quad (53)$$

where R_{cr} is the so called critical rate, R_0 the cutoff rate, and $E_{sp}(R)$ the sphere packing exponent, all functions of the channel $\{W\}$ only.

For the classical Shannon random coding ensemble the error exponent is tight for any deterministic sequence of codes only for $R \geq R_{cr}$, while this is not the case for low rates $R < R_{cr}$: in fact in this case expurgation techniques lead to the existence of sequences of codes guaranteeing higher error exponents. There are conjectures ([36], [15]) about the actual achievable error exponent (the so called reliability function of the channel) at any rate $R \in [0, C]$, but still no completely proved results.

Nevertheless it was proved in [20] that $E(R)$ is tight for the average code from the classical random coding ensemble at any rate, i.e.

$$\lim_{N \rightarrow \infty} -\frac{\log \overline{P(e|\Phi_N)}}{N} = E(R), \quad \forall 0 \leq R \leq C. \quad (54)$$

Moreover, when dealing with a channel which is symmetric with respect to the action of a Galois field \mathbb{F}_q (as for instance a binary-input symmetric-output channel), it is well known that (54) holds true for the \mathbb{F}_q -linear coding ensemble. The proof of this fact, although probably never explicitly published yet ([15]), can be obtained with a slight modification of Gallager's proof in [20]. Indeed, a closer look at [20] shows that the fundamental ingredients of that proof in the special case of \mathbb{F}_q -symmetrical channels are uniform distribution of the codewords over \mathbb{F}_q^N and their pairwise independence. As these two properties are preserved when moving from the random coding ensemble to the \mathbb{F}_q -linear one, almost the same proof of [20] can be carried on showing that (54) continues to hold true in this case.

We conjecture that the G -error exponent is tight in the latter sense, i.e. that

$$\lim_{N \rightarrow +\infty} -\frac{\log \overline{P(e|\Phi_N)}^{(R, \alpha_G(R))}}{N} = E_G(R). \quad (55)$$

We have not yet a complete proof of (55), but only some partial results, and we want to explain them in the simple special case when $G = \mathbb{Z}_4$.

Let $0 < R \leq C_{\mathbb{Z}_4}$. The \mathbb{Z}_4 -coding ensemble is the sequence of independent random variables $(\Phi_N)_N$, with each Φ_N uniformly distributed over $\text{Hom}(\mathbb{Z}_4^{k_N}, \mathbb{Z}_4^N)$, where $k_N := \lfloor \frac{RN}{\log 4} \rfloor$. The \mathbb{Z}_4 -random coding exponent of the channel is

$$E_{\mathbb{Z}_4}(R) = \min \{E_1(R/2), E_2(R)\},$$

where, as usual, $E_2(R)$ and $E_1(R)$ are, respectively, the random coding exponent of the \mathbb{Z}_4 -symmetric channel, and of its $2\mathbb{Z}_4$ -symmetric subchannel. In this case partition (35) reduces to

$$\mathbb{Z}_4^{k_N} = \{\mathbf{0}\} \cup H_{\mathbf{k}_N, 1} \cup H_{\mathbf{k}_N, 2}$$

where $H_{\mathbf{k}_N,1} = 2\mathbb{Z}_4^N \setminus \{\mathbf{0}\}$, $H_{\mathbf{k}_N,2} = \mathbb{Z}_4^{k_N} \setminus 2\mathbb{Z}_4^{k_N}$. Consider the random encoders

$$\Phi_{N,1} := \Phi_N|_{H_{\mathbf{k}_N,1} \cup \{\mathbf{0}\}}, \quad \Phi_{N,2} := \Phi_N|_{H_{\mathbf{k}_N,2} \cup \{\mathbf{0}\}}.$$

We have, by successively applying the UEP, property (2) of ML decoding, and Jensen inequality (notice that function $\mathbb{R}^d \ni \mathbf{x} \mapsto \max_{1 \leq i \leq d} x_i \in \mathbb{R}$ is convex),

$$\begin{aligned} \overline{P(e|\Phi_N)}^{(R, \alpha_G(R))} &= \overline{P(e|\Phi_N, \mathbf{0})}^{(R, \alpha_G(R))} \\ &\geq \max\{P(e|\Phi_{N,1}, \mathbf{0}), P(e|\Phi_{N,2}, \mathbf{0})\}^{(R, \alpha_G(R))} \\ &\geq \max\left\{\overline{P(e|\Phi_{N,1}, \mathbf{0})}^{(R, \alpha_G(R))}, \overline{P(e|\Phi_{N,2}, \mathbf{0})}^{(R, \alpha_G(R))}\right\}. \end{aligned} \quad (56)$$

Now we clearly have that $\Phi_{N,1}\mathbf{x} = \Phi_N\mathbf{x}$ is uniformly distributed over $2\mathbb{Z}_4^N$ for every $\mathbf{x} \in H_{\mathbf{k}_N,1}$ and $\Phi_{N,2}\mathbf{x} = \Phi_N\mathbf{x}$ is uniformly distributed over \mathbb{Z}_4^N for every $\mathbf{x} \in H_{\mathbf{k}_N,2}$. Moreover $\Phi_{N,1}\mathbf{x}$ and $\Phi_{N,1}\mathbf{y}$ are independent for every $\mathbf{x}, \mathbf{y} \in H_{\mathbf{k}_N,1}$ such that $\mathbf{x} \neq \mathbf{y}$. Indeed $(\Phi_{N,1})_N$ is the binary linear ensemble (identifying $2\mathbb{Z}_4$ with the binary field \mathbb{F}_2), so that from the previous observations we know that the random coding exponent $E_1(R/2)$ is tight for the term $\overline{P(e|\Phi_{N,1}, \mathbf{0})}^{(R, \alpha_G(R))}$, i.e.

$$\lim_{N \rightarrow \infty} \frac{\log \overline{P(e|\Phi_{N,1}, \mathbf{0})}^{(R, \alpha_G(R))}}{N} = E_1(R/2), \quad 0 \leq R \leq C_1. \quad (57)$$

Instead, two r.v.s $\Phi_{N,2}\mathbf{x}$ and $\Phi_{N,2}\mathbf{y}$ are independent only for those $\mathbf{x}, \mathbf{y} \in H_{\mathbf{k}_N,2}$ such that $\mathbf{x} - \mathbf{y} \in H_{\mathbf{k}_N,2}$; otherwise, when $\mathbf{x} - \mathbf{y} \in H_{\mathbf{k}_N,1}$, then $\Phi_{N,2}\mathbf{x}$ has uniform distribution over the coset $\Phi_{N,2}\mathbf{y} + 2\mathbb{Z}_4^N$. In this case Gallager's arguments cannot be directly applied to obtain a tightness result at low rates for the term $\overline{P(e|\Phi_{N,2}, \mathbf{0})}^{(R, \alpha_G(R))}$ (though we conjecture they can be properly modified to get the desired result), so that we actually only have that

$$\lim_{N \rightarrow \infty} \frac{\log \overline{P(e|\Phi_{N,1}, \mathbf{0})}^{(R, \alpha_G(R))}}{N} = E_2(R), \quad R_{cr,2} \leq R \leq C_2, \quad (58)$$

where $R_{cr,2}$ denotes the critical rate of the \mathbb{Z}_4 -symmetrical channel. By combining (56) with (57) and (58), we obtain that

$$\lim_{N \rightarrow \infty} \frac{\log \overline{P(e|\Phi_N)}^{(R, \alpha_G(R))}}{N} = E_{\mathbb{Z}_4}(R), \quad \text{whenever } E_{\mathbb{Z}_4}(R) = E_1(R) \text{ or } R_{cr,2} \leq R \leq C_{\mathbb{Z}_4}. \quad (59)$$

We observe that the first condition in (59) is surely holding at very low rates: indeed

$$\lim_{R \rightarrow 0} E_1(R/2) = E_1(0) \leq E_2(0) = \lim_{R \rightarrow 0} E_2(R), \quad (60)$$

with strict inequality holding true in (60) for nontrivial channels. So we can conclude that, even for \mathbb{Z}_4 -symmetric channels for which $C_{\mathbb{Z}_4} = C_4$ so that there is no loss of capacity, there is a loss in the average error exponent at low rates when restricting from Shannon's random coding ensemble to the \mathbb{Z}_4 -code ensemble.

Similar considerations can be extended to a generic finite Abelian group G , showing that, when G does not admit Galois field structure (i.e. when G is not isomorphic to any \mathbb{Z}_p^r), then even if the G -capacity coincides with Shannon one, restricting to G -encoders causes a loss in the average error exponent at low rates.

4.4 The parity check ensemble

There is another way to represent Abelian group codes. Instead of using the encoder image representation, one can as well use kernel representations. We essentially obtain the same codes, however the probabilistic ensembles present certain differences.

Given a design rate R and a splitting $\alpha \in \mathcal{P}(\mathbf{r}^G)$, for each block length $N \in \mathbb{N}$ we define \mathbf{h}_N by

$$(\mathbf{h}_N)_{i,j} = \left\lceil \frac{RN(1 - \alpha_{i,j})}{j \log p_i} \right\rceil$$

Let $\mathcal{V}_{\mathbf{h}_N}$ the corresponding Abelian group having type \mathbf{h}_N . Consider a sequence of independent r.v.s Φ'_N uniformly distributed over $\text{Hom}(G^N, \mathcal{V}_{\mathbf{h}_N})$. Let $\mathcal{U}_N = \ker(\Phi'_N)$ the corresponding sequence of independent r.v. taking values in the set of subgroups of G^N , and finally let

$$\Phi_N : \mathcal{U}_N \hookrightarrow G^N$$

the immersion of \mathcal{U}_N in G^N . The corresponding ensemble will be denoted by $\mathcal{E}'(R, \alpha)$. Notice that for this ensemble the rate of Φ_N is a r.v. R_N ; indeed it can be proved that $P'_G \left(\lim_N R_N = R \right) = 1$ (P'_G denotes the probability with respect to this new ensemble).

Let $\overline{P(e|\Phi_N)}^{(R, \alpha')}$ denote the word error probability averaged over this ensemble. Using techniques very similar to those used to upperbound $\overline{P(e|\Phi_N)}^{(R, \alpha)}$ it is possible to prove the following estimation, which constitutes an analogous of Theorem 12.

Theorem 16 *Let (G, \mathcal{Y}, W) a G -symmetric MC. For every $R \in [0, \log |G|]$, $\alpha \in \mathcal{P}(\mathbf{r}^G)$,*

$$\overline{P(e|\Phi_N)}^{(R, \alpha')} \leq \sum_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G\text{-compatible}}} \exp(-NE_1(R_{\mathbf{l}}))$$

where $E_1(R)$ is the error exponent of the subchannel obtained by restricting the input to the subgroup $G_{\mathbf{l}}$ and

$$R_{\mathbf{l}} := R \sum_{i=1}^s \sum_{j=1}^{r_i} \frac{l_{i,j}}{j} \alpha_{i,j}.$$

5 \mathbb{Z}_{p^r} -codes for p^r -PSK do achieve capacity of the AWGN channel!

In this section we will consider MCs having as input set the m -PSK constellation

$$K_m = \{\xi_m^k, k = 0, \dots, m-1\}$$

where, we recall, $\xi_m := e^{\frac{2\pi}{m}i}$. Notice that K_m is a subgroup of the multiplicative subgroup of non-zero complex numbers \mathbb{C}^* .

The following definition captures many interesting channels, among which the 2-dimensional K_m -AWGN channel.

Definition 17 *A K_m additive isotropic decreasing noise (K_m -AIDN) channel is a memoryless channel (K_m, \mathcal{Y}, W) where*

- $\mathcal{Y} = (Y, \mathcal{B}, \mu)$ where Y is a closed subgroup of \mathbb{C}^* such that $K_m \leq Y$, \mathcal{B} is the corresponding Borel σ -algebra, and μ is a measure over \mathcal{B} ;

- the transition laws $W(y|x)$ only depend on the distance $|x - y|$ and such dependence is monotonically decreasing, i.e. there exists a decreasing function $\zeta : [0, +\infty) \rightarrow [0, +\infty)$ such that $W(y|x) = \zeta(|y - x|)$.

This rather abstract definition allows us to treat at once many different widely used symmetric channels with input K_m and either continuous or discrete output. Notice that from Def.17 it follows that any K_m -AIDN is \mathbb{Z}_m -symmetric, since \mathbb{Z}_m is a generating group for K_m , \mathbb{Z}_m isometrically acts on \mathcal{Y} (by rotations), and $W(gy|gx) = \zeta(|gy - gx|) = \zeta(|x - y|) = W(y|x)$. We show some examples of K_m -AIDN channels.

Example 9 Both the unquantized K_m -AWGN channel and the unquantized K_m -Laplacian channel are AIDN channels. \square

Next example shows how discrete output K_m -AIDN channels can be obtained from continuous output ones by a proper quantization.

Example 10 Suppose an unquantized K_m -AIDN channel (a K_m -AWGN for instance)

$$(K_m, \mathbb{C}^*, W) \quad (61)$$

is given, and let $\zeta : [0, +\infty) \rightarrow [0, +\infty)$ the decreasing function such that $W(y|x) = \zeta(|y - x|)$ (whose existence is guaranteed by the Def.17). For every positive integer m' such that $m \mid m'$, we can introduce a new, discrete output, K_m -AIDN channel by quantizing the output of (61) over Voronoi regions of the $K_{m'}$ constellation. Explicitly such channel is given by

$$(K_m, K_{m'}, W') \quad (62)$$

where the output $K_{m'}$ is equipped with the counting measure μ' , and transition laws are given by

$$W'(\xi_{m'}^k | \xi_m^j) := \int_V W(\xi_{m'}^k, t | \xi_m^j) d\mu(t) = \int_V \zeta(|\xi_{m'}^k t - \xi_m^j|) d\mu(t) = \int_V \zeta\left(\left|\xi_{m'}^{k-j\frac{m'}{m}} t - 1\right|\right) d\mu(t),$$

and V is the Voronoi region of $1 = \xi_{m'}^0 \in K_{m'}$, defined as

$$V := \left\{ \rho e^{i\theta} \in \mathbb{C} : \rho > 0, -\frac{2\pi}{2m'} \leq \theta \leq \frac{2\pi}{2m'} \right\}.$$

Clearly $K_m \leq K_{m'} \leq \mathbb{C}$, so that, in order to see that channel (62) fulfil Def.17, it remains to show that the second requirement is fulfilled. We start by noticing that for every $\rho > 0$, $\theta \in \mathbb{R}$,

$$\begin{aligned} \left| \xi_{m'}^k \rho e^{i\theta} - \xi_m^j \right|^2 &= \left(\rho \cos\left(\theta + \frac{2\pi}{m'}k\right) - \cos\left(j\frac{2\pi}{m}\right) \right)^2 + \left(\rho \sin\left(\theta + \frac{2\pi}{m'}k\right) - \sin\left(j\frac{2\pi}{m}\right) \right)^2 \\ &= \rho^2 + 1 - 2\rho \left(\cos\left(\theta + \frac{2\pi}{m'}k\right) \cos\left(j\frac{2\pi}{m}\right) + \sin\left(\theta + \frac{2\pi}{m'}k\right) \sin\left(j\frac{2\pi}{m}\right) \right) \\ &= \rho^2 + 1 + 2\rho \cos\left(\theta + \frac{2\pi}{m'}k - j\frac{2\pi}{m}\right). \end{aligned} \quad (63)$$

From (63), and from the fact that $\cos(x) = \cos(y)$ if and only if $x = \pm y \pmod{2\pi}$, it immediately follows that for every couple of values $k, k' \in \mathbb{Z}_{m'}$

$$\left| \xi_{m'}^k - \xi_m^j \right| = \left| \xi_{m'}^{k'} - \xi_m^j \right| \iff k - \frac{m'}{m}j = -k' + \frac{m'}{m}j \text{ or } k = k'. \quad (64)$$

Then for $k \neq k' \in \mathbb{Z}_{m'}$ such that $|\xi_{m'}^k - \xi_m^j| = |\xi_{m'}^{k'} - \xi_m^j|$, we have

$$\begin{aligned}
W'(\xi_{m'}^k | \xi_m^j) &= \int_V \zeta \left(\left| \xi_{m'}^{k-j\frac{m'}{m}} t - 1 \right| \right) d\mu(t) \\
&= \int_V \zeta \left(\left| \xi_{m'}^{-k'+j\frac{m'}{m}} t - 1 \right| \right) d\mu(t) \\
&= \int_V \zeta \left(\left| \xi_{m'}^{-k'+j\frac{m'}{m}} t^{-1} - 1 \right| \right) d\mu(t) \\
&= \int_V \zeta \left(\left| \xi_{m'}^{k'-j\frac{m'}{m}} t - 1 \right| \right) d\mu(t) = W'(\xi_{m'}^{k'} | \xi_m^j)
\end{aligned} \tag{65}$$

where we exploited (64), the symmetry property of Voronoi region $V = V^{-1}$, and finally (63). Equality (65) clearly implies that transition laws $W'(y|x)$ are function of the distance $|x - y|$, i.e. we can define a function $\zeta' : [0, +\infty) \rightarrow [0, +\infty)$ such that

$$\zeta'(|\xi_{m'}^k - \xi_m^j|) := \int_V \zeta \left(\left| \xi_{m'}^{k-j\frac{m'}{m}} t - 1 \right| \right) d\mu(t) = W'(\xi_{m'}^k | \xi_m^j),$$

arbitrarily interpolating $\zeta'(z)$ for values of z not included in the set $\{|\xi_{m'}^k - \xi_m^j|, k \in \mathbb{Z}_{m'}, j \in \mathbb{Z}_m\}$. At this point we are only left to prove that ζ' can be chosen decreasing; clearly it suffices to show that

$$|\xi_{m'}^k - \xi_m^j| < |\xi_{m'}^{k'} - \xi_m^j| \implies \zeta'(|\xi_{m'}^k - \xi_m^j|) \geq \zeta'(|\xi_{m'}^{k'} - \xi_m^j|). \tag{66}$$

Due to (64) it is sufficient to show (66) for value of k and k' such that $j\frac{m'}{m} \leq k, k' \leq j\frac{m'}{m} + \frac{m'}{2}$. Notice that if

$$0 \leq k - j\frac{m'}{m} < k' - j\frac{m'}{m} \leq \frac{m'}{2}, \quad -\frac{2\pi}{2m'} \leq \theta \leq \frac{2\pi}{2m'}, \quad \rho > 0$$

then

$$\begin{aligned}
\left| \xi_{m'}^k \rho e^{i\theta} - \xi_m^j \right|^2 &= \rho^2 + 1 + 2\rho \cos \left(\theta + \frac{2\pi}{m'} \left(k - j\frac{m'}{m} \right) \right) \\
&\geq \rho^2 + 1 + 2\rho \cos \left(\theta + \frac{2\pi}{m'} \left(k' - j\frac{m'}{m} \right) \right) \\
&= \left| \xi_{m'}^{k'} \rho e^{i\theta} - \xi_m^j \right|^2,
\end{aligned}$$

from which it follows that

$$\begin{aligned}
W'(\xi_{m'}^k | \xi_m^j) &= \int_V \zeta \left(\left| \xi_{m'}^k t - \xi_m^j \right| \right) d\mu(t) \\
&\geq \int_V \zeta \left(\left| \xi_{m'}^{k'} t - \xi_m^j \right| \right) d\mu(t) \\
&= W'(\xi_{m'}^{k'} | \xi_m^j)
\end{aligned}$$

where we made use of the decreasing property of ζ . □

We conclude our series of examples with the following one.

Example 11 Let $S^1 = \{e^{i\theta}\} \subset \mathbb{C}$ be the complex unitary circumference. Consider an unquantized K_m -AIDN channel of type (61). We define a new channel by projecting the output \mathbb{C}^* onto S^1 . Explicitly we consider the channel

$$(K_m, \mathcal{Y} = (S^1, \mathcal{B}', \mu'), \{W(\cdot|x)\}_{x \in K_m} \in \mathcal{P}(\mathcal{Y})) \quad (67)$$

where μ' is the Lebesgue measure of S^1 , and

$$W(y|x) := \int_0^{+\infty} \zeta(|\rho y - x|) \rho d\mu''(t)$$

and μ'' is the Lebesgue measure of \mathbb{R} .

The verification that (67) is an AIDN channel is almost the same as that of Example 10.

□

Now fix a prime number p and a positive integer r ; throughout the rest of the present section the base of log (and thus of the entropy function H) will be p . For a function $f : \mathcal{Y} \rightarrow \mathbb{R}$ we write $\{f > 0\}$ to denote the set $\{y \in \mathcal{Y} : f(y) > 0\}$.

In the following we will deal with K_{p^r} -AIDN channels; we will prove that for this class of channels the Shannon capacity C_{p^r} and the \mathbb{Z}_{p^r} -capacity $C_{\mathbb{Z}_{p^r}}$ do coincide. Recall that, by definition

$$C_{\mathbb{Z}_{p^r}} = \min_{l=1, \dots, r} \frac{r}{l} C_l ,$$

where C_l is the Shannon capacity of the p^l -th channel, i.e. the channel obtained by constraining the input on the K_{p^l} constellation.

Hence, our result is equivalent to say that

$$rC_{p^l} \geq lC_{p^r} , \quad \forall l, r : 1 \leq sl \leq r . \quad (68)$$

Notice that a simple inductive argument shows that (68) is equivalent to

$$qC_{p^{q+1}} \leq (q+1)C_{p^q} , \quad \forall q = 1, \dots, r-1 \quad (69)$$

The rest of section will be devoted to the proof of (69). The result will be achieved through a series of technical intermediate steps. We start with some notation.

Given a K_{p^q} -AIDN channel $(K_{p^q}, \mathcal{Y}, W)$ we define some connected probability densities which will play a key role in the following:

- for every $y \in \mathcal{Y}$, $1 \leq q \leq r$,

$$\lambda_q(y) := \frac{1}{p^q} \sum_{x \in K_{p^q}} W(y|x) = \frac{1}{p^q} \sum_{j=0}^{p^q-1} W(y \xi_{p^q}^j | 1) ;$$

(second equality follows from the \mathbb{Z}_{p^q} -symmetry of the channel);

- for every $1 \leq q \leq r-1$ and $y \in \mathcal{Y}$ such that $\lambda_{q+1}(y) > 0$,

$$\nu_q(y) := \frac{1}{p \lambda_{q+1}(y)} \left(\lambda_q(y \xi_{p^{q+1}}^0), \lambda_q(y \xi_{p^{q+1}}^1), \dots, \lambda_q(y \xi_{p^{q+1}}^{p-1}) \right) ;$$

- for every $1 \leq q \leq r$ and $y \in \mathcal{Y}$ such that $\lambda_q(y) > 0$,

$$\omega_q(y) := \frac{1}{p^q \lambda_q(y)} \left(W(y|\xi_{p^q}^0), W(y|\xi_{p^q}^1), \dots, W(y|\xi_{p^q}^{p^q-1}) \right) .$$

Notice that:

- $\lambda_q \in \mathcal{P}(\mathcal{Y})$;
- for any fixed $y \in \mathcal{Y}$ such that $\lambda_{q+1}(y) > 0$, $\omega_q(y) \in \mathcal{P}(p^q)$;
- for any fixed $y \in \mathcal{Y}$ such that $\lambda_q(y) > 0$, $\nu_q(y) \in \mathcal{P}(p)$;
- $K_{p^{q+1}} = \bigcup_{0 \leq k < p} \xi_{p^{q+1}}^k K_{p^q}$, and therefore, for $y \in \mathcal{Y}$,

$$\lambda_{q+1}(y) = \frac{1}{p} \sum_{k=1}^p \lambda_q \left(y \xi_{p^{q+1}}^k \right) . \quad (70)$$

For any $q = 1, \dots, r$ consider the p^q -th subchannel. Since this subchannel is \mathbb{Z}_{p^q} -symmetric (in fact it is a K_{p^q} AIDN channel), its Shannon capacity C_{p^q} is obtained by uniform distribution over the input set K_{p^q} . The corresponding distribution over the output set \mathcal{Y} is described by

$$P_Y(y) = \sum_{x \in K_{p^q}} p^{-q} W(y|x) = \lambda_q(y) .$$

So

$$C_{p^q} = H(\lambda_q) - H(W(\cdot|1)) . \quad (71)$$

Therefore (69) is equivalent to

$$H(W(\cdot|1)) + qH(\lambda_{q+1}) \leq (q+1)H(\lambda_q) , \quad q = 1, \dots, r-1 . \quad (72)$$

Next lemma shows how the entropies of the families of probability laws $\omega_q(y)$ and $\nu_q(y)$ come out in (72).

Lemma 18 For every $q = 1, \dots, r-1$,

- $$H(W(\cdot|1)) = H(\lambda_q) - q + \int_{\{\lambda_q > 0\}} \lambda_q(x) H(\omega_q(x)) d\mu(x) ; \quad (73)$$

- $$H(\lambda_q) = H(\lambda_{q+1}) - 1 + \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(x) H(\nu_q(x)) d\mu(x) . \quad (74)$$

Proof:

We have

$$\begin{aligned}
H(W(\cdot|1)) &= - \int_{\mathcal{Y}} W(y|1) \log W(y|1) d\mu(y) \\
&= - \frac{1}{p^q} \sum_{k=0}^{p^q-1} \int_{\mathcal{Y}} W(y|\xi_{p^q}^k|1) \log W(y|\xi_{p^q}^k|1) d\mu(y) = - \frac{1}{p^q} \sum_{k=0}^{p^q-1} \int_{\{\lambda_q > 0\}} W(y|\xi_{p^q}^k|1) \log W(y|\xi_{p^q}^k|1) d\mu(y) \\
&= - \int_{\{\lambda_q > 0\}} \left[\frac{1}{p^q} \sum_{k=0}^{p^q-1} W(y|\xi_{p^q}^k) \right] \log \lambda_q(x) d\mu(y) - \int_{\{\lambda_q > 0\}} \lambda_q(y) \sum_{k=0}^{p^q-1} \frac{W(y|\xi_{p^q}^k)}{p^q \lambda_q(y)} \log \frac{W(y|\xi_{p^q}^k)}{\lambda_q(y)} d\mu(y) \\
&= - \int_{\{\lambda_q > 0\}} \lambda_q(y) \log \lambda_q(y) d\mu(y) - \int_{\{\lambda_q > 0\}} \lambda_q(y) \sum_{k=0}^{p^q-1} (\omega_q(y))_k \log(p^q (\omega_q(y))_k) d\mu(y) \\
&= H(\lambda_q) - q + \int_{\{\lambda_q > 0\}} \lambda_q(y) H(\omega_q(y)) d\mu(y) ,
\end{aligned} \tag{75}$$

and

$$\begin{aligned}
H(\lambda_q) &= - \int_{\mathcal{Y}} \lambda_q(y) \log \lambda_q(y) d\mu(y) \\
&= - \frac{1}{p} \sum_{k=0}^{p-1} \int_{\{\lambda_{q+1} > 0\}} \lambda_q(y|\xi_{p^{q+1}}^k) \log \lambda_q(y|\xi_{p^{q+1}}^k) d\mu(y) \\
&= - \int_{\{\lambda_{q+1} > 0\}} \frac{1}{p} \sum_{k=0}^{p-1} \lambda_q(y|\xi_{p^{q+1}}^k) \log \lambda_{q+1}(y) d\mu(y) - \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) \sum_{k=0}^{p-1} \frac{\lambda_q(y|\xi_{p^{q+1}}^k)}{p \lambda_{q+1}(y)} \log \frac{\lambda_q(y|\xi_{p^{q+1}}^k)}{\lambda_{q+1}(y)} d\mu(y) \\
&= - \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) \log \lambda_{q+1}(y) - \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) \sum_{k=0}^{p-1} (\nu_q(y))_k \log(p (\nu_q(y))_k) d\mu(y) \\
&= H(\lambda_{q+1}) - 1 + \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) H(\nu_q(y)) d\mu(y) .
\end{aligned} \tag{76}$$

Lemma 18 shows that (72) is equivalent to

$$q \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) H(\nu_q(y)) d\mu(y) \geq \int_{\{\lambda_q > 0\}} \lambda_q(y) H(\omega_q(y)) d\mu(y) , \quad \forall q = 1, \dots, r-1 . \tag{77}$$

We will prove (77) by estimating the two entropies appearing in the integrals.

Now fix an arbitrary $y \in \mathcal{Y}$ and an integer $1 \leq q < r$, and consider the set of likelihood values

$$W_q(y) := \left\{ W(y|\xi_{p^q}^0), W(y|\xi_{p^q}^1), \dots, W(y|\xi_{p^q}^{p^q-1}) \right\} = \left\{ W(y|\xi_{p^q}^0|1), W(y|\xi_{p^q}^{p^q-1}|1), \dots, W(y|\xi_{p^q}^1|1) \right\} .$$

Notice that

$$W_{q+1}(y) = \left\{ W(y|\xi_{p^{q+1}}^0|1), W(y|\xi_{p^{q+1}}^1|1), \dots, W(y|\xi_{p^{q+1}}^{p^{q+1}-1}|1) \right\} = \bigcup_{j=0}^{p-1} W_q(y|\xi_{p^{q+1}}^j) .$$

The geometry of the $K_{p^{q+1}}$ constellation implies that the ordering of the set $W_{q+1}(y)$ has a very particular structure.

Lemma 19 *For every $1 \leq q < r$ and $y \in \mathcal{Y}$, there is a partition*

$$W_{q+1}(y) = \bigcup_{k=1}^{p^q} W_q^k(y), \quad W_q^k(y) = \{w_{q,0}^k(y), w_{q,1}^k(y), \dots, w_{q,p-1}^k(y)\}$$

such that:

- $w_{q,j}^k(y) \in W_q(\xi_{p^q}^j y), \quad \forall k = 0, \dots, p^q - 1, \forall j = 0, \dots, p - 1;$
- $0 \leq k < k' < p^q \implies w_{q,i}^k(y) \geq w_{q,i}^{k'}(y), \quad \forall i, j = 0, \dots, p - 1.$ (78)

Proof:

By the definition of an AIDN channel, $W(y|x)$ is a decreasing function of the Euclidean distance $|y - x|$, the decreasing ordering of the set $W_{q+1}(y)$ coincides with the increasing ordering of the set of distances $\{|y - x|, x \in K_{p^{q+1}}\}$. Let

$$y = \rho e^{i\theta}, \quad \varphi_j = j \frac{2\pi}{p^{q+1}}, \quad j \in \mathbb{Z}_{p^{q+1}}.$$

Then

$$\begin{aligned} |y - \xi_{p^{q+1}}^j|^2 &= (\rho \cos \theta - \cos \varphi_j)^2 + (\rho \sin \theta - \sin \varphi_j)^2 \\ &= \rho^2 + 1 - 2\rho(\cos \theta \cos \varphi_j + \sin \theta \sin \varphi_j) \\ &= \rho^2 + 1 + 2\rho \cos(\theta - \varphi_j) \end{aligned}$$

Let $j^* \in \mathbb{Z}_{p^{q+1}}$ such that

$$|\theta - \varphi_{j^*}| \leq \theta - \varphi_j, \quad \forall j \in \mathbb{Z}_{p^{q+1}}$$

Then

$$\varphi_{j^*} \leq \theta \leq \varphi_{j^*} + \frac{1}{2} \frac{2\pi}{p^{q+1}} \quad (79)$$

or

$$\varphi_{j^*} - \frac{1}{2} \frac{2\pi}{p^{q+1}} \leq \theta \leq \varphi_{j^*}. \quad (80)$$

Suppose that (79) holds true. Then

$$\cos(\theta - \varphi_{j^*}) \geq \cos(\theta - \varphi_{j^*+1}) \geq \cos(\theta - \varphi_{j^*-1}) \geq \cos(\theta - \varphi_{j^*+2}) \geq \dots \geq \cos(\theta - \varphi_{j^* - \lfloor \frac{p^q}{2} \rfloor}). \quad (81)$$

From (81) it follows that, for odd p ,

$$\begin{aligned} W_q^0(y) &= \left\{ W(y|\xi_{p^{q+1}}^{j^*}), W(y|\xi_{p^{q+1}}^{j^*+1}), W(y|\xi_{p^{q+1}}^{j^*-1}), \dots, W(y|\xi_{p^{q+1}}^{j^* - \lfloor \frac{p^q}{2} \rfloor}) \right\} \\ W_q^1(y) &= \left\{ W(y|\xi_{p^{q+1}}^{j^* + \lceil \frac{p^q}{2} \rceil}), W(y|\xi_{p^{q+1}}^{j^* - \lceil \frac{p^q}{2} \rceil}), \dots, W(y|\xi_{p^{q+1}}^{j^* + p}) \right\} \\ &\vdots \\ W_q^{p^q-2}(y) &= \left\{ W(y|\xi_{p^{q+1}}^{j^* + \lfloor \frac{p^q}{2} \rfloor - p}), W(y|\xi_{p^{q+1}}^{j^* - \lfloor \frac{p^q}{2} \rfloor + p}), \dots, W(y|\xi_{p^{q+1}}^{j^* + \lfloor \frac{p^q}{2} \rfloor - \lfloor \frac{p^q}{2} \rfloor}) \right\} \\ W_q^{p^q-1}(y) &= \left\{ W(y|\xi_{p^{q+1}}^{j^* - \lfloor \frac{p^q}{2} \rfloor + \lfloor \frac{p^q}{2} \rfloor}), W(y|\xi_{p^{q+1}}^{j^* + \lfloor \frac{p^q}{2} \rfloor - \lceil \frac{p^q}{2} \rceil}), \dots, W(y|\xi_{p^{q+1}}^{j^* - \lfloor \frac{p^q}{2} \rfloor}) \right\}. \end{aligned} \quad (82)$$

But (82) implies the desired result since for every k the set of ξ 's exponents of the elements of $W_q^k(y)$ contains exactly one element from each equivalence class of integers modulo p .

The cases when (80) holds true instead of (79), and $p = 2$ are analogous. \blacksquare

Notice that for $j = 0, \dots, p-1$

$$W_q(y\xi_{p^{q+1}}^j) = \left\{ w_{q,j}^0(y) \geq w_{q,j}^1(y) \geq \dots \geq w_{q,j}^{p^q-1}(y) \right\} .$$

So, for every $y \in \mathcal{Y}$ such that $\lambda_q(y\xi_{p^{q+1}}^j) > 0$, if we define

$$\overline{\omega}_q(y, j) := \frac{1}{p^q \lambda_q(y\xi_{p^{q+1}}^j)} \left(w_{q,j}^0(y), w_{q,j}^1(y), \dots, w_{q,j}^{p^q-1}(y) \right) ,$$

we clearly have

$$H(\overline{\omega}_q(y, j)) = H(\omega_q(y\xi_{p^{q+1}}^j)) ,$$

since $\omega_q(y\xi_{p^{q+1}}^j)$ and $\overline{\omega}_q(y, j)$ simply differ for a permutation.

Consider now the p -adic expansion map

$$\theta : \{0, \dots, p^q - 1\} \rightarrow \{0, \dots, p-1\}^q ,$$

defined as follows: if $s \in \{0, \dots, p^q - 1\}$, we can write, in a unique way

$$s = \sum_{k=0}^{q-1} \rho_k p^k$$

for suitable elements $\rho_k \in \{0, \dots, p-1\}$. We then define

$$\theta(s) := (\rho_0, \dots, \rho_{q-1}) .$$

It is a standard fact that θ is a bijection. Now let $Z(y, j)$ be a random variable on $\{0, \dots, p^q - 1\}$ with distribution $\overline{\omega}_q(y, j)$ and let

$$Y(y, j) = (Y_1(y, j), \dots, Y_q(y, j)) := \theta \circ Z(y, j) .$$

For $\alpha = 1, \dots, q$, let $\delta_q^\alpha(y, j)$ be the distribution of $Y_\alpha(y, j)$ on $\{0, \dots, p-1\}$. A straightforward computation shows that

$$\delta_q^\alpha(y, j) = \frac{1}{p^q \lambda_q(y\xi_{p^{q+1}}^j)} \left(\sum_{h=0}^{p^\alpha-1} \sum_{\tilde{h}=0}^{p^{q-\alpha}-1} w_j^{\tilde{h}p^{\alpha+1} + sp^\alpha + h}(y) \Big|_{s=0, \dots, p-1} \right) . \quad (83)$$

Lemma 20 For every $1 \leq \alpha \leq q$,

$$H(\omega_q(y\xi_{p^{q+1}}^j)) \leq \sum_{\alpha=1}^q H(\delta_q^\alpha(y, j)) \quad (84)$$

Proof:

We have

$$\begin{aligned} H(\omega_q(y\xi_{p^{q+1}}^j)) &= H(\overline{\omega}_q(y, j)) = H(Z(y, j)) = H(\theta \circ Z(y, j)) = H(Y(y, j)) \\ &\leq \sum_{\alpha=1}^q H(Y_\alpha(y, j)) = \sum_{\alpha=1}^q H(\delta_q^\alpha(y, j)) , \end{aligned}$$

where we first used the fact that θ is a bijection, then apply chain rule for entropy, and finally the conditional entropy bound (see [7] for instance). \blacksquare

Next step of our proof consists in upperbounding the entropies $H(\delta_q^\alpha(x))$ with the entropy $H(\nu_q(x))$, for every $1 \leq \alpha \leq q < r$ and for every $j \in \{0, \dots, p-1\}$ and $y \in \mathcal{Y}$ such that $\lambda_q(y\xi_{p^{q+1}}^j) > 0$.

We start by stating a simple result characterizing the so called (generalized) 'permutahedron' of a given point in the n -dimensional Euclidean space. Let us introduce some notation. For any $K \subseteq \mathbb{R}^n$ the convex hull of K is defined as the smallest convex subset of \mathbb{R}^n containing K : it will be denoted by $\text{co}(K)$. The set $\text{co}(K)$ can be characterized as the intersection of all the convex sets K' such that $\mathbb{R}^n \supseteq K' \supset K$. A \mathcal{V} -polytope in \mathbb{R}^n is the convex hull of finite set $K \subset \mathbb{R}^n$. A \mathcal{H} -polytope in \mathbb{R}^n is a bounded intersection of closed halfspaces ($\{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n a_i x_i \leq a_0\}$, $a_i \in \mathbb{R}$ for $0 \leq i \leq n$). Notice that, since every halfspace is convex, then every \mathcal{H} -polytope is convex too; moreover it can be easily proved that every \mathcal{H} -polytope is the convex hull of its boundary. There is a general fundamental result (see [37] for instance) stating that an arbitrary set $P \subset \mathbb{R}^n$ is a \mathcal{V} -polytope if and only if it is an \mathcal{H} -polytope: we will therefore simply call it polytope.

In the following we will deal with a special class of polytopes: given a point $\mathbf{x} \in \mathbb{R}^n$, we shall consider $\text{co}(S_n \mathbf{x})$, i.e. the convex hull of the set of all component permutations of \mathbf{x} : this is sometimes called the (generalized) permutahedron of x . By the theorem we cited above, $\text{co}(S_n \mathbf{x})$ can be characterized as an \mathcal{H} -polytope, and next result explicitly gives such characterization.

Lemma 21 *Let $\mathbf{w} \in \mathbb{R}^n$ be such that*

$$w_1 \geq w_2 \geq \dots \geq w_n . \quad (85)$$

Then

$$\text{co}(S_n \mathbf{w}) = A$$

where

$$A := \bigcap_{J \subset \{1, \dots, n\}} \left\{ \sum_{i \in J} x_i \leq \sum_{i=1}^{|J|} w_i \right\} \cap \left\{ \sum_{i=1}^n x_i = \sum_{i=1}^n w_i \right\} \subset \mathbb{R}^n$$

Proof:

To prove $\text{co}(S_n \mathbf{w}) \subseteq A$ it suffices to note that, for every $\sigma \in S_n$, $\sigma \mathbf{x} \in A$: in fact, it is easy to check that, due to (85), every constraint is satisfied. Since A is convex and because of the definition of $\text{co}(S_n \mathbf{w})$ it immediately follows that $\text{co}(S_n \mathbf{w}) \subseteq A$.

We now prove the converse inclusion, $A \subseteq \text{co}(S_n \mathbf{w})$, by induction (we use the strong form of the induction principle). Clearly the statement is true for $n = 1$. Suppose that our claim is true for every $m \leq n$ for some given $n \in \mathbb{N}$. Let $\mathbf{w} \in \mathbb{R}^{n+1}$ such that $w_1 \geq w_2 \geq \dots \geq w_{n+1}$. For every $J \subset \{1, \dots, n+1\}$ consider the facet A_J of A defined by

$$A_J := \bigcap_{\substack{I \subset \{1, \dots, n+1\} \\ I \neq J}} \left\{ \sum_{i \in I} x_i \leq \sum_{i=1}^{|I|} w_i \right\} \cap \left\{ \sum_{i \in J} x_i = \sum_{i=1}^{|J|} w_i \right\} \cap \left\{ \sum_{i=1}^{n+1} x_i = \sum_{i=1}^{n+1} w_i \right\} .$$

We observe that

$$\pi_J A_J \subseteq B_J , \quad \pi_{J^c} A_J \subseteq C_J , \quad (86)$$

where π_J and π_{J^c} are the projections of \mathbb{R}^{n+1} onto the linear subspaces $\{x_i = 0, i \in J^c\}$ and $\{x_i = 0, i \in J\}$ respectively, and

$$B_J := \bigcap_{I \subset J} \left\{ \sum_{i \in I} x_i \leq \sum_{i=1}^{|I|} w_i \right\} \cap \left\{ \sum_{i \in J} x_i = \sum_{i=1}^{|J|} w_i \right\} \cap \bigcap_{i \in J^c} \{x_i = 0\}$$

$$C_J := \bigcap_{I \subset J^c} \left\{ \sum_{i \in I} x_i \leq \sum_{i=|J|+1}^{|J|+|I|} w_i \right\} \cap \left\{ \sum_{i \in J^c} x_i = \sum_{i=|J|+1}^{n+1} w_i \right\} \cap \bigcap_{i \in J} \{x_i = 0\}.$$

In fact, the former inclusion in (86) is trivial since B_J is defined as the intersection of a subset of the halfspaces whose intersection defines A_J , while for the latter it suffices to observe that, for each $I \subset J^c$,

$$\mathbf{x} \in A_J \Rightarrow \sum_{i \in I \cup J} x_i \leq \sum_{i=1}^{|I|+|J|} x_i, \quad \sum_{i \in J} x_i = \sum_{i=1}^{|J|} x_i \Rightarrow \sum_{i \in I} x_i = \sum_{i \in I \cup J} x_i - \sum_{i \in J} x_i \leq \sum_{i=|I|+1}^{|I|+|J|} x_i.$$

Now let $\theta_J \in S_{n+1}$ be any permutation such that

$$\theta_J(\{1, \dots, |J|\}) = J,$$

and let $S_J := \{\sigma \in S_{n+1} : \sigma|_{\{|J|+1, \dots, n+1\}} \equiv id\}$, $S_{J^c} := \{\sigma \in S_{n+1} : \sigma|_{\{1, \dots, |J|\}} \equiv id\}$. Notice that S_J commutes with S_{J^c} in the sense that $\sigma\rho = \rho\sigma$, for all $\sigma \in S_J$ and $\rho \in S_{J^c}$. We also define $\phi_J : \pi_J \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{|J|}$ and $\phi_{J^c} : \pi_{J^c} \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{|J^c|}$ as the standard isomorphisms. By applying the inductive hypothesis to $\Phi_J \pi_J \theta_J \mathbf{w}$ and $\Phi_{J^c} \pi_{J^c} \theta_J \mathbf{w}$ respectively, and then immersing back the results in \mathbb{R}^{n+1} by Φ_J^{-1} and $\Phi_{J^c}^{-1}$ respectively, you have that

$$B_J \subseteq \text{co}(\pi_J \theta_J S_J \mathbf{w}), \quad C_J \subseteq \text{co}(\pi_{J^c} \theta_J S_{J^c} \mathbf{w}). \quad (87)$$

For every $\mathbf{x} \in A_J$ we have $\pi_J \mathbf{x} \in B_J$ and $\pi_{J^c} \mathbf{x} \in C_J$ from (86) and then (87) implies that $\lambda' \in \mathcal{P}(S_J)$ and $\lambda'' \in \mathcal{P}(S_{J^c})$ exist such that

$$\begin{aligned} \mathbf{x} &= \pi_J \mathbf{x} + \pi_{J^c} \mathbf{x} \\ &= \sum_{\sigma \in S_J} \lambda'(\sigma) \pi_J \theta_J \sigma \mathbf{w} + \sum_{\rho \in S_{J^c}} \lambda''(\rho) \pi_{J^c} \theta_J \rho \mathbf{w} \\ &= \sum_{\substack{\sigma \in S_J \\ \rho \in S_{J^c}}} \lambda'(\sigma) \lambda''(\rho) \theta_J \sigma \rho \mathbf{w} \\ &= \sum_{\sigma \in \theta_J S_J S_{J^c}} \lambda(\sigma) \sigma \mathbf{w} \in \text{co}(S_{n+1} \mathbf{w}), \end{aligned}$$

with $\lambda \in \mathcal{P}(\theta_J S_J S_{J^c}) \subseteq \mathcal{P}(S_{n+1})$ defined by $\lambda(\theta_J \sigma \rho) := \lambda'(\sigma) \lambda''(\rho)$. So, for every $J \subset \{1, \dots, n+1\}$, we have proved that

$$A_J \subseteq \text{co}(S_{n+1} \mathbf{w}),$$

but then

$$A = \text{co}(\partial A) = \text{co} \left(\bigcup_{J \subset \{1, \dots, n+1\}} A_J \right) \subseteq \text{co}(S_{n+1} \mathbf{w}).$$

■

Lemma 22 Suppose n^2 real numbers $\{a_i^k, i, k = 0, \dots, n-1\}$ are given, such that

$$k < k' \implies a_j^k \leq a_i^{k'}, \quad j, l = 0, \dots, n-1. \quad (88)$$

Define the two vectors

$$\mathbf{x} = \left(\sum_{i=0}^{n-1} a_i^0, \sum_{i=0}^{n-1} a_i^1, \dots, \sum_{i=0}^{n-1} a_i^{n-1} \right), \quad \mathbf{v} = \left(\sum_{k=0}^{n-1} a_0^k, \sum_{k=0}^{n-1} a_1^k, \dots, \sum_{k=0}^{n-1} a_{n-1}^k \right).$$

Then $\mathbf{v} \in \text{co}(S_n \mathbf{x})$, i.e. \mathbf{v} is a convex combination of permutations of \mathbf{x} .

Proof: (88) implies that

$$x_0 \geq x_1 \geq \dots \geq x_{n-1}$$

and, for every $J \subset \{1, \dots, n-1\}$,

$$\sum_{i \in J} v_i \leq \sum_{i=0}^{|J|-1} x_i.$$

So Lemma 21 can be applied to show that $\mathbf{v} \in \text{co}(S_n \mathbf{x})$. ■

We can now prove the following inequality.

Lemma 23 For every $0 \leq \alpha < q < r$, and $y \in \mathcal{Y}$ such that $\lambda_{q+1}(y) > 0$,

$$H \left(\sum_{0 \leq j < p: \lambda_q(y \xi_{p^{q+1}}^j) > 0} \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} \delta_q^\alpha(y, j) \right) \leq H(\boldsymbol{\nu}_q(y)) \quad (89)$$

Proof: We will show that

$$\boldsymbol{\nu}_q(y) \in \text{co} \left(S_p \left(\sum_{0 \leq j < p: \lambda_q(y \xi_{p^{q+1}}^j) > 0} \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} \delta_q^\alpha(y, j) \right) \right). \quad (90)$$

Then (89) simply follows by the concavity of the entropy function.

From Lemma 19 and from (83) it follows that

$$\sum_{0 \leq j < p: \lambda_q(y \xi_{p^{q+1}}^j) > 0} \lambda_q(y \xi_{p^{q+1}}^j) \delta_q^\alpha(y, j) = \left(\sum_{j=0}^{p-1} \sum_{h=0}^{p^{\alpha-1} p^{q-\alpha-1}-1} \sum_{\tilde{h}=0}^{p^{\alpha-1} p^{q-\alpha-1}-1} w_j^{sp^\alpha+h+\tilde{h}p^{\alpha+1}}(y), s = 0, \dots, p-1 \right)$$

while, by definition,

$$p^{q+1} \lambda_{q+1}(y) \boldsymbol{\nu}_q(y) = \left(\sum_{i=0}^{p^q-1} w_{q,0}^i(y), \sum_{i=0}^{p^q-1} w_{q,1}^i(y), \dots, \sum_{i=0}^{p^q-1} w_{q,p-1}^i(y) \right).$$

If we define, for $0 \leq j, s \leq p-1$,

$$\alpha_j^s := \sum_{h=0}^{p^{\alpha-1} p^{q-\alpha-1}-1} \sum_{\tilde{h}=0}^{p^{\alpha-1} p^{q-\alpha-1}-1} w_{q,j}^{sp^\alpha+h+\tilde{h}p^{\alpha+1}}(y),$$

then

$$p^q \sum_{0 \leq j < p: \lambda_q(y \xi_{p^{q+1}}^j) > 0} \lambda_q(y \xi_{p^{q+1}}^j) \delta_q^\alpha(y, j) = \left(\sum_{j=0}^{p-1} a_j^0 \sum_{j=0}^{p-1} a_j^1, \dots, \sum_{j=0}^{p-1} a_j^{p-1} \right),$$

while

$$p^{q+1} \lambda_{q+1}(y) \nu_q(y) = \left(\sum_{s=0}^{p-1} a_s^0, \sum_{s=0}^{p-1} a_s^1, \dots, \sum_{s=0}^{p-1} a_s^{p-1} \right)$$

Fix a couple $(k, k') \in \{0, \dots, p-1\}$ such that $k < k'$: from (78) we have

$$w_{q,j}^{k p^\alpha + h + \tilde{h} p^{\alpha+1}}(y) \geq w_{q,i}^{k' p^\alpha + h + \tilde{h} p^{\alpha+1}}(y),$$

for every $j, i \in \{0, \dots, p-1\}$, $h \in \{0, \dots, p^{\alpha-1} - 1\}$, $\tilde{h} \in \{0, \dots, p^{q-\alpha-1} - 1\}$, and thus

$$\begin{aligned} a_j^k &= \sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_j^{k p^\alpha + h + \tilde{h} p^{\alpha+1}}(y) \\ &\leq \sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_i^{k' p^\alpha + h + \tilde{h} p^{\alpha+1}}(y) = a_i^{k'}. \end{aligned}$$

So the coefficients $\{a_j^k, j, k = 0, \dots, p-1\}$ satisfy (88) and then Lemma 22 can be applied to conclude that

$$p^{q+1} \lambda_{q+1}(y) \nu_q(y) \in \text{co} \left(S_p \left(p^q \sum_{0 \leq j < p: \lambda_q(y \xi_{p^{q+1}}^j) > 0} \lambda_q(y \xi_{p^{q+1}}^j) \delta_q^\alpha(y, j) \right) \right), \quad (91)$$

which in turn implies (90), since we have supposed $\lambda_{q+1}(y) > 0$. ■

Finally, we are ready to prove the following fundamental result.

Theorem 24 *For every $1 \leq q < r$,*

$$q C_{p^{q+1}} \leq (q+1) C_{p^q}. \quad (92)$$

Proof: We already noticed that (92) is equivalent to

$$q \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) H(\nu_q(y)) d\mu(y) \geq \int_{\{\lambda_q > 0\}} \lambda_q(y) H(\omega_q(y)) d\mu(y). \quad (93)$$

Fix an arbitrary $y \in \{\lambda_{q+1}(y) > 0\}$. Successively applying (84), the concavity of the entropy function H and (89), we obtain

$$\begin{aligned} \sum_{\substack{0 \leq j < p: \\ \lambda_q(y \xi_{p^{q+1}}^j) > 0}} \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} H(\omega_q(y \xi_{p^{q+1}}^j)) &\leq \sum_{\alpha=1}^q \left[\sum_j \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} H(\delta_q^\alpha(y, j)) \right] \\ &\leq \sum_{\alpha=1}^q H \left(\sum_j \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} \delta_q^\alpha(y, j) \right) \\ &\leq \sum_{\alpha=1}^q H(\nu_q(y)) \\ &= q H(\nu_q(y)). \end{aligned} \quad (94)$$

Thus

$$\frac{1}{p} \sum_{\substack{0 \leq j < p: \\ \lambda_q(y\xi_{p^{q+1}}^j) > 0}} \lambda_q(y\xi_{p^{q+1}}^j) H(\omega_q(y\xi_{p^{q+1}}^j)) \leq q\lambda_{q+1}(y) H(\nu_q(y)) \quad \forall y \in \{\lambda_{q+1} > 0\}, \quad (95)$$

which implies, since $\mathbb{1}_{\{\lambda_{q+1} > 0\}}(y) \geq \mathbb{1}_{\{\lambda_q > 0\}}(y\xi_{p^{q+1}}^j)$,

$$\begin{aligned} \int_{\{\lambda_q > 0\}} H(\omega_q(y)) d\mu(y) &= \int_{\mathcal{Y}} \lambda_q(y) H(\omega_q(y)) \mathbb{1}_{\{\lambda_q > 0\}}(y) d\mu(y) \\ &= \frac{1}{p} \sum_{j=0}^{p-1} \int_{\mathcal{Y}} \lambda_q(y\xi_{p^{q+1}}^j) H(\omega_q(y\xi_{p^{q+1}}^j)) \mathbb{1}_{\{\lambda_q > 0\}}(y\xi_{p^{q+1}}^j) d\mu(y) \\ &= \int_{\mathcal{Y}} \frac{1}{p} \sum_{\substack{0 \leq j < p: \\ \lambda_q(y\xi_{p^{q+1}}^j) > 0}} \lambda_q(y\xi_{p^{q+1}}^j) H(\omega_q(y\xi_{p^{q+1}}^j)) \mathbb{1}_{\{\lambda_{q+1} > 0\}}(y) d\mu(y) \\ &\leq q \int_{\mathcal{Y}} \lambda_{q+1}(y) H(\nu_q(y)) \mathbb{1}_{\{\lambda_{q+1} > 0\}}(y) d\mu(y) \\ &= q \int_{\{\lambda_{q+1} > 0\}} \lambda_{q+1}(y) H(\nu_q(y)) d\mu(y) . \end{aligned} \quad (96)$$

■

We summarize the results of the present section in the following:

Corollary 25 *For any prime p and positive integer r , every K_{p^r} -AIDN channel is such that*

$$\hat{C}_{\mathbb{Z}_{p^r}} = C_{p^r} . \quad (97)$$

Combining Corollary 25 with Corollary 15, we can finally state a result first conjectured by Loeliger in [25].

Corollary 26 *\mathbb{Z}_{p^r} -(-free) codes achieve capacity of the p^r -PSK AWGN channel.*

6 An example when $C_G < C$

In the previous section we have shown that for a wide class of \mathbb{Z}_{p^r} -symmetric channels with p^r -PSK as input \mathbb{Z}_{p^r} -capacity and Shannon capacity do coincide, thus implying by Corollary 15 that \mathbb{Z}_{p^r} -codes do suffice to achieve Shannon capacity of such channels. At this point the question arising is whether it is the case for any higher dimensional GU constellation admitting generating group isomorphic to \mathbb{Z}_{p^r} . The answer is negative as we will show in this section. In fact we will provide a whole family of counterexamples based on the three-dimensional constellations introduced in Example 5 of Section 2. We will prove that \mathbb{Z}_{2^r} -capacity of the AWGN channel with input constrained on some of these constellations is strictly less than the corresponding Shannon capacity, thus leading to an effective algebraic obstruction to the use of \mathbb{Z}_{2^r} -codes. This motivates the investigation of non Abelian group codes for such constellations.

We start by fixing some notation. Let r be an arbitrary positive integer to be considered fixed throughout this section. We consider the family of three-dimensional GU constellations $K_{2^r}^\beta$, parameterized by $\beta \in (0, +\infty)$ and defined as

$$K_{2^r}^\beta := \left\{ x_k = \left(\sqrt{\frac{1}{1+\beta^2}} e^{\frac{2\pi}{2^r} ki}, \sqrt{\frac{\beta^2}{1+\beta^2}} (-1)^k \right), \quad k = 0, 1, \dots, 2^r - 1 \right\} \subset \mathbb{C} \times \mathbb{R} \simeq \mathbb{R}^3.$$

We recall that the symmetry group of $K_{2^r}^\beta$ is isomorphic to the dihedral group D_{2^r} , and that $K_{2^r}^\beta$ admits two non isomorphic generating groups: the cyclic one \mathbb{Z}_{2^r} and the dihedral one D_{2^r-1} . Let us fix a standard deviation value $\sigma > 0$, and consider the corresponding family of $K_{2^r}^\beta$ -AWGN channels $(K_{2^r}^\beta, \mathbb{R}^3, W)$, with $W(y|x) = \frac{1}{(2\pi\sigma^2)^{3/2}} e^{-\frac{\|y-x\|^2}{2\sigma^2}}$. For $s = 1, \dots, r$ we will use the notation $C_{2^s}(\beta)$ for the capacity of the $K_{2^s}^\beta$ -AWGN channel, while $C_{\mathbb{Z}_{2^r}}(\beta)$ will be the \mathbb{Z}_{2^r} -capacity of the $K_{2^r}^\beta$ -AWGN channel, i.e.

$$C_{\mathbb{Z}_{2^r}}(\beta) = \min_{1 \leq s \leq r} \frac{r}{s} C_{2^s}(\beta).$$

We start our analysis by considering the limit case as β goes to 0. For $\beta = 0$, $K_{2^r}^\beta$ degenerates into an \mathbb{R}^3 embedding of the 2^r -PSK constellation, so that we can extend our definition of $K_{2^r}^\beta$ to the case $\beta = 0$ in a natural way:

$$K_{2^r}^0 := \left\{ x_k = \left(e^{\frac{2\pi}{2^r} ki}, 0 \right), \quad k = 0, 1, \dots, 2^r - 1 \right\} \subset \mathbb{C} \times \mathbb{R} \simeq \mathbb{R}^3.$$

Notice that clearly $K_{2^r}^0$ is not a 3-dimensional constellation since it does not span \mathbb{R}^3 . It is a trivial fact that, since orthogonal components of the additive Gaussian noise are mutually independent, for every $1 \leq s \leq r$ $C_{2^s}(0)$ coincides with the capacity of the K_{2^s} -AWGN channel, i.e. the 2-dimensional AWGN channel with input constrained over the 2^s -PSK constellation. Thus, all the results of last section hold true for the $K_{2^r}^0$ -AWGN channel: in particular we have \mathbb{Z}_{2^r} -capacity and Shannon one coinciding, i.e.

$$C_{\mathbb{Z}_{2^r}}(0) = C_{2^r}(0). \quad (98)$$

Similar arguments can be applied, for every given $\beta > 0$, to the 2^{r-1} -th subconstellation

$$\left\{ \left(\sqrt{\frac{1}{1+\beta^2}} e^{\frac{2\pi}{2^{r-1}} ki}, \sqrt{\frac{\beta^2}{1+\beta^2}} \right), \quad k = 0, 1, \dots, 2^{r-1} - 1 \right\}$$

which coincides with a 3-dimensional embedding of the constellation $\sqrt{\frac{1}{1+\beta^2}} K_{2^{r-1}}$, i.e. the 2^{r-1} -PSK rescaled by the homotopy $x \mapsto \sqrt{\frac{1}{1+\beta^2}} x$. This observation, combined with the equivalence of AWGN-channels with the same signal to noise ratio, and again the independence of orthogonal components of the Gaussian noise, allows us to apply the results of the previous section to state that

$$(r-1)C_{2^s}(\beta) \geq sC_{2^{r-1}}(\beta), \quad 1 \leq s \leq r-1. \quad (99)$$

Thus, for every given $\beta \in (0, +\infty)$, in order to check whether $C_{2^r}(\beta)$ and $C_{\mathbb{Z}_{2^r}}(\beta)$ do coincide or not we are only left to compare the two capacities $C_{2^r}(\beta)$ and $C_{2^{r-1}}(\beta)$, i.e.

$$C_{\mathbb{Z}_{2^r}}(\beta) = C_{2^r}(\beta) \iff (r-1)C_{2^r}(\beta) \leq rC_{2^{r-1}}(\beta).$$

If we now let the parameter β go to $+\infty$, the constellation $K_{2^r}^\beta$ approaches an \mathbb{R}^3 embedding of the 2-PAM modulation, with the 2^{r-1} even labeled points $\{x_{2k}, k = 0, \dots, x_{2^{r-1}-1}\}$ collapsed into the point $(0, 0, 1)$, and the odd labeled ones $\{x_{2k+1}, \dots, 2^{r-1}-1\}$ into the point $(0, 0, -1)$. Let us define this limit constellation as

$$K^\infty := \{(0, 0, 1), (0, 0, -1)\} .$$

We denote Shannon capacity of the K^∞ -AWGN channel by $C(\infty)$ and notice that, for every finite standard deviation value σ , we have

$$C(\infty) > 0 ,$$

while every subchannel of K^∞ trivially has zero Shannon capacity. We now want to evaluate the limit of both capacities $C_{2^r}(\beta)$ and $C_{2^{r-1}}(\beta)$ as β goes to infinity. Intuitively, as $K_{2^r}^\beta$ is approaching $K_{2^r}^\infty$, we can expect that respectively $C_{2^r}(\beta) \xrightarrow{\beta \rightarrow \infty} C(\infty)$ and $C_{2^s}(\beta) \xrightarrow{\beta \rightarrow \infty} 0$ for every $s < r$. In fact this is true as can be formally proved in the following way. We start by noticing that

$$\begin{aligned} \sum_{x \in K_{2^r}^\beta} \frac{1}{2^r} W(y|x) \log \left(\frac{W(y|x)}{\frac{1}{2^r} \sum_{z \in K_{2^r}^\beta} W(y|z)} \right) &\leq \sum_{x \in K_{2^r}^\beta} \frac{1}{2^r} \sum_{z \in K_{2^r}^\beta} \frac{1}{2^r} W(y|x) \log \left(\frac{W(y|x)}{W(y|z)} \right) \\ &= \frac{1}{2^{2r}} \sum_{x, z \in K_{2^r}^\beta} W(y|x) \log e \left(-\frac{\|y-x\|^2}{2\sigma^2} + \frac{\|y-z\|^2}{2\sigma^2} \right) \\ &\leq \frac{1}{2^{2r}} \sum_{x, z \in K_{2^r}^\beta} W(y|x) \frac{\log e}{2\sigma^2} (-\|y-x\|^2 + (\|y-x\| + \|z-x\|)^2) \\ &\leq \frac{1}{2^{2r}} \sum_{x, z \in K_{2^r}^\beta} W(y|x) \frac{\log e}{2\sigma^2} (\|y-x\|^2 + 2\|x-z\|^2) \\ &\leq \frac{1}{2^r} \sum_{x \in K_{2^r}^\beta} W(y|x) \frac{\log e}{2\sigma^2} (\|y-x\|^2 + 8) \end{aligned}$$

where the first inequality is due to the convexity of the function $x \rightarrow \log \frac{1}{x}$, the second one to the triangular inequality, the third one comes from the fact that $2ab \leq a^2 + b^2$ for every $a, b \in \mathbb{R}$, and the last one from the fact that x and z both lie on a sphere of radius 1, so that $\|x-z\| \leq \|x\| + \|z\| \leq 2$. Since

$$\frac{1}{2^r} \sum_{x \in K_{2^r}^\beta} \int_{\mathbb{C}^*} W(y|x) \frac{\log e}{2\sigma^2} (\|y-x\|^2 + 8) d\mu(y) = \log e \left(\frac{1}{2} + \frac{4}{\sigma^2} \right) < +\infty$$

we can apply Lebesgue's dominated convergence theorem (see [30]) in order to exchange the limit and the integral signs in evaluating the expressions $\lim_{\beta \rightarrow +\infty} C_{2^s}(\beta)$ for any $s \leq r$. By this argument and the continuity of transition densities $W(y|x) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|y-x\|^2}{2\sigma^2}}$, we get

$$\begin{aligned} \lim_{\beta \rightarrow +\infty} C_{2^r}(\beta) &= \lim_{\beta \rightarrow +\infty} \sum_{x \in K_{2^r}^\beta} \frac{1}{2^r} \int_{\mathcal{Y}} W(y|x) \log \left(\frac{W(y|x)}{\frac{1}{2^r} \sum_{z \in K_{2^r}^\beta} W(y|z)} \right) d\mu(y) \\ &= \int_{\mathcal{Y}} \frac{1}{2^r} \lim_{\beta \rightarrow +\infty} \sum_{x \in K_{2^r}^\beta} W(y|x) \log \left(\frac{W(y|x)}{\frac{1}{2^r} \sum_{z \in K_{2^r}^\beta} W(y|z)} \right) d\mu(y) \quad (100) \\ &= \int_{\mathcal{Y}} \frac{1}{2} \sum_{x \in K^\infty} W(y|x) \log \left(\frac{W(y|x)}{\frac{1}{2} \sum_{z \in K^\infty} W(y|z)} \right) d\mu(y) = C(\infty) \end{aligned}$$

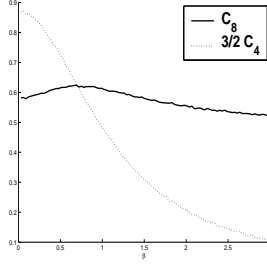


Figure 4: Shannon capacity and \mathbb{Z}_8 -capacity of K_8^β as a function of β

and, for every $1 \leq s < r$

$$\begin{aligned}
\lim_{\beta \rightarrow +\infty} C_{2^s}(\beta) &= \lim_{\beta \rightarrow +\infty} \int_{\mathcal{Y}} \frac{1}{2^s} \sum_{j=0}^{2^s-1} W(y|x_{2^{r-s}j}) \log \left(\frac{W(y|x_{2^{r-s}j})}{\frac{1}{2^s} \sum_{k=0}^{2^s-1} W(y|x_{2^{r-s}k})} \right) d\mu(y) \\
&= \int_{\mathcal{Y}} \frac{1}{2^s} \sum_{j=0}^{2^s-1} \lim_{\beta \rightarrow +\infty} W(y|x_{2^{r-s}j}) \log \left(\frac{W(y|x_{2^{r-s}j})}{\frac{1}{2^s} \sum_{k=0}^{2^s-1} W(y|x_{2^{r-s}k})} \right) d\mu(y) \quad (101) \\
&= \int_{\mathcal{Y}} W(y|(0,0,1)) \log \left(\frac{W(y|(0,0,1))}{W(y|(0,0,1))} \right) d\mu(y) = 0.
\end{aligned}$$

Thus a continuity argument applied to $C_{2^r}(\beta)$ and $C_{2^{r-1}}(\beta)$ implies the existence of $\bar{\beta} = \bar{\beta}(\sigma) \geq 0$ such that

$$C_{2^{r-1}}(\beta) < C_{2^r}(\beta), \quad \forall \beta > \bar{\beta}.$$

As a consequence we have that

$$C_{\mathbb{Z}_{2^r}}(\beta) = C_{2^r}(\beta) \iff \beta \leq \bar{\beta}. \quad (102)$$

As an immediate consequence of Theorem 5, (102) tell us that for every $\sigma > 0$ algebraic obstructions surely occur for $\beta > \bar{\beta}(\sigma)$. We can conclude that *for sufficiently high -but finite- values of β \mathbb{Z}_{2^r} -codes do not achieve Shannon capacity of the $K_{2^r}^\beta$ -AWGN channel of any arbitrary given signal to noise ratio.* We observe that it can be proved that, for $r > 2$,

$$(r-1)C_{2^r}(0) < rC_{2^{r-1}}(0).$$

A continuity argument implies then that $\bar{\beta} > 0$, i.e. *for sufficiently small -but positive- values of β , \mathbb{Z}_{2^r} -codes do achieve capacity of the $K_{2^r}^\beta$ -AWGN channel.*

Figure 4 reports the behaviour of $C_8(\beta)$ and $C_{\mathbb{Z}_8}(\beta)$ as a function of the parameter β (Montecarlo simulations).

Summarizing, in this section we have provided an example of Abelian G -symmetric channel –the $K_{2^r}^\beta$ -AWGN channel for $\beta > \bar{\beta}$ – for which G -codes are not sufficient to achieve Shannon capacity. It remains an open question whether or not for high values of β the capacity of the $K_{2^r}^\beta$ AWGN-channels can still be achieved by $D_{2^{r-1}}$ -codes, i.e. codes which are subgroups of $D_{2^{r-1}}^N$. Our feeling is that it could be possible: it seems to us that, roughly speaking, the structure of the dihedral group is more suitable to be adapted to $K_{2^r}^\beta$ when β

goes to infinity, since D_{2^r-1} contains a binary subgroup with corresponding subconstellation $\left\{ \left(\sqrt{\frac{1}{1+\beta^2}}, \sqrt{\frac{\beta^2}{1+\beta^2}} \right), \left(\sqrt{\frac{1}{1+\beta^2}} e^{\frac{2\pi}{2^r}i}, -\sqrt{\frac{\beta^2}{1+\beta^2}} \right) \right\}$ approaching $K^\infty = \{(0, 1)(0, -1)\}$ as β goes to infinity. More in general, one can ask which geometrically uniform constellations S admit eventually non-Abelian generating groups G such that Shannon capacity of the S -AWGN channels can be achieved by G -codes.

7 Conclusions

In this paper we developed a Shannon theory for group codes over symmetric memoryless channels, when the generating group G is an arbitrary finite Abelian group. Our results generalize the classical theory for binary linear codes over symmetric channels. The main example we have in mind is the AWGN channel with input restricted over a geometrically uniform constellation S admitting G as generating group and either soft or quantized output. We have individuated a new threshold value for the rates at which reliable transmission is possible with G -codes, which we called the G -capacity C_G , defined as the solution of an optimization problem involving Shannon capacities of the channels obtained by restricting the input to some of the subgroups of G . We have shown that at rates below C_G the average ML word error probability of the ensemble of G -codes goes to zero exponentially fast with the block length, with exponent at least equal to the G -channel coding exponent $E_G(R)$, while at rates beyond C_G the word error probability of any G -code is bounded from below by a strictly positive constant. We have proved that for the AWGN channel with m -PSK constellation as input (and m the power of a prime) the G -capacity C_G does coincide with the Shannon capacity C , so that in this case we have shown that reliable transmission at any rate $R < C$ can in fact be reached using group codes over \mathbb{Z}_m . Finally we have exhibited a counterexample when $C_G < C$: it consists of the AWGN channel with as input a particular three-dimensional constellation admitting \mathbb{Z}_m as generating group.

Among the still open problems we recall:

- giving a full proof that $E_G(R)$ is tight for the average G -code, and analyzing the error exponent of a typical G -code from the ensemble;
- studying new geometrically uniform constellations;
- extending the theory to non-Abelian groups.

Especially last point seems to us a great challenge for future research.

References

- [1] A. Barg, G. D. Forney, Jr., “Random Codes: Minimum Distances and Error Exponents”, *IEEE Trans. Inform. Theory*, vol. 48, pp. 2568-2573, 2001.
- [2] S. Benedetto, R. Garello, M. Mondin, G. Montorsi “Geometrically uniform partitions of $L \times$ MPSK constellations and related binary trellis codes”, *IEEE Trans. Inf. Theory*, vol. 39, pp.1773-1798, 1993.
- [3] S. Benedetto, R. Garello, M. Mondin, G. Montorsi “Geometrically uniform TCM codes based on $L \times$ MPSK constellations”, *IEEE Trans. Inf. Theory*, vol. 40, pp.137-152, 1994.
- [4] A. Bennatan, D. Burshetein, “On The Application of LDPC Codes to Arbitrary Discrete Memoryless Channels”, *IEEE Trans. Inf. Theory*, vol. 50, pp.417-438, Mar. 2004.

- [5] G. Caire, E. Biglieri, “Linear block codes over cyclic groups”, *IEEE Trans. Inf. Theory*, vol. 41, pp.1246-1256, 1995.
- [6] G. Como, F. Fagnani, “Ensembles of Codes over Abelian Groups”, in Proceedings of ISIT 2005 (Adelaide, SA, Australia), pp. 1788-1792, 5-9 Sept. 2005.
- [7] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.
- [8] U. Erez, G. Miller, “The ML Decoding Performance of LDPC Ensembles Over \mathbb{Z}_q ”, *IEEE Trans. Inform. Theory*, vol. 51, pp. 1871-1879, 2005.
- [9] F. Fagnani, R. Garello, B. Scanavino, S. Zampieri, “Geometrically Uniform Parallel Concatenated Coded Modulation Schemes – Part 1: Analysis”, submitted to *IEEE Trans. Inform. Theory*, 2004.
- [10] F. Fagnani, R. Garello, B. Scanavino, S. Zampieri, “Geometrically Uniform Parallel Concatenated Coded Modulation Schemes – Part 2: Design”, submitted to *IEEE Trans. Inform. Theory*, 2004.
- [11] F. Fagnani, F. Garin, “Analysis of Serial Concatenation Schemes for Non-binary Modulations”, in Proceedings of ISIT 2005 (Adelaide, SA, Australia), pp. 745–749, 5-9 Sept. 2005.
- [12] F. Fagnani, S. Zampieri, “Minimal Syndrome Formers for Group Codes”, *IEEE Trans. Inform. Theory*, vol. 45, pp. 1-31, 1998.
- [13] F. Fagnani, S. Zampieri, “System Theoretic Properties of Convolutional Codes Over Rings”, *IEEE Trans. Inform. Theory*, vol. 47, pp. 2256-2274, 2001.
- [14] F. Fagnani, S. Zampieri, “Minimal and systematic convolutional codes over finite Abelian groups”, *Linear Alg. its Applic.*, vol. 378, pp. 31-59, 2004.
- [15] G. D. Forney, Jr., “Geometrically Uniform Codes”, *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241-1260, 1991.
- [16] G. D. Forney, Jr., M.D. Trott, “The dynamics of group codes: state spaces, trellis diagrams and canonical encoders”, *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491-1513, 1993.
- [17] G. D. Forney, Jr., M.D. Trott, “The dynamics of group codes: Dual Abelian Group Codes and Systems”, *IEEE Trans. Inform. Theory*, vol. 50, pp. 2935-2965, 2004.
- [18] R. G. Gallager, *Low Density Parity Check Codes*, MIT Press, Cambridge MA, 1963.
- [19] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
- [20] R. G. Gallager, “The random coding bound is tight for the average code”, *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 244-246, 1973.
- [21] R. Garello, G. Montorsi, S. Benedetto, D. Divsalar, F. Pollara, “Labelings and encoders with the uniform bit error property with applications to serially concatenated trellis codes”, *IEEE Trans. Inform. Theory*, vol. 48, pp. 123-136, 2002.

- [22] T. W. Hungerford, *Algebra*, Springer Verlag, New York, 1974.
- [23] I. Ingemarsson, “Commutative Group Codes for the Gaussian Channel”, *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 215-219, 1973.
- [24] R. Johannesson, Z.-X. Wan, E. Wittenmark, “Some structural properties of convolutional codes over rings”, *IEEE Trans. Inform. Theory*, vol. 44, pp. 839-845, 1998.
- [25] H.-A. Loeliger, “Signal Sets Matched To Groups”, *IEEE Trans. Inform. Theory*, vol. 37, n. 6, pp. 1675-1679, Nov. 1991.
- [26] H.-A. Loeliger, G. D. Forney, T. Mittelholzer, M. D. Trott, “Minimality and Observability of Group Systems”, *Linear Alg. its Applic.*, vol. 205-206, pp. 937-963, 1994.
- [27] H.-A. Loeliger, T. Mittelholzer, “Convolutional Codes Over Groups”, *IEEE Trans. Inform. Theory*, vol. 42, n. 6, pp. 1660-1686, 1996.
- [28] D.J.C. MacKay, “Good Error Correcting Codes Based On Very Sparse Matrices”, *IEEE Trans. Inf. Theory*, vol. 45, pp.399-431, Mar. 1999.
- [29] G. Miller, D. Burshetein , “Bounds on the Maximum Likelihood Decoding Error Probability of Low-Density Parity-Check Codes”, *IEEE Trans. Inform. Theory*, vol. 47, pp.2696-2710, Nov. 2001.
- [30] W. Rudin, “Real and Complex Analysis”, McGraw-Hill, New York, 1966.
- [31] N. Shulman, M. Feder, “Random Coding Techniques for Nonrandom Codes”, *IEEE Trans. Inform. Theory*, vol. 45, NO.6, pp. 2001-2004, 1999.
- [32] D. Slepian, “Group Codes for the Gaussian Channel”, *Bell System Technical Journal*, vol. 47, pp. 575-602, April 1968.
- [33] D. Slepian, “On Neighbor Distances and Symmetry in Group Codes”, *IEEE Trans. Inform. Theory*, vol. 17, pp. 630-632, September 1971.
- [34] D. Sridhara, T.E. Fuja, “LDPC Codes Over Rings for PSK Modulation”, *IEEE Trans. Inform. Theory*, vol. 51, NO.9, pp. 3209-3220, 2005.
- [35] G. Ungerboeck, “Channel Coding with Multilevel/Phase Signals”, *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55-67, 1982.
- [36] A. J. Viterbi, J. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, New York, 1979.
- [37] M. Ziegler, *Lecture notes on polytopes*, Springer, New York, 1995.