The capacity of finite Abelian group codes over symmetric memoryless channels

Citation: Como, G., and F. Fagnani. "The Capacity of Finite Abelian Group Codes Over Symmetric Memoryless Channels." Information Theory, IEEE Transactions on 55.5 (2009): 2037-2054. © 2009 Institute of Electrical and Electronics Engineers

As Published: http://dx.doi.org/10.1109/TIT.2009.2015992

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: http://hdl.handle.net/1721.1/51996

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

The Capacity of Finite Abelian Group Codes Over Symmetric Memoryless Channels

Giacomo Como and Fabio Fagnani

Abstract—**The capacity of finite Abelian group codes over symmetric memoryless channels is determined. For certain important examples, such as m-PSK constellations over additive white Gaussian noise (AWGN) channels, with m a prime power, it is shown that this capacity coincides with the Shannon capacity; i.e., there is no loss in capacity using group codes. (This had previously been known for binary-linear codes used over binary-input output-symmetric memoryless channels.) On the other hand, a counterexample involving a three-dimensional geometrically uniform constellation is presented in which the use of Abelian group codes leads to a loss in capacity. The error exponent of the average group code is determined, and it is shown to be bounded away from the random-coding error exponent, at low rates, for finite Abelian groups not admitting Galois field structure.**

Index Terms—**Capacity, channel coding theorem, error exponent, geometrically uniform constellation, group codes, m-PSK, nonbinary codes.**

I. Introduction

It is a well-known fact that binary-linear codes suffice to reach capacity on binary-input output-symmetric channels [1]–[3]. Moreover, by averaging over the ensemble of linear codes, the same error exponent is achieved as by averaging over the ensemble of all codes. The same has been proven to hold true [4] for group codes over finite Abelian groups admitting Galois field structure.

In this paper, we investigate the same question for group codes employed over nonbinary channels exhibiting symmetries with respect to the action of a finite Abelian group G. The main example we have in mind is the additive white Gaussian noise (AWGN) channel with input set restricted to a finite geometrically uniform (GU) constellation [5] (m-PSK, for instance) and with possibly hard- or soft-decision decoding rule. In [6], it was conjectured that group codes should suffice in this case to achieve capacity exactly as in the field case. On the other hand, in [4] it was conjectured that group codes do not achieve the random-coding exponent if the group G does not admit Galois field structure. To our knowledge, there has not been any progress towards any of these directions.

However, interest in group codes has not decreased in these years. Indeed, they provide the possibility to use more spectrally efficient signal constellations while keeping many good qualities of binary-linear codes. More specifically, on the one hand, group codes have congruent Voronoi region, invariant distance profiles, and enjoy the uniform error property. On the other hand, the nice structure of the corresponding minimal encoders, syndrome formers, and trellis representations makes group codes appealing for low-memory encoding and low-complexity iterative decoding schemes. We refer to [7]–[20] and references therein for an overview of the many research lines on group codes which have been developing during recent years. Observe that coset codes over finite fields allow to achieve capacity and the random-coding error exponent of any memoryless channel [2]. However, whenever the group structure does not match the symmetry of the channel (e.g., binary coset codes on 2^r-PSK AWGN channels, for $r > 2$), or if the channel is not symmetric, coset codes in general fail to be GU, do not enjoy the uniform error property, and have noninvariant distance profiles.

Recently, group codes have made their appearance also in the context of turbo concatenated schemes [21], [22] and of low-density parity-check (LDPC) codes [23]–[26]. In the binary case, an important issue, for these types of high-performance coding schemes, is the evaluation of the gap to Shannon capacity, as well as the rate of convergence to zero of the word- and bit-error rate. For regular LDPC codes such gaps have been evaluated quite precisely [23]–[28] and it has been shown that, when the density parameters are allowed to increase, these schemes tend to attain the performance of generic binary-linear codes. In [24], [25], the authors extend such an analysis to LDPC codes over the cyclic group $\mathbb{Z}_q$, but they have to restrict themselves to the case of prime q. We believe that, without first a complete understanding of our original question, namely, if group codes do themselves allow to reach capacity and the correct error exponent, LDPC codes over general Abelian groups cannot be properly analyzed, since it cannot be understood whether the gap to capacity is due to the group structure or to the sparseness of the syndrome representation. In [29], a fundamental analysis of LDPC codes over Abelian groups is proposed, based on the general results for group codes presented in this paper.

Our work focuses on the case of finite Abelian groups and is organized as follows. In Section II, we introduce all relevant notation, we briefly resume the Shannon–Gallager theory concerning the capacity and error exponents of memoryless channels and basic concepts concerning GU constellations, and

Manuscript received December 07, 2005; revised June 16, 2008. Current version published April 22, 2009. The material in this paper was been presented in part at the IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia, September 2005..

G. Como was with the Dipartimento di Matematica, Politecnico di Torino, Torino, Italy. He is now with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: giacomo@mit.edu).

F. Fagnani is with the Dipartimento di Matematica, Politecnico di Torino, 10126 Torino, Italy (e-mail: fabio.fagnani@polito.it).

Communicated by L. M. G. M. Tolhuizen, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2009.2015992

0018-9448/$25.00 © 2009 IEEE

we formally state the main question whether group codes can achieve capacity of a symmetric channel.

In Section III, we consider memoryless channels which are symmetric with respect to the action of cyclic groups $\mathbb{Z}_{p^r}$ of prime power order, and we determine (in a computationally effective way) the capacity achievable by group codes over such channels. This capacity is called the $\mathbb{Z}_{p^r}$-capacity and equals the minimum of the normalized Shannon capacities of the channels obtained by restricting the input to all nontrivial subgroups of $\mathbb{Z}_{p^r}$. The results are contained in Theorem 5 which is an inverse coding theorem for group codes and in Theorem 7 which exhibits an average result working on the ensemble of group codes. The error exponent for the average group code is determined as well. It is shown that for $r > 1$, the average $\mathbb{Z}_{p^r}$-code is bounded away from the random-coding exponent at least at low rates, confirming a conjecture of Dobrushin [4].

In Section IV, we show that for the p^r-PSK AWGN channel, the $\mathbb{Z}_{p^r}$-symmetric capacity and the classical Shannon capacity do coincide so that group codes allow to achieve capacity in this case. This proves a conjecture of Loeliger [6].

In Section V, we present a counterexample based on a three-dimensional GU constellation where, in contrast, the two capacities are shown to differ from each other. It remains an open problem whether using non-Abelian generating groups the Shannon capacity can be achieved in this case.

Finally, in Section VI, we generalize the theory to channels symmetric with respect to the action of arbitrary finite Abelian generating groups.

II. Problem Statement

In this section, all relevant notation and definition are introduced, and a formal statement of the problem is presented.

A. Notation

Throughout the paper, the functions $\exp : \mathbb{R} \to \mathbb{R}$ and $\log : (0, +\infty) \to \mathbb{R}$ have to be considered with respect to the same, arbitrary chosen, base $a \in (1, +\infty)$, unless explicit mention to the contrary. For a subset $A \subseteq B$, $\mathbb{1}_A : B \to \{0, 1\}$ will denote the indicator function of A, defined by $\mathbb{1}_A(x) = 1$ if $x \in A$, $\mathbb{1}_A(x) = 0$ if $x \notin A$.

For two groups G and H, we shall write $G \simeq H$ to mean they are isomorphic, while $H \leq G$ will mean that H is a subgroup of G. Unless otherwise stated, we shall use the multiplicative notation for a generic group G, with 1_G denoting the null element. When restricted to Abelian case, we shall switch to the additive notation with 0 denoting the null element.

Given a finite set A, we shall consider the simplex $\mathcal{P}(A) := \{\boldsymbol{\theta} : A \to [0, +\infty) | \sum_a \boldsymbol{\theta}(a) = 1\}$ of probability measures on A. The discrete entropy function $H : \mathcal{P}(A) \to \mathbb{R}$ is defined by $H(\boldsymbol{\theta}) := -\sum_{\boldsymbol{\theta}(a)>0} \boldsymbol{\theta}(a) \log \boldsymbol{\theta}(a)$. Similarly, for a continuous space B, we shall denote $\mathcal{P}(B)$ the set of probability densities on B and define the entropy function $H : \mathcal{P}(B) \to [-\infty, +\infty]$ by $H(\mu) := -\int_B \mu(x) \log \mu(x) \mathrm{d}x$. Given $\boldsymbol{x} \in A^n$, its A-*type* (or empirical frequency) is the probability measure $\boldsymbol{\theta}_A(\boldsymbol{x}) \in \mathcal{P}(A)$ given by $\boldsymbol{\theta}_A(\boldsymbol{x}) := 1/n \sum_{1 \leq i \leq n} \mathbb{1}_{\{x_i\}}$. Define the set of types of all n-tuples by $\mathcal{P}_n(A) := \boldsymbol{\theta}_A(A^n)$, and let $\mathcal{P}_{\mathbb{N}}(A) := \cup_n \mathcal{P}_n(A)$ be the set of all A-types. The number of A-types $|\mathcal{P}_n(A)| = \binom{n+|A|-1}{|A|-1}$ is a quantity growing polynomially fast in n. In contrast, the set of n-tuples of a given type $\boldsymbol{\theta}$, denoted by

$$A^n_{\boldsymbol{\theta}} := \{\boldsymbol{x} \in A^n \text{ s.t. } \boldsymbol{\theta}_A(\boldsymbol{x}) = \boldsymbol{\theta}\}$$

has cardinality $|A^n_{\boldsymbol{\theta}}| = \binom{n}{n\boldsymbol{\theta}} := n!/\prod_a (n\boldsymbol{\theta}(a))!$ growing exponentially fast with n.

B. Coding Theory for Memoryless Channels

Throughout the present paper, memoryless channels (MCs) will be considered, which are described by a triple $(\mathcal{X}, \mathcal{Y}, W)$, where $\mathcal{X}$ is the input set, $\mathcal{Y}$ is the output set and, for every x in $\mathcal{X}$, $W(\cdot|x)$ is a probability density on $\mathcal{Y}$ describing the conditional distribution of the output given that the input x has been transmitted. The input set $\mathcal{X}$ will always be assumed finite, while the output set $\mathcal{Y}$ will often be identified with the n-dimensional Euclidean space $\mathbb{R}^n$. Nevertheless, all the results presented in this paper continue to hold when $\mathcal{Y}$ is a discrete space as well; in this case, it is simply needed to replace Lebesgue integrals with sums over $\mathcal{Y}$.[1]

We shall consider the nth extension of an MC $(\mathcal{X}, \mathcal{Y}, W)$, having input set $\mathcal{X}^n$ and output set $\mathcal{Y}^n$ and transition probability densities $W_n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{j=1}^n W(y_j|x_j)$. This motivates the name memoryless, the various transmissions being probabilistically independent once the input signals have been fixed.

As usual, a block code is any subset $\mathcal{C} \subseteq \mathcal{X}^n$, while a decoder is any measurable mapping $\mathcal{D} : \mathcal{Y}^n \to \mathcal{C}$. A coding scheme consists of a pair of a code and a decoder. n is the block length, while $R = \log |\mathcal{C}|/n$ will denote the transmission rate.

The probabilistic model of transmission is obtained by assuming that the transmitted codeword is a random variable (r.v.) $\boldsymbol{X}$ uniformly distributed over $\mathcal{C}$, and that the channel-output r.v. $\boldsymbol{Y}$ has conditional probability density $W_n(\cdot|\boldsymbol{X})$ given $\boldsymbol{X}$. An error occurs when the output $\boldsymbol{Y}$ is incorrectly decoded, i.e., it is the event $\{\mathcal{D}(\boldsymbol{Y}) \neq \boldsymbol{X}\}$. The error probability of the coding scheme $(\mathcal{C}, \mathcal{D})$ is therefore given by

$$p_e(\mathcal{C}, \mathcal{D}) := \frac{1}{|\mathcal{C}|} \sum_{\boldsymbol{x} \in \mathcal{C}} p_e(\mathcal{C}, \mathcal{D}|\boldsymbol{x})$$

where $p_e(\mathcal{C}, \mathcal{D}|\boldsymbol{x}) := \int_{\mathcal{Y}^n} \mathbb{1}_{\mathcal{C}\setminus\{\boldsymbol{x}\}}(\mathcal{D}(\boldsymbol{y})) W_n(\boldsymbol{y}|\boldsymbol{x}) \mathrm{d}\boldsymbol{y}$ is the error probability conditioned on the transmission of the codeword $\boldsymbol{x}$.

It is well known that, given a code $\mathcal{C}$, the decoder minimizing the error probability is the maximum-likelihood (ML) one $\mathcal{D}_{\mathrm{ML}}(\boldsymbol{y}) := \arg\max_{\boldsymbol{x} \in \mathcal{C}} W_n(\boldsymbol{y}|\, \boldsymbol{x})$, solving cases of nonuniqueness by assigning to $\mathcal{D}_{\mathrm{ML}}(\boldsymbol{y})$ a value $\boldsymbol{x} \in \mathcal{C}$ arbitrarily chosen from the set of maxima of $W_n(\boldsymbol{y}|\boldsymbol{x})$. Throughout the paper, we shall always assume that ML decoding is used, and use the notation $p_e(\mathcal{C})$ and $p_e(\mathcal{C}|\boldsymbol{x})$ for $p_e(\mathcal{C}, \mathcal{D}_{ML})$ and $p_e(\mathcal{C}, \mathcal{D}_{ML}|\boldsymbol{x})$, respectively.

[1] In fact, all the results hold true when $\mathcal{Y}$ is a Borel space [32] and integrations are carried on with respect to an abstract σ-finite reference measure, with respect to which all conditioned output measures are absolutely continuous.

In order to state the classical channel-coding theorem we are only left with defining the capacity and the random-coding exponent of an MC $(\mathcal{X}, \mathcal{Y}, W)$. The former is defined as

$$C := \max_{p \in \mathcal{P}(\mathcal{X})} \sum_{x \in \mathcal{X}} p(x) \int_{\mathcal{Y}} W(y|x) \log \frac{W(y|x)}{\sum\limits_{z \in \mathcal{X}} p(z) W(y|z)} \mathrm{d}y. \tag{1}$$

The latter is given, for $R \in [0, \log|\mathcal{X}|]$, by

$$E(R) := \max_{0 \le \rho \le 1} \max_{p \in \mathcal{P}(\mathcal{X})} (E_0(\rho, p) - \rho R) \tag{2}$$

where, for every $\rho \in [0,1]$ and $p \in \mathcal{P}(\mathcal{X})$

$$E_0(\rho, p) := -\log \left(\int_{\mathcal{Y}} \left(\sum_{x \in \mathcal{X}} p(x) W(y|x)^{1/1+\rho} \right)^{1+\rho} \mathrm{d}y \right). \tag{3}$$

A well-known fact (see [2], [3]) is that

$$E(R) > 0 \Leftrightarrow R < C. \tag{4}$$

Moreover, the random-coding exponent $E(R)$ is continuous, monotonically decreasing, and convex in the interval $[0, C)$, while the dependence of both C and $E(R)$ on the channel is continuous.

Given a design rate $R \in [0, \log|\mathcal{X}|]$ and block length n, the random-coding ensemble is obtained by considering a random collection $\mathcal{C}_n$ of $\lceil \exp(Rn) \rceil$ possibly nondistinct n-tuples, sampled independently from $\mathcal{X}^n$, each with distribution $\bigotimes_{1 \le j \le n} \boldsymbol{\mu}^*$, where $\boldsymbol{\mu}^*$ in $\mathcal{P}(\mathcal{X})$ is the optimal input distribution in (2). $\overline{p_e(\mathcal{C}_n)}^R$ will denote the average error probability with respect to such a probability distribution.

We can now state the Shannon–Gallager coding theorem for MCs.

Theorem: Assume a MC $(\mathcal{X}, \mathcal{Y}, W)$ is given, having capacity C and random-coding exponent $E(R)$. Then the following holds.

(a)
$$\overline{p_e(\mathcal{C}_n)}^R \le \exp(-nE(R)).$$

In particular, this implies that the average error probability tends to 0 exponentially fast for $n \longrightarrow +\infty$, provided that the rate of the codes 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 $p_e(\mathcal{C}) \ge A_R$.

C. Symmetric Memoryless Channels and Group Codes

In this paper, we shall focus on MCs exhibiting symmetries, and on codes matching such symmetries.

In order to formalize the notion of symmetry, a few concepts about group actions need to be recalled. Given a finite group G, with identity 1_G, and a set A, we say that G acts on A if, for every $g \in G$, there 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 $ga = b$. 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 $a_0 \in A$. Finally, the action of a group G on a measure space A is said to be isometric if it consists of measure-preserving bijections. In particular, when A is a finite set, all group actions are isometric. In contrast, when $A = \mathbb{R}^n$, this becomes a real restriction and is satisfied if the maps $a \mapsto ga$ are isometries of $\mathbb{R}^n$, i.e., maps preserving the Euclidean distance.

Definition 1: Let G be a group. An 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 $g \in G, x \in \mathcal{X}, y \in \mathcal{Y}$.

The simplest example of a G-symmetric MC is the following one, while a much richer family of symmetric MCs based on GU signal constellations will be presented in Section II-D.

Example 1: (*m-ary symmetric channel*). Consider a finite set $\mathcal{X}$ of cardinality $m \ge 2$ and some $\varepsilon \in [0,1]$. The *m-ary symmetric channel* is described by the triple $(\mathcal{X}, \mathcal{X}, W)$, where $W(y|x) = 1 - \varepsilon$ if $y = x$ and $P(y|x) = \varepsilon/(m-1)$ otherwise. This channel returns the transmitted input symbol x as output with probability $1 - \varepsilon$, while with probability ε a wrong symbol is received, uniformly distributed over the set $\mathcal{X} \setminus \{x\}$. The special case $m = 2$ corresponds to the binary-symmetric channel (BSC). The m-ary symmetric channel exhibits the highest possible level of symmetry. Indeed, it is G-symmetric for every group G of order $|G| = m$. To see this, it is sufficient to observe that every group acts simply and transitively on itself. Notice that whenever $m = p^r$ for some prime p and positive integer r, the group G can be chosen to be $\mathbb{Z}_p^r$ which is compatible with the structure of the Galois field $\mathbb{F}_{p^r}$.

A first 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 (1) and (2) can be chosen to be the uniform distribution over the input set $\mathcal{X}$.

Since the input of a G-symmetric MC can be identified with the group G itself, block codes for such channels are subsets $\mathcal{C} \subseteq G^n$. However, it is natural to consider a subclass of codes matching the symmetry of the channel: they are known as group codes.

Definition 2: For a finite group G, a *G-code* is a subgroup $\mathcal{C} \le G^n$.

G-codes enjoy many properties when used over G-symmetric MCs. In particular, [5] they have congruent Voronoi (ML decoding) regions, and invariant distance profiles. As a consequence, the uniform error property (UEP) holds true, namely, the error probability does not depend on the transmitted codeword: $p_e(\mathcal{C}|\boldsymbol{x}) = p_e(\mathcal{C}|\boldsymbol{x}')$ for every $\boldsymbol{x}, \boldsymbol{x}'$ in $\mathcal{C}$.

Another important property is that their ML error probability of a G-code can be bounded by a function of its type-spectrum only. For a code $\mathcal{C} \subseteq G^n$ and a type $\boldsymbol{\theta}$ in $\mathcal{P}(G)$, let $S_{\mathcal{C}}(\boldsymbol{\theta}) := |\mathcal{C} \cap G^n_{\boldsymbol{\theta}}|$ be the number of codewords $\boldsymbol{x}$ of $\mathcal{C}$ of type $\boldsymbol{\theta}$. The following estimate is proved using techniques similar to those in [33]. It will be used in Section III-B while proving the direct coding theorem for $\mathbb{Z}_{p^r}$-codes.

Lemma 3: Let G be a finite group, $(G, \mathcal{Y}, W)$ a G-symmetric MC, and $\mathcal{C} \subseteq G^n$ a code such that $1_{G^n} \in \mathcal{C}$. Then

$$p_e(\mathcal{C}|1_{G^n}) \leq \frac{1}{|G|^n} \sum_{\boldsymbol{z} \in G^n} \int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}) \left(\sum_{\substack{\boldsymbol{\theta} \in \mathcal{P}_n(G) \\ \boldsymbol{\theta} \neq \delta_{1_G}}} \frac{S_{\mathcal{C}}(\boldsymbol{\theta})}{\binom{n}{n\boldsymbol{\theta}}} \sum_{\boldsymbol{x} \in G^n_{\boldsymbol{\theta}}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}\boldsymbol{x}) \right)^{\rho} \mathrm{d}\boldsymbol{y}. \quad (5)$$

Proof: See Appendix A. □

Observe that Lemma 3 does not assume $\mathcal{C}$ to be a G-code. However, when $\mathcal{C}$ is a G-code, (5) provides an estimate to $p_e(\mathcal{C})$ by the UEP.

A fundamental question arising is whether G-codes allow to achieve the capacity of a G-symmetric MC. This is known to be the case for binary-linear codes over binary-input output-symmetric channels. Moreover, as shown in [4], the same continues to hold true whenever the group G has the property that every element g in G has the same order, i.e., when G is isomorphic to $\mathbb{Z}_p^r$ for some prime p and positive integer r. However, in [6] Loeliger conjectured that $\mathbb{Z}_m$-codes should suffice to achieve capacity on the m-PSK AWGN channel even for non-prime m. In the present paper, Loeliger's conjecture will be proved to be true for m equal to a prime power. More generally, the capacity achievable by G-codes over G-symmetric channels will be characterized for any finite Abelian group G, and a counterexample will be presented showing that, when G is not isomorphic to $\mathbb{Z}_p^r$, G-codes may fail to achieve Shannon capacity.

D. Geometrically Uniform Signal Constellations

A finite d-dimensional *constellation* is a finite subset $S \subset \mathbb{R}^d$ spanning $\mathbb{R}^d$; i.e., every $\boldsymbol{x} \in \mathbb{R}^d$ can be written as $\boldsymbol{x} = \sum_{s \in S} \alpha_s s$ with $\alpha_s \in \mathbb{R}$. We shall restrict ourselves to the study of finite constellations $S \subset \mathbb{R}^d$ with barycenter $\mathbf{0}$, i.e., such that $\sum_{s \in S} s = \mathbf{0}$: these minimize the average per-symbol energy over the class of 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 finite GU constellation admits a generating group (see [34] for a counterexample). However, in what follows we shall always assume that the constellations we are dealing with, do admit generating groups, and, actually, Abelian ones.

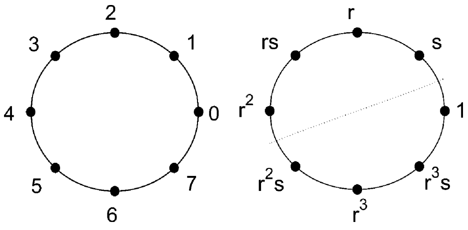

Fig. 1. 8-PSK constellation with the two labelings $\mathbb{Z}_8$ and D_4.

Let S be a finite d-dimensional GU constellation equipped with a generating group G. Define the *S-AWGN channel* as the d-dimensional unquantized AWGN channel with input set S, output $\mathbb{R}^d$, and Gaussian transition densities given by

$$W(y|x) = \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-||y-x||^2/2\sigma^2}.$$

The S-AWGN channel is G-symmetric.

A well-known fact (see [7]) is that every finite GU constellation S lies on a sphere. With no loss of generality, we shall assume that the radius of this sphere is 1.

The above construction of G-symmetric channels with a finite GU constellation S as input can be extended to a much wider class of channels. Indeed, one could consider the hard-decoded version of the S-AWGN channel, obtained by quantizing the output over the Voronoi regions of S through the map

$$Q : \mathbb{R}^d \to S \quad Q(x) = \underset{s \in S}{\operatorname{argmin}} ||x - s||.$$

Moreover, all the theory can be generalized to MCs having a GU finite constellation S as input and transition densities $W(y|x)$ which are functions of the Euclidean distance $||y - x||$ only. As an example, one can consider the Laplacian channel with transition probability densities given by

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

where $\lambda > 0$ is a parameter and $\Gamma(t) := \int_0^{+\infty} x^{t-1} e^{-x} \mathrm{d}x$ is the well-known Euler's Γ function.

In the following we present some examples of finite GU constellations admitting an Abelian generating group. We start with the simplest example, a binary constellation.

Example 2: (2-PAM). The 2-PAM constellation is defined by

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

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. □

We now consider m-PSK constellations, which are the main practical example of finite GU constellations.

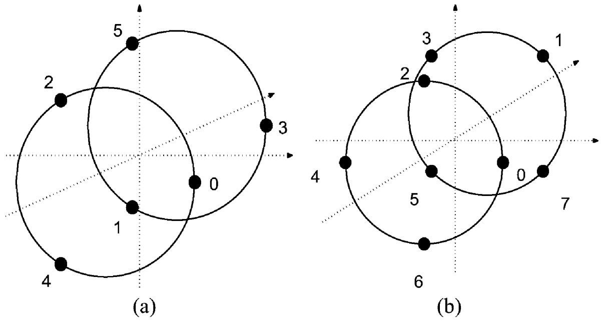

Fig. 2. (a) $\mathbb{Z}_6$-labeled 2-PAM ×3-PSK; (b) $\mathbb{Z}_8$-labeled K_8^β constellation.

Example 3: (*m-PSK*) For any integer $m \geq 2$, define $\xi_m := e^{i2\pi/m}$. The m-PSK constellation is

$$K_m := \left\{\xi_m^k, 1 \leq k \leq m\right\}.$$

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 $\mathbb{Z}_m$, the Abelian group of integers modulo m, as a generating group. When m is even there is another generating group (see [5], [6]): the dihedral group $D_{m/2}$, which is noncommutative for $m \geq 6$. It follows that the m-PSK AWGN channel is both $\mathbb{Z}_m$-symmetric and (for even m) $D_{m/2}$-symmetric. The constellation K_8 with the two possible labelings $\mathbb{Z}_8$ and D_4 is shown in Fig. 1. □

The next example shows how higher dimensional GU constellations can be obtained as Cartesian products of lower dimensional ones. This example will be considered in Section VI, to show how the G-capacity can be evaluated for Abelian group codes whose order is not a power of a prime.

Example 4: (*Cartesian product constellation*). For any integer $m > 2$ consider the family of three-dimensional GU constellations parameterized by $\beta \in (0, +\infty)$

$$K_{m\times 2}^\beta := \left\{\frac{1}{\sqrt{1+\beta^2}}\left(\xi_m^k, (-1)^l\beta\right) | 0 \leq k \leq 2,\, l = 0, 1\right\}.$$

Fig. 2(a) shows the special case $m = 3$. It is 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, AWGN channels with input m-PSK ×2-PAM are $\mathbb{Z}_{2m}$-symmetric. □

Finally, we provide an example of an "effectively" three-dimensional constellation, i.e., one which is not obtained as the Cartesian product of lower dimensional ones. This constellation will be used as a counterexample in Section V.

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

$$K_m^\beta = \left\{\left(\sqrt{\frac{1}{1+\beta^2}}\xi_m^k, \sqrt{\frac{\beta^2}{1+\beta^2}}(-1)^k\right), 1 \leq k \leq m\right\}.$$

An example with $m = 8$ is shown in Fig. 2(b): observe that even-labeled points and odd-labeled ones have an offset of $\pi/4$. 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, K_m^β-AWGN channels are both $\mathbb{Z}_m$-symmetric and $D_{m/2}$-symmetric. □

III. The Coding Theorem for $\mathbb{Z}_{p^r}$-Codes on $\mathbb{Z}_{p^r}$-Symmetric Memoryless Channels

Given a prime p and a positive integer r, let $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ be a $\mathbb{Z}_{p^r}$-symmetric MC, whose input has been identified with the group $\mathbb{Z}_{p^r}$ itself with no loss of generality. For $1 \leq l \leq r$, consider the MC $(p^{r-l}\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ obtained by restricting the input of the original MC to the subgroup $p^{r-l}\mathbb{Z}_{p^r}$. We shall denote by C_l the Shannon capacity of such a channel, and by $E_l(R)$ its error exponent.

Definition 4: The $\mathbb{Z}_{p^r}$*-capacity* of the MC $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ is

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

its $\mathbb{Z}_{p^r}$-error exponent is

$$E_{\mathbb{Z}_{p^r}}(R) := \min_{1\leq l\leq r} E_l\left(\frac{l}{r}R\right).$$

It is easily observed that $E_{\mathbb{Z}_{p^r}}(R) > 0$ if and only if $R < C_{\mathbb{Z}_{p^r}}$. In the remainder of this section, the quantity $C_{\mathbb{Z}_{p^r}}$ will be shown to be exactly the capacity achievable by $\mathbb{Z}_{p^r}$-codes over the $\mathbb{Z}_{p^r}$-symmetric MC $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$. In particular, in Section III-A it will be proven that reliable transmission with $\mathbb{Z}_{p^r}$-codes is not possible at any rate beyond $C_{\mathbb{Z}_{p^r}}$. In Section III-B, a random-coding argument will be used in order to show that $\mathbb{Z}_{p^r}$-codes of arbitrarily small error probability exist at any rate below $C_{\mathbb{Z}_{p^r}}$, and that $E_{\mathbb{Z}_{p^r}}(R)$ is a lower bound to the error exponent of the average $\mathbb{Z}_{p^r}$-code of rate R. Section III-C will deal with issues of tightness of $E_{\mathbb{Z}_{p^r}}(R)$.

A. The Converse Coding Theorem for $\mathbb{Z}_{p^r}$-Codes

Let $\mathcal{C} \leq \mathbb{Z}_{p^r}^n$ be some $\mathbb{Z}_{p^r}$-code of length n and rate R. Standard algebraic arguments (see [14], for instance) allow to show that

$$\mathcal{C} \simeq \bigoplus_{1\leq s\leq r} \mathbb{Z}_{p^s}^{K_s}$$

for some nonnegative integers K_s satisfying

$$\frac{1}{n}\sum_{1\leq s\leq r} sK_s \log p = R.$$

For every $1 \leq l \leq r$, we consider the code $\mathcal{C}_l := \mathcal{C} \cap p^{r-l}\mathbb{Z}_{p^r}^n$ obtained by restricting the original code $\mathcal{C}$ to the subgroup $p^{r-l}\mathbb{Z}^n_{p^l}$. This is tantamount to considering only those codewords of $\mathcal{C}$ of order not exceeding p^l. It follows that

$$\mathcal{C}_l \simeq \bigoplus_{1\le s<l} \mathbb{Z}^{K_s}_{p^s} \bigoplus_{l\le s\le r} \mathbb{Z}^{K_s}_{p^l}.$$

By denoting the rate of the subcode $\mathcal{C}_l$ by R_l, we get

$$\begin{aligned} R_l &= \frac{1}{n}\sum_{1\le s<l} sK_s\log p + \frac{1}{n}\sum_{l\le s\le r} lK_s\log p \\ &\ge \frac{1}{n}\sum_{1\le s<l} \frac{l}{r}sK_s\log p + \frac{1}{n}\sum_{l\le s\le r} \frac{s}{r}lK_s\log p \\ &= \frac{l}{r}R. \end{aligned} \tag{6}$$

We now apply the inverse channel coding theorem to the code $\mathcal{C}_l$ and to the MC $(p^{r-l}\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ obtained by restricting the input of the original MC to the subgroup $p^{r-l}\mathbb{Z}_{p^r}$. Recalling that C_l denotes the Shannon capacity of such a channel, we get that, if $R_l > C_l$, then the error probability of the code $\mathcal{C}_l$ is bounded away from zero by some constant independent of the block length n. Since $\mathcal{C}_l \subseteq \mathcal{C}$, and since both are $\mathbb{Z}_{p^r}$-codes, by the UEP we get that

$$p_e(\mathcal{C}) = p_e(\mathcal{C}|0) \ge p_e(\mathcal{C}_l|0) = p_e(\mathcal{C}_l).$$

It thus follows that, if $R_l > C_l$, then the error probability of the original $\mathbb{Z}_{p^r}$-code $\mathcal{C}$ itself is bounded away from zero independently from its block length.

By repeating the argument above for all $1 \le l \le r$ and using (6), the following theorem is proved.

Theorem 5: Let $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ be a $\mathbb{Z}_{p^r}$-symmetric MC; denote its $\mathbb{Z}_{p^r}$-capacity by $C_{\mathbb{Z}_{p^r}}$. Then, for every design rate $R > C_{\mathbb{Z}_{p^r}}$ there exists a constant $A_R > 0$ such that

$$p_e(\mathcal{C}) \ge A_R$$

for every $\mathbb{Z}_{p^r}$-code $\mathcal{C}$ of rate not smaller than R.

B. A Coding Theorem for $\mathbb{Z}_{p^r}$-codes

Theorem 5 provides a necessary condition for reliable transmission using $\mathbb{Z}_{p^r}$-codes on $\mathbb{Z}_{p^r}$-symmetric channels: for this to be possible, the rate must not exceed the $\mathbb{Z}_{p^r}$-capacity $C_{\mathbb{Z}_{p^r}}$. However, it is not clear at all whether any rate below $C_{\mathbb{Z}_{p^r}}$ can actually be achieved by means of $\mathbb{Z}_{p^r}$-codes. In principle, there could be other algebraic constraints coming into the picture, which have been overlooked in our analysis. In fact, we shall see that this is not the case: the condition $R < C_{\mathbb{Z}_{p^r}}$ will be shown to be sufficient for reliable transmission using $\mathbb{Z}_{p^r}$-codes over a $\mathbb{Z}_{p^r}$-symmetric MC.

Given a design rate R in $(0, \log p^r)$, we introduce the $\mathbb{Z}_{p^r}$-code ensemble as follows. For every block length n, we set $h := \lfloor(1 - R/\log p^r)n\rfloor$, and consider a random parity-check operator Φ_n uniformly distributed over the the set $\hom(\mathbb{Z}^n_{p^r}, \mathbb{Z}^h_{p^r})$ of all homomorphisms from $\mathbb{Z}^n_{p^r}$ to $\mathbb{Z}^h_{p^r}$. Finally, let $\mathcal{C}_n := \ker\Phi_n$ be the random $\mathbb{Z}_{p^r}$-code obtained as the kernel of Φ_n, i.e., the set of all those n-tuples $\boldsymbol{x}$ in $\mathbb{Z}^n_{p^r}$ such that $\Phi_n\boldsymbol{x} = \mathbf{0}$. Observe that the rate of $\mathcal{C}_n$ is deterministically not smaller than R. We are interested in estimating the average error probability $\overline{p_e(\mathcal{C}_n)}$ of the parity-check ensemble of $\mathbb{Z}_{p^r}$-codes of design rate R.

A first step in our analysis consists in evaluating the average type-spectrum. For any type $\boldsymbol{\theta}$ in $\mathcal{P}(\mathbb{Z}_{p^r})$, let $S_n(\boldsymbol{\theta}) := S_{\mathcal{C}_n}(\boldsymbol{\theta})$ be the number of codewords in the random code $\mathcal{C}_n$ of type $\boldsymbol{\theta}$. We have

$$S_n(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in(\mathbb{Z}_{p^r})^n_{\boldsymbol{\theta}}} \mathbb{1}_{\mathcal{C}_n}(\boldsymbol{x}) = \sum_{\boldsymbol{x}\in(\mathbb{Z}_{p^r})^n_{\boldsymbol{\theta}}} \mathbb{1}_{\{\Phi_n\boldsymbol{x}=\mathbf{0}\}}. \tag{7}$$

The expected value of $S_n(\boldsymbol{\theta})$ can be evaluated as follows.

Lemma 6: For every $\boldsymbol{\theta}$ in $\mathcal{P}_n(\mathbb{Z}_{p^r})$, the average type spectrum of the parity-check ensemble of $\mathbb{Z}_{p^r}$-codes of design rate R is given by

$$\overline{S_n(\boldsymbol{\theta})} = \binom{n}{n\boldsymbol{\theta}} p^{-hs(\boldsymbol{\theta})}$$

where, for $\boldsymbol{\theta}$ in $\mathcal{P}(\mathbb{Z}_{p^r})$, $s(\boldsymbol{\theta})$ denotes the smallest integer $l \ge 0$ such that $\boldsymbol{\theta}(a) = 0$ for all $a \notin p^{r-l}\mathbb{Z}_{p^r}$.

Proof: Consider the standard basis $\{\delta_i\}_{1\le i\le n}$ of $\mathbb{Z}^n_{p^r}$. Then, an equivalent condition for Φ_n to be uniformly distributed over $\hom(\mathbb{Z}^n_{p^r}, \mathbb{Z}^h_{p^r})$ is that the r.v.'s $\Phi_n\delta_i$, $1 \le i \le n$, are mutually independent and uniformly distributed over $\mathbb{Z}^h_{p^r}$.

Let us now fix an n-tuple $\boldsymbol{x}$ of type $\boldsymbol{\theta} \in \mathcal{P}_n(\mathbb{Z}_{p^r})$. From the definition of $s(\boldsymbol{\theta})$ it follows that x_j belongs to $p^{r-s(\boldsymbol{\theta})}\mathbb{Z}_{p^r}$ for all $1 \le j \le n$, and that $x_i \notin p^{r-s(\boldsymbol{\theta})-1}\mathbb{Z}_{p^r}$ for some $1 \le i \le n$. It follows that the r.v. $X_i := x_i\Phi_n\delta_i$ is uniformly distributed over $p^{r-s(\boldsymbol{\theta})}\mathbb{Z}^h_{p^r}$, while the r.v. $X_{-i} := \sum_{j\neq i} x_j\Phi_n\delta_j$ takes values in $p^{r-s(\boldsymbol{\theta})}\mathbb{Z}^h_{p^r}$ and is independent from X_i. Therefore, $\Phi_n\boldsymbol{x} = \sum_{1\le i\le n} x_i\Phi_n\delta_i = X_i + X_{-i}$ is uniformly distributed over $p^{r-s}\mathbb{Z}^h_{p^r}$. So, in particular, $\mathbb{P}(\Phi_n\boldsymbol{x} = \mathbf{0}) = p^{-hs(\boldsymbol{\theta})}$.

Hence, for any type $\boldsymbol{\theta}$ in $\mathcal{P}_n(\mathbb{Z}_{p^r})$ we have

$$\mathbb{E}\left[S_n(\boldsymbol{\theta})\right] = \sum_{\boldsymbol{x}\in(\mathbb{Z}_{p^r})^n_{\boldsymbol{\theta}}} \mathbb{P}(\Phi_n\boldsymbol{x}=\mathbf{0}) = \binom{n}{n\boldsymbol{\theta}} p^{-hs(\boldsymbol{\theta})},$$

the first equality above following from (7) and the linearity of expectation. □

Lemmas 6 and 3 allow us to prove the following fundamental estimate on the average error probability of the parity-check ensemble of $\mathbb{Z}_{p^r}$-codes.

Theorem 7: Let $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ be a $\mathbb{Z}_{p^r}$-symmetric MC, and let $E_{\mathbb{Z}_{p^r}}(R)$ be its $\mathbb{Z}_{p^r}$-error exponent. Then the average error probability of the $\mathbb{Z}_{p^r}$-code ensemble of design rate R satisfies

$$\overline{p_e(\mathcal{C}_n)} \le r\exp\left(-nE_{\mathbb{Z}_{p^r}}(R)\right).$$

Proof: For all $1 \le s \le r$ let

$$\mathcal{C}^s_n := \{\mathbf{0}\}\bigcup\left(\mathcal{C}_n\bigcap\left(\bigcup\nolimits_{s(\boldsymbol{\theta})=s}(\mathbb{Z}_{p^r})^n_{\boldsymbol{\theta}}\right)\right)$$

be the subcode of $\mathcal{C}_n$ consisting of the all-zero codeword and of all the codewords of $\mathcal{C}_n$ whose type $\boldsymbol{\theta}$ is such that $s(\boldsymbol{\theta}) = s$. Observe that $\mathcal{C}^s_n \subseteq p^{r-s}\mathbb{Z}^n_{p^r}$. By the UEP and the union bound we have

$$\overline{p_e(\mathcal{C}_n)} = \overline{p_e(\mathcal{C}_n|\mathbf{0})} \le \sum_{1\le s\le r} \overline{p_e(\mathcal{C}^s_n|\mathbf{0})}. \tag{8}$$

$$\begin{aligned}
\overline{p_e(\mathcal{C}_n^s|\mathbf{0})} &\leq \frac{1}{p^{sn}}\sum_{\boldsymbol{z}}\int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\overline{\left(\sum_{\boldsymbol{\theta}}\frac{S_n(\boldsymbol{\theta})}{\binom{n}{n\boldsymbol{\theta}}}\sum_{\boldsymbol{x}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}{+}\boldsymbol{x})\right)^{\rho}}\mathrm{d}\boldsymbol{y}\\
&\leq \frac{1}{p^{sn}}\sum_{\boldsymbol{z}}\int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\left(\sum_{\boldsymbol{\theta}}\frac{\overline{S_n(\boldsymbol{\theta})}}{\binom{n}{n\boldsymbol{\theta}}}\sum_{\boldsymbol{x}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}{+}\boldsymbol{x})\right)^{\rho}\mathrm{d}\boldsymbol{y}\\
&\leq \frac{1}{p^{sn}}\int_{\mathcal{Y}^n}\left(\sum_{\boldsymbol{z}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\right)\left(\frac{1}{p^{sh}}\sum_{\boldsymbol{z}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\right)^{\rho}\mathrm{d}\boldsymbol{y}\\
&= p^{s\rho(n-h)}\int_{\mathcal{Y}^n}\left(\frac{1}{p^{sn}}\sum_{\boldsymbol{z}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\right)^{1+\rho}\mathrm{d}\boldsymbol{y}
\end{aligned}$$

For every $1 \leq s \leq r$ and $0 \leq \rho \leq 1$, by applying Lemma 3 to the code $\mathcal{C}_n^s$ and the MC $(p^{r-s}\mathbb{Z}_{p^r}, \mathcal{Y}, W)$, then Jensen's inequality and Lemma 6, we get the bound on $p_e(\mathcal{C}_n^s|\mathbf{0})$ displayed at the top of the page, where the summation index $\boldsymbol{z}$ runs over $p^{r-s}\mathbb{Z}_{p^r}^n$, $\boldsymbol{\theta}$ over types in $\mathcal{P}_n(\mathbb{Z}_{p^r})$ such that $s(\boldsymbol{\theta}) = s$, and $\boldsymbol{z}$ over $(\mathbb{Z}_{p^r})^n_{\boldsymbol{\theta}}$, the set of type-$\boldsymbol{\theta}$ n-tuples. Observe that $p^{s\rho(n-h)} \leq \exp(n\rho s/rR)$ while, since the channel is stationary and memoryless

$$\int_{\mathcal{Y}^n}\left(\frac{1}{p^{sn}}\sum_{\boldsymbol{z}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\right)^{1+\rho}\mathrm{d}\boldsymbol{y} = \left(\int_{\mathcal{Y}}\left(\frac{1}{p^{s}}\sum_{z} W^{1/1+\rho}(y|z)\right)^{1+\rho}\mathrm{d}y\right)^{n}.$$

Therefore, we get

$$\overline{p_e\left(\mathcal{C}_n^s|\mathbf{0}\right)} \leq \exp\left(n\rho\frac{s}{r}R - nE_s^0\left(\boldsymbol{u}_s,\rho\right)\right)$$

where $E_s^0(\cdot,\cdot)$ denotes the Gallager exponent of the MC $(p^{r-s}\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ (as defined in (3)) and $\boldsymbol{u}_s$ is the uniform distribution over $p^{r-s}\mathbb{Z}_{p^r}$. Since the MC $(p^{r-s}\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ is $p^{r-s}\mathbb{Z}_{p^r}$-symmetric, $\boldsymbol{u}_s$ is the optimal input distribution. Then, by optimizing the exponent $E_s^0(\boldsymbol{u}_s, \rho) - \rho s/rR$ over ρ in $[0, 1]$, we get the error exponent $E_s(r/sR)$, so that

$$\overline{p_e(\mathcal{C}_n^s|\mathbf{0})} \leq \exp\left(nE_s\left(\frac{s}{r}R\right)\right).$$

The claim now follows by combining the above inequality with (8), and recalling Definition 4. □

Standard probabilistic arguments allow us to prove the following corollary of Theorem 7, estimating the asymptotic error exponent of the typical $\mathbb{Z}_{p^r}$-code.

Corollary 8: Let $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ be a $\mathbb{Z}_{p^r}$-symmetric MC of $\mathbb{Z}_{p^r}$-capacity $C_{\mathbb{Z}_{p^r}}$ and $\mathbb{Z}_{p^r}$-error exponent $E_{\mathbb{Z}_{p^r}}(R)$. Then, for every $0 < R < C_{\mathbb{Z}_{p^r}}$, we have

$$\liminf_{n\in\mathbb{N}} -\frac{1}{n}\log p_e(\mathcal{C}_n) \geq E_{\mathbb{Z}_{p^r}}(R)$$

with probability one over the $\mathbb{Z}_{p^r}$-coding ensemble of design rate R.

Proof: With no loss of generality, we can restrict ourselves to rates $0 \leq R < C_{\mathbb{Z}_{p^r}}$, since otherwise $E_{\mathbb{Z}_{p^r}}(R) = 0$ and the claim is trivial as $p_e(\mathcal{C}_n) \leq 1$. For any $0 < \varepsilon < E_{\mathbb{Z}_{p^r}}(R)$, $n \in \mathbb{N}$ define the event

$$A_n^\varepsilon := \{p_e(\mathcal{C}_n) \geq r\exp(-n(E_{\mathbb{Z}_{p^r}}(R) - \varepsilon))\}.$$

By applying Theorem 7 and the Markov inequality, we obtain

$$\begin{aligned}
\mathbb{P}(A_n^\varepsilon) &\leq \mathbb{P}\left(p_e(\mathcal{C}_n) \geq \frac{1}{r}\exp(n\varepsilon)\overline{p_e(\mathcal{C}_n)}\right)\\
&\leq r\exp(-n\varepsilon).
\end{aligned}$$

Then $\sum_n \mathbb{P}(A_n^\varepsilon) \leq \sum_n \exp(-n\varepsilon) < +\infty$, and the Borel–Cantelli lemma implies that with probability one the event A_n^ε occurs for finitely many n in $\mathbb{N}$. Therefore, with probability one $\liminf_n -1/n\log p_e(\mathcal{C}_n) \geq E_{\mathbb{Z}_{p^r}}(R) - \varepsilon$. Finally, the claim follows from the arbitrariness of ε in $(0, E_{\mathbb{Z}_{p^r}}(R))$. □

Corollary 9: Let $(\mathbb{Z}_{p^r}, \mathcal{Y}, W)$ be a $\mathbb{Z}_{p^r}$-symmetric MC, and let $C_{\mathbb{Z}_{p^r}}$ be its $\mathbb{Z}_{p^r}$-capacity. Then, for all $0 \leq R < C_{\mathbb{Z}_{p^r}}$ there exist $\mathbb{Z}_{p^r}$-codes $\mathcal{C}$ of rate not smaller than R and arbitrarily low error probability.

C. On Tightness of the Error Exponent

Theorem 7 provides an exponential upper bound on the average error probability of the parity-check ensemble of $\mathbb{Z}_{p^r}$-codes on a $\mathbb{Z}_{p^r}$-symmetric MC. Corollary 8 states that the same error exponent is asymptotically achieved by a typical code sequence sampled from the $\mathbb{Z}_{p^r}$-code ensemble.

A natural question arising is whether these bounds are tight. We conjecture that $E_{\mathbb{Z}_{p^r}}(R)$ is the correct error exponent for the average $\mathbb{Z}_{p^r}$-code at any rate $0 < R < C_{\mathbb{Z}_{p^r}}$, i.e., that

$$\lim_{n} -\frac{1}{n}\log\overline{p_e(\mathcal{C}_n)} = E_{\mathbb{Z}_{p^r}}(R). \tag{9}$$

No proof of (9) in its full generality will be presented here. Rather, we shall confine ourselves to consider the high-rate and the low-rate regimes.

Theorem 10: For any nontrivial $\mathbb{Z}_{p^r}$-symmetric MC, there exist some $0 < R_0 \leq R_1 < C_{\mathbb{Z}_{p^r}}$ such that (9) holds true for the $\mathbb{Z}_{p^r}$-code ensemble of design rate $R \in (0, R_0)\cup(R_1, C_{\mathbb{Z}_{p^r}})$.

Proof: First we concentrate on the high-rate regime. For rates R close enough to $C_{\mathbb{Z}_{p^r}}$, from the continuity of the exponents $E_s(R)$, it follows that $E_{\mathbb{Z}_{p^r}}(R) = E_s(s/rR)$ for one of the channels $(p^{r-s}\mathbb{Z}_{p^r}, \mathcal{Y}, P)$ whose normalized capacity

r/sC_s coincides with the $\mathbb{Z}_{p^r}$-capacity $C_{\mathbb{Z}_{p^r}}$. It is known that close to capacity the random-coding exponent coincides with the sphere-packing exponent [2], [3]. Then, by applying the sphere-packing bound to the subcode $\mathcal{C}_n \cap p^{r-s}\mathbb{Z}_{p^r}^n$ (whose rate is not smaller than s/rR), we get that, for all rates R not smaller than some $0 < R_1 < C_{\mathbb{Z}_{p^r}}$

$$\begin{aligned} E_{\mathbb{Z}_{p^r}}(R) &= E_s(\frac{s}{r}R) \\ &\geq \limsup_n -\frac{1}{n}\log\overline{p_e(\mathcal{C}_n \cap p^{r-s}\mathbb{Z}_{p^r}^n)} \\ &\geq \limsup_n -\frac{1}{n}\log\overline{p_e(\mathcal{C}_n)}. \end{aligned} \tag{10}$$

We shall now concentrate on showing the validity of (9) in the low-rate regime. First, observe that at rate $R = 0$

$$E_1(0) \leq E_2(0) \leq \cdots \leq E_r(0), \tag{11}$$

the inequalities above being strict on nontrivial $\mathbb{Z}_{p^r}$-symmetric MCs. From the continuity of the error exponents as functions of the rate R, it follows that for any nontrivial $\mathbb{Z}_{p^r}$-symmetric MC

$$E_{\mathbb{Z}_{p^r}}(R) = E_1\left(\frac{1}{r}R\right), \quad \forall R \leq R_0, \tag{12}$$

for some $R_0 > 0$.

Notice that $\mathcal{C}_n \cap p^{r-1}\mathbb{Z}_{p^r}^n$ coincides with the $\mathbb{Z}_p$-linear code ensemble of rate $\frac{1}{r}R$. It is known [35] that $E_1(\frac{1}{r}R)$ is the correct error exponent for the average $\mathbb{Z}_p$-linear code. In fact, the arguments developed in [36] in order to prove tightness of the error exponent for the average code sampled from the random coding ensemble only require pairwise independence of the random codewords. In the $\mathbb{Z}_p$-linear ensemble the events $\{\boldsymbol{x} \in \mathcal{C}_n\}$ and $\{\boldsymbol{w} \in \mathcal{C}_n\}$ are independent whenever $\boldsymbol{x}$ and $\boldsymbol{w}$ are linearly independent in $\mathbb{Z}_p^n$. Since the number of elements of $p^{r-1}\mathbb{Z}_{p^r}^n$ linearly dependent on any $\boldsymbol{x}$ in $p^{r-1}\mathbb{Z}_{p^r}^n$ is at most p, the arguments of [36] can still be used to show that

$$\limsup_n -\frac{1}{n}\log\overline{p_e(\mathcal{C}_n \cap p^{r-1}\mathbb{Z}_{p^r}^n)} \leq E_1(\frac{1}{r}R) = E_{\mathbb{Z}_{p^r}}(R).$$

Then, since $p_e(\mathcal{C}_n) \geq p_e(\mathcal{C}_n \cap p^{r-1}\mathbb{Z}_{p^r}^n)$, from (12) it follows that

$$\limsup_n -\frac{1}{n}\log\overline{p_e(\mathcal{C}_n)} \leq E_{\mathbb{Z}_{p^r}}(R), \quad \forall R \leq R_0. \tag{13}$$

Finally, the claim follows from (10), (13) and Theorem 7. □

Notice that, for $r \geq 2$, strict inequalities in (11) imply that

$$E_{\mathbb{Z}_{p^r}}(R) < E_r(R), \quad R \leq R_0. \tag{14}$$

Therefore, for $r \geq 2$ on any nontrivial $\mathbb{Z}_{p^r}$-symmetric MC, the average $\mathbb{Z}_{p^r}$-code exhibits poorer performance than the average code (i.e., a code sampled from the random-coding ensemble). This result had been first conjectured in [4], where the author hypothesized that the random-coding exponent of any G-symmetric MC is achieved by the average G-code only if $G \simeq \mathbb{Z}_p^r$ for prime p, namely when G admits Galois field structure.

However, it can be shown that, at low rates, $E_{\mathbb{Z}_{p^r}}(R)$ is not the correct error exponent for the $\mathbb{Z}_{p^r}$-code ensemble. In fact, similarly to the random-coding ensemble and the linear-coding ensemble [35], it can be shown that at low rates the error exponent of a typical $\mathbb{Z}_{p^r}$-code is higher than $E_{\mathbb{Z}_{p^r}}(R)$. This is because the average error probability is affected by an asymptotically negligible fraction of codes with poor behavior. In other words, at low rates the bound of Corollary 8 is not tight. In a forthcoming work [30], we shall show that the typical $\mathbb{Z}_{p^r}$-code achieves the expurgated error exponent on many $\mathbb{Z}_{p^r}$-symmetric MCs of interest, including the p^r-PSK AWGN channel. Since it is known that the random-coding ensemble does not achieve the expurgated error exponent with probability one, this will show that at low rates hierarchies for the average and the typical error exponent can be reversed: while the average random code behaves better than the average group code, the typical group code exhibits better performance than the typical random code.

IV. $\mathbb{Z}_{p^r}$-Codes Achieve Capacity on the p^r-PSK AWGN Channel

This section will focus on the p^r-PSK AWGN channel, for which it will be shown that the $\mathbb{Z}_{p^r}$-capacity $C_{\mathbb{Z}_{p^r}}$ coincides with the Shannon capacity C. As a consequence, $\mathbb{Z}_{p^r}$-codes are capacity-achieving for this important family of symmetric MCs, thus confirming a conjecture of Loeliger [6].

Throughout this section p will be some given prime number, r a fixed positive integer. The base of log (and thus of the entropy function H) will be p. For m in $\mathbb{N}$, $\xi_m = e^{2\pi/mi} \in \mathbb{C}$ will denote a primitive mth root of 1. $(\mathbb{Z}_{p^r}, \mathbb{C}, W)$ will denote the p^r-PSK AWGN channel, with input $\mathcal{X}$ identified with $\mathbb{Z}_{p^r}$, output $\mathcal{Y}$ identified with the complex field $\mathbb{C}$, and transition probability densities accordingly given by $W(y|x) = 1/2\sigma^2 e^{-||y-\xi_{p^r}^x||^2/2\sigma^2}$.

Recall that, by Definition 4, $C_{\mathbb{Z}_{p^r}} = \min_{1\leq l\leq r} r/lC_l$, where C_l is the Shannon capacity of the MC $(r/l\mathbb{Z}_{p^r}, \mathbb{C}, W)$, i.e., the AWGN channel with input restricted to the p^l-PSK constellation. Hence, the condition $C = C_{\mathbb{Z}_{p^r}}$ is equivalent to $rC_l \geq lC_r$ for all $1 \leq s, l \leq r$. A simple inductive argument shows that this is in turn equivalent to

$$qC_{q+1} \leq (q+1)C_q, \quad \forall 1 \leq q \leq r-1. \tag{15}$$

The rest of the section will be devoted to the proof of (15). The result will be achieved through a series of technical intermediate steps.

We start by introducing some related probability densities which will play a key role in the sequel:

- for every $1 \leq q \leq r$, λ_q in $\mathcal{P}(\mathbb{C})$ defined by
$$\lambda_q(y) := \frac{1}{p^q}\sum_{x\in p^{r-q}\mathbb{Z}_{p^r}} W(y|x) = \frac{1}{p^q}\sum_{j=0}^{p^q-1} W(y\xi_{p^q}^j|0)$$
(with the second equality above following from the symmetry of the MC);
- for every $1 \leq q \leq r-1$ and $y \in \mathbb{C}$, $\boldsymbol{\nu}_q(y)$ in $\mathcal{P}(\mathbb{Z}_p)$ defined by
$$[\boldsymbol{\nu}_q(y)](a) := \frac{\lambda_q(y\xi_{p^{q+1}}^a)}{p\lambda_{q+1}(y)}; \tag{16}$$

- for every $1 \leq q \leq r$ and $y \in \mathbb{C}$ a probability distribution $\boldsymbol{\omega}_q(y)$ in $\mathcal{P}(p^{r-q}\mathbb{Z}_{p^r})$ defined by

$$[\boldsymbol{\omega}_q(y)](x) := \frac{1}{p^q \lambda_q(y)} W(y|x).$$

For any $1 \leq q \leq r$, consider the p^q-PSK AWGN channel $(p^{r-q}\mathbb{Z}_{p^r}, \mathbb{C}, W)$. Since it is symmetric, its Shannon capacity C_q is achieved by a uniform distribution over the input $p^{r-q}\mathbb{Z}_{p^r}$. The corresponding output probability density is given by

$$\sum_{x \in p^{r-q}\mathbb{Z}_{p^r}} p^{-q} W(y|x) = \lambda_q(y)$$

so that

$$C_q = H(\lambda_q) - H\left(W(\cdot|0)\right). \tag{17}$$

Therefore, (15) is equivalent to

$$H\left(W(\cdot|0)\right) + qH(\lambda_{q+1}) \leq (q+1)H(\lambda_q)\,,\ 1 \leq q < r. \tag{18}$$

The following result relates the entropies of the discrete probability distributions $\boldsymbol{\omega}_q(y)$ and $\boldsymbol{\nu}_q(y)$ to those of the continuous densities λ_q and $W(\cdot|0)$.

Lemma 11: For every $1 \leq q < r$

$$H(W(\cdot|0)) = H(\lambda_q) - q + \int_{\mathbb{C}} \lambda_q(y) H(\boldsymbol{\omega}_q(y)) \mathrm{d}y\,; \tag{19}$$

$$H(\lambda_q) = H(\lambda_{q+1}) - 1 + \int_{\mathbb{C}} \lambda_{q+1}(y) H(\boldsymbol{\nu}_q(y)) \mathrm{d}y. \tag{20}$$

Proof: See Appendix B, Subsection A. □

As a consequence of Lemma 11 we have that (18) is equivalent to

$$q \int_{\mathbb{C}} \lambda_{q+1}(y) H\left(\boldsymbol{\nu}_q(y)\right) \mathrm{d}y \geq \int_{\mathbb{C}} \lambda_q(y) H\left(\boldsymbol{\omega}_q(y)\right) \mathrm{d}y \tag{21}$$

for all $1 \leq q \leq r-1$.

We pass now to the core of the argument which relies on geometric considerations. For $1 \leq q < r$, fix an arbitrary point y in the output set $\mathbb{C}$, and consider the multiset of likelihood values for the input p^q-PSK, given by

$$\begin{aligned}\mathcal{W}_q(y) &:= \{W(y|0), W(y|1), \dots, W(y|p^q-1)\}\\ &= \left\{W(y|0), W(y\xi_{p^q}|0), \dots, W(y\xi_{p^q}^{p^q-1}|0)\right\}.\end{aligned}$$

Since the p^{q+1}-PSK constellation is the disjoint union of p copies of the p^q-PSK constellation each rotated by an angle multiple of $2\pi/p^{q+1}$, we have

$$\mathcal{W}_{q+1}(y) = \bigcup_{0 \leq j < p} \mathcal{W}_q(y\xi_{p^{q+1}}^j)\,. \tag{22}$$

The geometry of the p^{q+1}-PSK constellation implies that the ordering of the multiset of likelihoods $\mathcal{W}_{q+1}(y)$ satisfies a fundamental nesting property with respect to the partition (22). Roughly speaking, this property consists in that all the subsets $\mathcal{W}_q(y\xi_{p^{q+1}}^j)$ contain the same amount of highest values of the set $\mathcal{W}_{q+1}(y)$. More precisely, if $w_{q,l}^k$ is the kth highest value in $\mathcal{W}_q(y\xi_{p^{q+1}}^l)$ for some $1 \leq k \leq p^q$ and $0 \leq l \leq p-1$, then each of the subsets $\mathcal{W}_q(y\xi_{p^{q+1}}^j)$ contains at least $k-1$ elements not smaller than $w_{q,l}^k$. This is formalized in the following lemma.

Lemma 12: For every $1 \leq q < r$ and $y \in \mathcal{Y}$, there exists a partition

$$\mathcal{W}_{q+1}(y) = \bigcup_{1 \leq k \leq p^q} \mathcal{W}_q^k(y)$$

where each multiset $\mathcal{W}_q^k(y) = \{w_{q,0}^k, w_{q,1}^k, \dots, w_{q,p-1}^k\}$ is such that, for all $0 \leq j,\ i \leq p-1$, $w_{q,j}^k$ belongs to $\mathcal{W}_q(y\xi_{p^{q+1}}^j)$, and

$$0 \leq k < k' < p^q \implies w_{q,i}^k(y) \geq w_{q,j}^{k'}(y). \tag{23}$$

Proof: See Appendix B, Subsection B. □

Observe that, with the notation introduced in Lemma 12

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

If we consider the probability distribution $\boldsymbol{\omega}_{q,j}(y)$ in $\mathcal{P}(\mathbb{Z}_{p^q})$ defined by

$$\left[\boldsymbol{\omega}_{q,j}(y)\right](k) := \frac{1}{p^q \lambda_q(y\xi_{p^{q+1}}^j)} w_{q,j}^k$$

we have that the entropies $H\left(\boldsymbol{\omega}_{q,j}(y)\right)$ and $H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)\right)$ do coincide, as $\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)$ and $\boldsymbol{\omega}_{q,j}(y)$ simply differ by a permutation of $\mathbb{Z}_{p^q}$.

Consider now the p-adic expansion map $\zeta : \mathbb{Z}_{p^q} \to \mathbb{Z}_p^q$, defined as follows: if $s \in \mathbb{Z}_{p^q}$ is such that $s = \sum_{0 \leq k < q} \rho_k p^k$ for $0 \leq \rho_k < p$, then $\zeta(s) := (\rho_0, \dots, \rho_{q-1})$. It is a standard fact that ζ is a bijection. Let $Z(y,j)$ be a $\mathbb{Z}_{p^q}$-valued random variable with distribution $\boldsymbol{\omega}_{q,j}(y)$ and let $\boldsymbol{Y}(y,j) = (Y_1(y,j), \dots, Y_q(y,j)) := \zeta \circ Z(y,j)$ the corresponding $\mathbb{Z}_p^q$-valued random variable. For $1 \leq \alpha \leq q$, let $\boldsymbol{\delta}_q^\alpha(y,j) \in \mathcal{P}(\mathbb{Z}_p)$ be the probability distribution of $Y_\alpha(y,j)$. A straightforward computation shows that

$$\left[\boldsymbol{\delta}_q^\alpha(y,j)\right](s) = \frac{1}{p^q \lambda_q(y\xi_{p^{q+1}}^j)} \sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_{q,j}^{\tilde{h}p^{\alpha+1}+sp^\alpha+h}. \tag{24}$$

We can now prove the following upper bound on the entropy $H\left(\boldsymbol{\omega}_{q,j}(y)\right)$.

Lemma 13: For every $1 \leq \alpha \leq q$

$$H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)\right) \leq \sum_{1 \leq \alpha \leq q} H\left(\boldsymbol{\delta}_q^\alpha(y,j)\right). \tag{25}$$

Proof: We have

$$\begin{aligned}H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)\right) &= H\left(\boldsymbol{\omega}_{q,j}(y)\right) = H\left(Z(y,j)\right)\\ &= H\left(\boldsymbol{Y}(y,j)\right)\\ &\leq \sum_{1 \leq \alpha \leq q} H\left(Y_\alpha(y,j)\right)\\ &= \sum_{1 \leq \alpha \leq q} H\left(\boldsymbol{\delta}_q^\alpha(y,j)\right)\end{aligned}$$

where we first used the fact that $\boldsymbol{Y}(y,j) = \zeta \circ Z(y,j)$ where ζ is a bijection, then apply chain rule for entropy, and finally the conditional entropy bound (see [37] for instance). □

The next step of our argument consists in showing that the probability distribution $\boldsymbol{\nu}_q(y)$ in $\mathcal{P}(\mathbb{Z}_p)$—as defined in (16)—is a convex combination of $\boldsymbol{\delta}_q^{\alpha}(y)$ for $1 \leq \alpha \leq q$, so that—by Jensen's inequality—its entropy estimates from above the corresponding convex combination of the entropies of $\boldsymbol{\delta}_q^{\alpha}(y)$. The proof of Lemma 14 below is based on certain properties of the so called "permutahedron" of a given point in the n-dimensional Euclidean space, which are derived in Appendix B, Subsection C.

Lemma 14: For every $1 \leq \alpha < q < r$, and $y \in \mathbb{C}$, we have

$$H\left(\sum_{j\in\mathbb{Z}_p} \frac{\lambda_q(y\xi_{p^{q+1}}^j)}{p\lambda_{q+1}(y)}\boldsymbol{\delta}_q^{\alpha}(y,j)\right) \leq H\left(\boldsymbol{\nu}_q(y)\right). \tag{26}$$

Proof: See Appendix B, Subsection C. □

We are finally in the position to prove the following fundamental result.

Theorem 15: For every positive integer q, let C_q be the Shannon capacity of the p^q-PSK AWGN channel. Then

$$qC_{q+1} \leq (q+1)C_q\,. \tag{27}$$

Proof: Fix an arbitrary output $y \in \mathbb{C}$. By successively applying (26), the Jensen inequality, and (25), we obtain

$$\begin{aligned} qH\left(\boldsymbol{\nu}_q(y)\right) &= \sum_{1\leq\alpha\leq q} H\left(\boldsymbol{\nu}_q(y)\right)\\ &\geq \sum_{1\leq\alpha\leq q} H\left(\sum_{j\in\mathbb{Z}_p}\frac{\lambda_q(y\xi_{p^{q+1}}^j)}{p\lambda_{q+1}(y)}\boldsymbol{\delta}_q^{\alpha}(y,j)\right)\\ &\geq \sum_{1\leq\alpha\leq q}\sum_{j\in\mathbb{Z}_p}\frac{\lambda_q(y\xi_{p^{q+1}}^j)}{p\lambda_{q+1}(y)}H\left(\boldsymbol{\delta}_q^{\alpha}(y,j)\right)\\ &\geq \sum_{j\in\mathbb{Z}_p}\frac{\lambda_q(y\xi_{p^{q+1}}^j)}{p\lambda_{q+1}(y)}H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)\right). \end{aligned}$$

Therefore

$$\begin{aligned} &\int_{\mathbb{C}} \lambda_q(y)H\left(\boldsymbol{\omega}_q(y)\right)\mathrm{d}y\\ &\quad= \int_{\mathbb{C}} \frac{1}{p}\sum_{j\in\mathbb{Z}_p}\lambda_q(y\xi_{p^{q+1}}^j)H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)\right)\mathrm{d}y\\ &\quad\leq q\int_{\mathbb{C}} \lambda_{q+1}(y)H\left(\boldsymbol{\nu}_q(y)\right)\mathrm{d}y. \end{aligned}$$

Thus, (21) holds true for all $1 \leq q \leq r-1$, and this has previously been shown to be equivalent to the claim. □

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

Corollary 16: For any prime p and positive integer r, the $\mathbb{Z}_{p^r}$-capacity of the p^r-PSK AWGN channel coincides with its Shannon capacity, i.e., $C_{\mathbb{Z}_{p^r}} = C_r$.

Combining Corollary 16 with Corollary 9, we can finally state a result first conjectured by Loeliger in [6].

Corollary 17: $\mathbb{Z}_{p^r}$-codes achieve the capacity of the p^r-PSK AWGN channel.

We observe that the key step for the validity of the results of this section is Lemma 12. In fact, while all the other derivations do not depend on the particular $\mathbb{Z}_{p^r}$-symmetric channel, Lemma 12 heavily relied on the geometry of the p^r-PSK constellation. Hence, for all $\mathbb{Z}_{p^r}$-symmetric channels for which Lemma 12 holds, Theorem 15 and Corollary 16 continue to be true. This is for instance the case for hard-decoded p^r-PSK AWGN channels and for the p^r-ary symmetric channel of Example 1.

V. A Symmetric Channel for Which Group Codes Do Not Achieve Capacity

In the previous section, it was shown that, for the p^r-PSK AWGN channel, $\mathbb{Z}_{p^r}$-capacity and Shannon capacity do coincide. At this point, the question arising is whether this is the case for any higher dimensional GU constellation admitting a generating group isomorphic to $\mathbb{Z}_{p^r}$. In this section, we shall show that the answer is negative in general. In fact, we shall provide a whole family of counterexamples based on the 3-D constellations introduced in Example 5. We shall prove that $\mathbb{Z}_{2^r}$-capacity of the AWGN channel with input constrained to 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.

For some integer $r \geq 3$, we consider the family of GU constellations $K_{2^r}^{\beta}$, parameterized by $\beta \in [0, +\infty)$ and defined by

$$K_{2^r}^{\beta} := \left\{ x_k := \sqrt{\frac{1}{1+\beta^2}}\left(e^{2\pi/2^r ki}, (-1)^k\beta\right),\ 1\leq k\leq 2^r \right\}.$$

Observe that $K_{2^r}^{\beta}$ is 3-D for $\beta > 0$, and 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 nonisomorphic 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 $\left(K_{2^r}^{\beta}, \mathbb{R}^3, W\right)$, whose $\mathbb{Z}_{2^r}$-capacity will be denoted by $C_{\mathbb{Z}_{2^r}}(\beta)$. For $1 \leq s \leq r$, $C_{2^s}(\beta)$ will denote the capacity of the AWGN channel with input restricted to the subconstellation $\{x_{k2^{r-s}} | 1 \leq k \leq 2^s\}$, so that

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

We start our analysis by considering the limit case $\beta = 0$. In this case, $K_{2^r}^{0}$ coincides with an $\mathbb{R}^3$ embedding of the 2^r-PSK constellation and it is clearly not 3-D since it does not span $\mathbb{R}^3$. Since orthogonal components of the AWGN are mutually independent, for every $1 \leq s \leq r$, $C_{2^s}(0)$ coincides with the Shannon capacity of the 2^s-PSK-AWGN channel. Thus, all the results of Section IV hold true: in particular, the $\mathbb{Z}_{2^r}$-capacity and the Shannon capacity coincide, i.e.,

$$C_{\mathbb{Z}_{2^r}}(0) = C_{2^r}(0)\,. \tag{28}$$

Similar arguments can be applied, for every given $\beta > 0$, to the subconstellation

$$\left\{\left(\sqrt{\frac{1}{1+\beta^2}}e^{2\pi/2^{r-1}ki}, \sqrt{\frac{\beta^2}{1+\beta^2}}\right),\ 1\leq k\leq 2^{r-1}\right\}$$

coinciding with a 3-D embedding of a rescaled 2^{r-1}-PSK. Applying the results of the previous section, we get that

$$(r-1)C_{2^s}(\beta) \geq sC_{2^{r-1}}(\beta), \quad 1 \leq s \leq r-1. \tag{29}$$

Thus, for every $\beta \in (0,+\infty)$, in order to check whether $C_{2^r}(\beta)$ and $C_{\mathbb{Z}_{2^r}}(\beta)$ do coincide, one is only left to compare the normalized capacities $C_{2^r}(\beta)$ and $r/r-1C_{2^{r-1}}(\beta)$.

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 constellation, with the 2^{r-1} even-labeled points $\{x_{2k} | 1 \leq k \leq 2^{r-1}\}$ collapsed into the point $(0,1)$, and the odd labeled ones $\{x_{2k-1} | 0 \leq k \leq 2^{r-1}\}$ into the point $(0,-1)$. Let us define this limit constellation as $K^{\infty} := \{(0,1),(0,-1)\}$. Notice that, for every finite standard deviation value $\sigma > 0$, the Shannon capacity of the K^{∞}-AWGN channel is strictly positive, while $C_{2^{r-1}}(\infty) = 0$, since it is the capacity of an MC with indistinguishable inputs. A continuity argument yields the following result.

Proposition 18: For every finite variance $\sigma^2 > 0$ and any integer $r \geq 2$, the family of $K_{2^r}^{\beta}$-AWGN channels satisfies

$$\lim_{\beta \longrightarrow \infty} C_{2^r}(\beta) = C(\infty) > 0, \qquad \lim_{\beta \longrightarrow \infty} C_{\mathbb{Z}_{2^r}}(\beta) = 0.$$

Proof: See Appendix C. □

Theorem 5 and Proposition 18 have the following immediate consequence.

Corollary 19: For all variances $\sigma^2 > 0$, there exists a positive finite $\overline{\beta}$ such that, for any $\beta > \overline{\beta}$, $\mathbb{Z}_{2^r}$-codes do not achieve Shannon capacity of the $K_{2^r}^{\beta}$-AWGN channel.

On the other hand, it can be proved that $(r-1)C_{2^r}(0) < rC_{2^{r-1}}(0)$, for all $r > 2$. Then, by a continuity argument it can be shown that, for sufficiently small values of β, $C_{2^r}(\beta) = C_{\mathbb{Z}_{2^r}(\beta)}$, so that $\mathbb{Z}_{2^r}$-codes do achieve capacity of the $K_{2^r}^{\beta}$-AWGN channel. Fig. 3 refers to the case $2^r = 8$: the normalized Shannon capacities $C_8(\beta)$ and $3/2C_4(\beta)$ are plotted as a functions of the parameter β (Monte Carlo simulations).

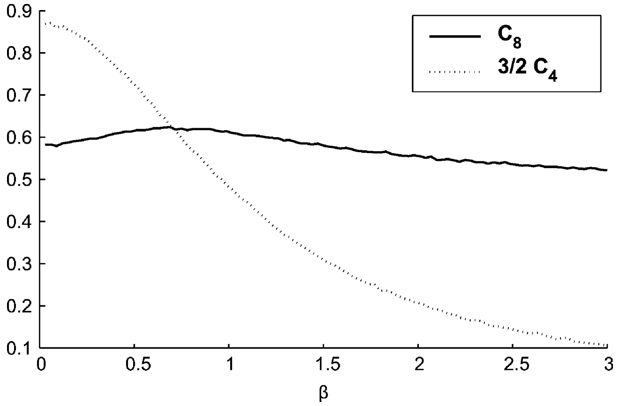

Fig. 3. Shannon capacity and $\mathbb{Z}_8$-capacity of K_8^{β}-AWGN channel as functions of β. It can be seen as $C_{\mathbb{Z}_8}(\beta) = \min\{C_8(\beta), 3/2C_4(\beta)\}$ coincides with $C_8(\beta)$ only for values of β below a certain threshold. The maxima of $C_8(\beta)$ and $C_{\mathbb{Z}_8}(\beta)$ are achieved for values of β close to this threshold, i.e., the two problems of optimizing respectively Shannon capacity and $\mathbb{Z}_8$-capacity seem to have similar solutions. The optimal values are strictly greater than the 8-PSK AWGN capacity.

VI. Arbitrary Finite Abelian Group

A. *The Algebraic Structure of Finite Abelian Groups*

In order to generalize the results of Section III, some basic facts about the structure of finite Abelian groups need to be recalled. We refer to standard textbooks in algebra ([38] for instance) for a more detailed treatment.

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

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

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

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

Write $\mu_M = p_1^{r_1} \cdots p_s^{r_s}$, where $p_1 < p_2 < \cdots < p_s$ are distinct primes and $r_1, \ldots, r_s$ are nonnegative 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

$$M = M_{(p_1^{r_1})} \oplus \cdots \oplus M_{(p_s^{r_s})}. \tag{30}$$

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

$$M_{(p_i^{r_i})} = \mathbb{Z}_{p_i}^{k_{i,1}} \oplus \mathbb{Z}_{p_i^2}^{k_{i,2}} \oplus \cdots \oplus \mathbb{Z}_{p_i^{r_i}}^{k_{i,r_i}}. \tag{31}$$

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

$$\boldsymbol{k}^M = \left(k_{i,j} \,| 1 \leq i \leq s, \, 1 \leq j \leq r_i^M\right)$$

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 $\sigma = (p_1, \ldots, p_s)$, we shall 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 shall denote by $M_{\boldsymbol{k}}$ the finite Abelian group having type $\boldsymbol{k}$.

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

If M and L are finite Abelian groups and $\phi \in \mathrm{Hom}(M,L)$, 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 noninjective if M is not σ^L-adapted or if any of the multiplicities in M is strictly larger than the corresponding in L.

B. The Inverse Channel Coding Theorem for Abelian G-Codes

Suppose we are given a finite Abelian group G having spectrum $\sigma^G = (p_1, \dots, p_s)$, multiplicity $\boldsymbol{r}^G = (r_1^G, \dots, r_s^G)$, and type $\boldsymbol{k}^G$. Consider a G-code $M \leq G^n$ of rate $R = 1/n \log |M|$. Clearly, M is σ^G-adapted and $r_i^M \leq r_i^G$ for all $1 \leq i \leq s$, since otherwise the immersion of M in G^n would not be injective. Then M can be decomposed as illustrated above in (30) and (31). Let us fix a matrix

$$\boldsymbol{L} = \left(L_{i,j} \in \mathbb{Z}^+ \mid 1 \leq i \leq s,\ 1 \leq j \leq r_i^G\right)$$

such that $L_{i,j} \leq j$ for every i and j. We shall say that $\boldsymbol{L}$ is an $\boldsymbol{r}^G$-compatible matrix. Define

$$M(\boldsymbol{L}) = \bigoplus_{1 \leq i \leq s} \bigoplus_{1 \leq j \leq r_i^G} p_i^{j-L_{i,j}} \mathbb{Z}_{p_i^j}^{k_{i,j}}. \qquad (32)$$

An immediate consequence of the previous considerations is that

$$M(\boldsymbol{L}) \subseteq G_{\boldsymbol{L}}^n, \quad G_{\boldsymbol{L}} := \bigoplus_{1 \leq i \leq s} \sum_{1 \leq j \leq r_i^G} p_i^{j-L_{i,j}} G_{(p_i^j)}.$$

These inclusions automatically give information-theoretic constraints to the possibility of reliable transmission using this type of codes. Denote by $R_{\boldsymbol{L}}$ the rate of $M(\boldsymbol{L})$ and by $C_{\boldsymbol{L}}$ the capacity of the subchannel having as input alphabet the subgroup $G_{\boldsymbol{L}}$. Then, a necessary condition for $p_e(M)$ not to be bounded away from 0 by some constant independent of n is that $R_{\boldsymbol{L}} \leq C_{\boldsymbol{L}}$ for every $\boldsymbol{r}^G$-compatible $\boldsymbol{L}$. This does not give explicit constraints yet to the rates R at which reliable transmission is possible using G-codes. For this, some extra work is needed using the structure of the Abelian groups $M(\boldsymbol{L})$. Notice that

$$R_{\boldsymbol{L}} = \frac{1}{n} \sum_{1 \leq i \leq s} \sum_{1 \leq j \leq r_i^G} L_{i,j} k_{i,j} \log p_i.$$

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

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

From the preceding definition, and recalling that $\log |M| = Rn$, we have $k_{i,j} = Rn\alpha_{i,j}/j \log p_i$.

Denote now by $\mathcal{P}(\boldsymbol{r}^G)$ the space of probability distributions $(\alpha_{i,j})$ on the set of pairs (i, j) with $1 \leq i \leq s$ and $1 \leq j \leq r_i^G$. We introduce the following definition.

Definition 20: Let G be a finite Abelian group of spectrum $\sigma^G = (p_1, \dots, p_s)$ and type $\boldsymbol{k}^G$. Let $(G, \mathcal{Y}, W)$ be a G-symmetric MC. For each $\boldsymbol{r}^G$-compatible matrix $\boldsymbol{L}$, let $C_{\boldsymbol{L}}$ be the capacity of the MC $(G_{\boldsymbol{L}}, \mathcal{Y}, W)$. The *$G$-capacity* of the MC $(G, \mathcal{Y}, W)$ is

$$C_G := \max_{\boldsymbol{\alpha} \in \mathcal{P}(\boldsymbol{r}^G)} \min_{\substack{\boldsymbol{L} \neq \boldsymbol{0} \\ \boldsymbol{r}^G\text{-comp.}}} \frac{C_{\boldsymbol{L}}}{\sum\limits_{1 \leq i \leq s} \sum\limits_{1 \leq j \leq r_i^G} \frac{L_{i,j}}{j} \alpha_{i,j}} \qquad (33)$$

where $\boldsymbol{L} \neq \boldsymbol{0}$ means that $L_{i,j} \neq 0$ for some i, j.

It clearly follows from our previous considerations that C_G is an upper bound to the rates at which reliable transmission is possible using G-codes. More precisely, we have the following result which is an immediate consequence of the inverse channel coding theorem.

Theorem 21: Consider a G-symmetric channel and let C_G be its G-capacity. Then, for every rate $R > C_G$ there exists a constant $A_R > 0$, such that the error probability of any G-code $\mathcal{C}$ of rate R satisfies

$$p_e(\mathcal{C}) \geq A_R.$$

C. A Coding Theorem for Abelian G-Codes

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

$$(\boldsymbol{h}_n)_{i,j} = \left\lceil \frac{Rn(1 - \alpha_{i,j})}{j \log p_i} \right\rceil.$$

Let $\mathcal{V}_{\boldsymbol{h}_n}$ be the Abelian group having spectrum σ^G and type $\boldsymbol{h}_n$. Consider a sequence of independent r.v.'s Φ_n uniformly distributed over $\mathrm{Hom}(G^n, \mathcal{V}_{\boldsymbol{h}_n})$. Let $\mathcal{C}_n := \ker(\Phi_n)$ be the corresponding sequence of random G-codes. We shall refer to such a random code construction as the G-coding ensemble of design rate R and splitting $\boldsymbol{\alpha}$. Notice that $\mathcal{C}_n$ has rate deterministically not smaller than R. Let $\overline{p_e(\mathcal{C}_n)}^{(R,\boldsymbol{\alpha})}$ denote the word error probability averaged over this ensemble. Theorem 7 admits the following generalization.

Theorem 22: Let $(G, \mathcal{Y}, W)$ be a G-symmetric MC. For every $R \in [0, \log |G|[$, $\boldsymbol{\alpha} \in \mathcal{P}(\boldsymbol{r}^G)$

$$\overline{p_e(\mathcal{C}_n)}^{(R,\boldsymbol{\alpha})} \leq \sum_{\substack{\boldsymbol{L} \neq \boldsymbol{0} \\ \boldsymbol{r}^G\text{-compatible}}} \exp\left(-nE_{\boldsymbol{L}}\left(R_{\boldsymbol{L}}\right)\right)$$

where $E_{\boldsymbol{L}}(R)$ is the error exponent of the MC $(G_{\boldsymbol{L}}, \mathcal{Y}, W)$, and

$$R_{\boldsymbol{L}} := R \sum_{1 \leq i \leq s} \sum_{1 \leq j \leq r_i} \frac{L_{i,j}}{j} \alpha_{i,j}.$$

By choosing $\boldsymbol{\alpha}^G \in \mathcal{P}(\boldsymbol{r}^G)$ such that

$$C_G = \min_{\substack{\boldsymbol{L} \neq \boldsymbol{0} \\ \boldsymbol{r}^G\text{-comp.}}} \frac{C_{\boldsymbol{L}}}{\sum\limits_{1 \leq i \leq s} \sum\limits_{1 \leq j \leq r_i^G} \frac{L_{i,j}}{j} \alpha_{i,j}}$$

one has that $\min_{\boldsymbol{L} \neq \boldsymbol{0}} E_{\boldsymbol{L}}(R_{\boldsymbol{L}}) > 0$ for all $R < C_G$. Therefore, Theorem 22 has the following corollary.

Corollary 23: Let $(G, \mathcal{Y}, W)$ be a G-symmetric MC of G-capacity C_G. Then, for every rate $0 < R < C_G$, there exists a G-code $\mathcal{C}$ of rate not smaller than R and arbitrarily low error probability.

Finally, for $0 < R < C$, it is possible to optimize the error exponent over all splittings $\boldsymbol{\alpha}$ in $\mathcal{P}(\boldsymbol{r}^G)$. This leads to the following definition of the G-coding error exponent of a MC $(G, \mathcal{Y}, W)$:

$$E_G(R) = \max_{\boldsymbol{\alpha} \in \mathcal{P}(\boldsymbol{r}^G)} \min_{\substack{\boldsymbol{L} \neq \boldsymbol{0} \\ \boldsymbol{r}^G\text{-comp.}}} E_{\boldsymbol{L}}\left(R \sum_{1 \leq i \leq s} \sum_{1 \leq j \leq r_i^G} \frac{L_{i,j}}{j} \alpha_{i,j}\right). \qquad (34)$$

By letting $\boldsymbol{\alpha}_G(R)$ in $\mathcal{P}(\boldsymbol{r}^G)$ be an optimal splitting in the maximization above, and using arguments similar to the proof of Corollary 8, the following corollary can be proved.

Corollary 24: Let $(G, \mathcal{Y}, W)$ be a G-symmetric MC of G-capacity C_G and G-coding exponent $E_G(R)$. Then, for all $0 < R < C_G$ we have

$$\liminf_n -\frac{1}{n}\log p_e(\mathcal{C}_n) \geq E_G(R)$$

with probability one over the G-coding ensemble of design rate R and optimal splitting $\boldsymbol{\alpha}^G(R)$.

D. Examples

In the sequel, three examples will be presented with explicit computations of C_G for Abelian groups G with particular algebraic structure. First we examine groups admitting Galois field structure, showing as in this case the G-capacity C_G coincides with the Shannon capacity C, as follows from classical linear-coding theory.

Example 6: Suppose that $G \simeq \mathbb{Z}_p^r$ for some prime p and positive integer r. Thus

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

Consequently, the only $\boldsymbol{r}^G$-compatible $\boldsymbol{L}$ is given by $\boldsymbol{L} = 1$ and therefore we have that in this case $C_G = C$, $E_{\mathbb{Z}_p^r}(R) = E(R)$. In other words, $\mathbb{Z}_p^r$-codes achieve both the capacity and the random-coding exponent of every $\mathbb{Z}_p^r$-symmetric MC. This had first been shown in [4]. In fact, in this case it is known that linear codes over the Galois field $\mathbb{F}_{p^r}$ suffice to achieve capacity random-coding exponent.

However, GU constellations admitting a generating group which is isomorphic to $\mathbb{Z}_p^r$ are affected by a constraint on their bandwidth efficiency. In fact, if S is a d-dimensional GU constellation admitting $\mathbb{Z}_p^r$ as a generating group, then standard arguments using group representation theory allow to conclude that

$$d \geq \begin{cases} r, & \text{if } p = 2 \\ 2r, & \text{if } p \geq 2. \end{cases} \qquad \square$$

In the next example, we show that when $G = \mathbb{Z}_{p^r}$, Definition 20 reduces to Definition 4 of Section III.

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

$$C_G = \min_{s=1,\dots,r} \frac{r}{s} C_s.$$

Notice first that in this case $\sigma^G = (p)$ and $\boldsymbol{r}^G = r$. A vector $\boldsymbol{L} = (L_1, \dots, L_r)$ is $\boldsymbol{r}^G$-compatible if and only if $L_j \leq j$ for every $j = 1, \dots, r$. Notice now that

$$G_{\boldsymbol{L}} = \sum_{1 \leq j \leq r} p^{j - L_j} G_{(p^j)} = \sum_{1 \leq j \leq r} p^{r - L_j}\mathbb{Z}_{p^r} = p^{r - L^*}\mathbb{Z}_{p^r}$$

where $L^* := \max_{1 \leq j \leq r} L_j$. Hence, $C_{\boldsymbol{L}} = C_{L^*}$.

Observe that $\mathcal{P}(\boldsymbol{r}^G)$ simply consists of the probability distributions $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_r)$. Suppose we are given some α in $\mathcal{P}(\boldsymbol{r}^G)$. We have that

$$\min_{\substack{\boldsymbol{L} \neq \boldsymbol{0} \\ \boldsymbol{r}^G\text{-comp.}}} \frac{C_{\boldsymbol{L}}}{\sum_{1 \leq j \leq r} \frac{L_j}{j}\alpha_j} = \min_{\rho=1}^{r} C_\rho \frac{1}{\max_{\substack{\boldsymbol{L} \neq \boldsymbol{0}\ \boldsymbol{r}^G\text{-comp.} \\ L^* = \rho}} \sum_{1 \leq j \leq r} \frac{L_j}{j}\alpha_j}.$$

Now

$$\max_{\substack{\boldsymbol{L} \neq \boldsymbol{0}\ \boldsymbol{r}^G\text{-comp.} \\ L^* = \rho}} \sum_{1 \leq j \leq r} \frac{L_j}{j}\alpha_j \geq \frac{\rho}{r}$$

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

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

Finally, the following example concerns one of the Cartesian product GU constellations introduced in Example 4.

Example 8: Consider the $K_{2\times 3}^\beta$ constellation introduced in Example 4 and an AWGN channel with input constrained to $K_{2\times 3}^\beta$. 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 $\boldsymbol{\alpha} \in \mathcal{P}(r^G)$ solution of the variational problem (33) defining $C_{\mathbb{Z}_6}$, as a function of the parameter 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 $\boldsymbol{r}^G = \left(r_1^G, r_2^G\right) = (1, 1)$. Equation (33) reduces to

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

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

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

Indeed, we have that

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

In Fig. 4, $\alpha_2^{\mathbb{Z}_6}(\beta)$ is plotted: notice how the optimal splitting follows the geometry of the constellation as $\alpha_2(\beta)$ is monotonically increasing in β with $\lim_{\beta \longrightarrow 0} \boldsymbol{\alpha}^{\mathbb{Z}_6}(\beta) = (0, 1)$ (as β goes to 0 $K_{2\times 3}(\beta)$ collapses onto constellation K_3) and $\lim_{\beta \longrightarrow +\infty} \boldsymbol{\alpha}^{\mathbb{Z}_6}(\beta) = (1, 0)$ (as β goes to $+\infty$ $K_{2\times 3}(\beta)$ collapses onto constellation 2-PAM). $\square$

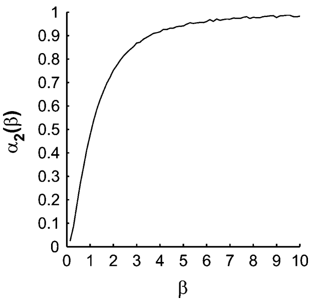

Fig. 4. The optimal splitting for the Cartesian product constellation $K^{\beta}_{2\times 3}$ as a function of β.

VII. Conclusion

In this paper, we have analyzed the information-theoretical limits of Abelian group codes over symmetric memoryless channels. Our results generalize the classical theory for binary-linear codes over binary-input symmetric-output channels. The main example we have in mind is the AWGN channel with input restricted over a geometrically uniform constellation S admitting G as a generating group and either soft or quantized output. We have characterized the threshold value for the rates at which reliable transmission is possible with G-codes, which we called the G-capacity C_G. The G-capacity is 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-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 input constrained to p^r-PSK constellations—with prime p–the G-capacity C_G coincides with the Shannon capacity C, so that in this case reliable transmission at any rate $R < C$ is in fact possible using $\mathbb{Z}_{p^r}$-codes.

Finally, we have exhibited a counterexample when $C_G < C$: it consists of the AWGN channel with as input a particular 3-D constellation admitting $\mathbb{Z}_{2^r}$ as a 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 the typical G-code;
- extending the theory to non-Abelian groups: indeed, it is known [6], [8] that GU constellations with Abelian generating group do not allow to achieve the unconstrained AWGN capacity.

Appendix A
Proof of Lemma 3

For the reader's convenience, all statements are repeated before their proof.

Lemma: Let G be a finite group, $(G, \mathcal{Y}, W)$ a G-symmetric MC, and $\mathcal{C} \subseteq G^n$ a code such that $1_{G^n} \in \mathcal{C}$. Then

$$p_e(\mathcal{C}|1_{G^n}) \le \frac{1}{|G|^n} \sum_{\boldsymbol{z}\in G^n} \int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}) \left(\sum_{\boldsymbol{\theta}\neq\delta_{1_G}} \frac{S_{\mathcal{C}}(\boldsymbol{\theta})}{\binom{n}{n\boldsymbol{\theta}}} \sum_{\boldsymbol{x}\in G^n_{\boldsymbol{\theta}}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}\boldsymbol{x}) \right)^{\rho} \mathrm{d}\boldsymbol{y}.$$

Proof: We start by recalling the Gallager bound [2]. Given an MC $(\mathcal{X}, \mathcal{Y}, W)$, and a code $\mathcal{C} \subseteq \mathcal{X}^n$, for every $\boldsymbol{x}$ in $\mathcal{C}$ and $\rho > 0$, the conditioned word error probability satisfies

$$p_e(\mathcal{C}|\boldsymbol{x}) \le \int_{\mathcal{Y}^n} W_n(\boldsymbol{y}|\boldsymbol{x})^{1/1+\rho} \left(\sum_{\boldsymbol{z}\in\mathcal{C}\setminus\{\boldsymbol{x}\}} W_n(\boldsymbol{y}|\boldsymbol{z})^{1/1+\rho} \right)^{\rho} \mathrm{d}\boldsymbol{y}.$$

From the given code $\mathcal{C}$ we generate the random code $\mathcal{C}' := \boldsymbol{Z}\Pi\mathcal{C}$, where Π is an r.v. uniformly distributed over the permutation group S_n (where $\pi \in S_n$ acts on $\boldsymbol{x} \in G^n$ by permuting its components, i.e., $(\pi\boldsymbol{x})_i := (\boldsymbol{x})_{\pi i}$) and $\boldsymbol{Z}$ is an r.v. uniformly distributed over G^n, independent from Π. Throughout the proof, we shall 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 code $\mathcal{C}'$ conditioned on the transmission of $\boldsymbol{Z}$ is equal to the word error probability of $\mathcal{C}$ conditioned on the transmission of 1_{G^n}. In fact, for every $\pi \in S_n$ we have that $1_{G^n} \in \pi\mathcal{C}$ and, since the channel is memoryless and stationary, the ML decision region $\Lambda_{\pi\mathcal{C}}$ for the codeword 1_{G^n} in the code $\pi\mathcal{C}$ coincides with $\pi\Lambda_{\mathcal{C}}$, where $\Lambda_{\mathcal{C}}$ denotes the ML decision region of 1_{G^n} in the code $\mathcal{C}$. Thus

$$\begin{aligned} p_e(\pi\mathcal{C}|1_{G^n}) &= 1 - \int_{\Lambda_{\pi\phi}} W_n(\boldsymbol{y}|1_{G^n})\mathrm{d}\boldsymbol{y} \\ &= 1 - \int_{\pi\Lambda_{\mathcal{C}}} W_n(\boldsymbol{y}|1_{G^n})\mathrm{d}\boldsymbol{y} \\ &= 1 - \int_{\Lambda_{\mathcal{C}}} W_n(\boldsymbol{y}|1_{G^n})\mathrm{d}\boldsymbol{y} = p_e(\mathcal{C}|1_{G^n}) \, . \end{aligned}$$

Similarly, for any $\boldsymbol{z} \in G^n$ we have $\boldsymbol{z} \in \boldsymbol{z}\mathcal{C}$ and, due to the G-symmetry of the channel, the ML decision region $\Lambda_{\boldsymbol{z}\mathcal{C}}$ of $\boldsymbol{z}$ in $\boldsymbol{z}\mathcal{C}$ coincides with $\boldsymbol{z}\Lambda_{\mathcal{C}}$, so that $p_e(\boldsymbol{z}\mathcal{C}|\boldsymbol{z}) = p_e(\mathcal{C}|1_{G^n})$. Therefore, we have

$$\mathbb{E}[p_e(\mathcal{C}'|\boldsymbol{Z})] = p_e(\mathcal{C}|1_{G^n}). \tag{35}$$

From (35), by applying the Gallager bound to each realization of the random code $\mathcal{C}'$, and observing that, for any $\boldsymbol{w} \in \mathcal{C}$, $\Pi\boldsymbol{w}$ is uniformly distributed over the set $G^n_{\boldsymbol{\theta}}$ of n-tuples of type $\boldsymbol{\theta}$ and independent from $\boldsymbol{Z}$, we get the equation at the top of the following page, with the summation index $\boldsymbol{w}$ running over $\mathcal{C}\setminus\{1_{G^n}\}$, $\boldsymbol{z}$ over G^n, $\boldsymbol{\theta}$ over $\mathcal{P}_n(G)\setminus\{\delta_{1_G}\}$, and $\boldsymbol{x}$ over $G^n_{\boldsymbol{\theta}}$. □

$$
\begin{aligned}
p_e(\mathcal{C}|1_{G^n}) &= \mathbb{E}[p_e(\mathcal{C}'|\boldsymbol{Z})]\\
&\leq \mathbb{E}\left[\int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{Z})\left(\sum_{\boldsymbol{w}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{Z}\Pi\boldsymbol{w})\right)^{\rho}\mathrm{d}\boldsymbol{y}\right]\\
&= \frac{1}{|G|^n}\sum_{\boldsymbol{z}}\int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\mathbb{E}\left(\sum_{\boldsymbol{w}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}\Pi\boldsymbol{w})\right)^{\rho}\mathrm{d}\boldsymbol{y}\\
&= \frac{1}{|G|^n}\sum_{\boldsymbol{z}}\int_{\mathcal{Y}^n} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z})\left(\sum_{\boldsymbol{\theta}}\frac{S_{\mathcal{C}}(\boldsymbol{\theta})}{\binom{n}{n\boldsymbol{\theta}}}\sum_{\boldsymbol{x}} W_n^{1/1+\rho}(\boldsymbol{y}|\boldsymbol{z}\boldsymbol{x})\right)^{\rho}\mathrm{d}\boldsymbol{y}
\end{aligned}
$$

APPENDIX B
PROOFS FOR SECTION IV

A. Proof of Lemma 11

Lemma: For every $1 \leq q < r$

$$
\begin{aligned}
H(W(\cdot|0)) &= H(\lambda_q) - q + \int_{\mathbb{C}} \lambda_q(y)H(\boldsymbol{\omega}_q(y))\mathrm{d}y\\
H(\lambda_q) &= H(\lambda_{q+1}) - 1 + \int_{\mathbb{C}} \lambda_{q+1}(y)H(\boldsymbol{\nu}_q(y))\mathrm{d}y.
\end{aligned}
$$

Proof: We have, for $\mathcal{K} := p^{r-q}\mathbb{Z}_{p^r}$

$$
\begin{aligned}
H\left(W(\cdot|0)\right) &= -\int_{\mathbb{C}} W(y|0)\log W(y|0)\mathrm{d}y\\
&= -\frac{1}{p^q}\sum_{k\in\mathcal{K}}\int_{\mathbb{C}} W(y\xi_{p^r}^k|0)\log W(y\xi_{p^r}^k|0)\mathrm{d}y\\
&= -\frac{1}{p^q}\sum_{k\in\mathcal{K}}\int_{\mathbb{C}} W(y|k)\log W(y|k)\mathrm{d}y\\
&= -\int_{\mathbb{C}} \lambda_q(y)\log\lambda_q(y)\mathrm{d}y\\
&\quad -\int_{\mathbb{C}} \lambda_q(y)\sum_{k\in\mathcal{K}}(\boldsymbol{\omega}_q(y))_k\log(p^q(\boldsymbol{\omega}_q(y))_k)\mathrm{d}y\\
&= H(\lambda_q) - q + \int_{\mathbb{C}} \lambda_q(y)H\left(\boldsymbol{\omega}_q(y)\right)\mathrm{d}y
\end{aligned}
$$

and

$$
\begin{aligned}
H(\lambda_q) &= -\int_{\mathbb{C}} \lambda_q(y)\log\lambda_q(y)\mathrm{d}y\\
&= -\frac{1}{p}\sum_{k\in\mathbb{Z}_p}\int_{\mathbb{C}} \lambda_q(y\xi_{p^{q+1}}^k)\log\lambda_q(y\xi_{p^{q+1}}^k)\mathrm{d}y\\
&= -\int_{\mathbb{C}} \frac{1}{p}\sum_{k\in\mathbb{Z}_p}\lambda_q(y\xi_{p^{q+1}}^k)\log\lambda_{q+1}(y)\mathrm{d}y\\
&\quad -\int_{\mathbb{C}} \lambda_{q+1}(y)\sum_{k\in\mathbb{Z}_p}\frac{\lambda_q(y\xi_{p^{q+1}}^k)}{p\lambda_{q+1}(y)}\\
&\qquad\cdot\log\frac{\lambda_q(y\xi_{p^{q+1}}^k)}{\lambda_{q+1}(y)}\mathrm{d}y\\
&= -\int_{\mathbb{C}} \lambda_{q+1}(y)\log\lambda_{q+1}(y)\\
&\quad -\int_{\mathbb{C}} \lambda_{q+1}(y)\sum_{k\in\mathbb{Z}_p}(\boldsymbol{\nu}_q(y))_k\log(p(\boldsymbol{\nu}_q(y))_k)\mathrm{d}y\\
&= H(\lambda_{q+1}) - 1 + \int_{\mathbb{C}} \lambda_{q+1}(y)H\left(\boldsymbol{\nu}_q(y)\right)\mathrm{d}y. \qquad \square
\end{aligned}
$$

B. Proof of Lemma 12

Lemma: For every $1 \leq q < r$ and $y \in \mathbb{C}$, there exists a partition

$$
\mathcal{W}_{q+1}(y) = \bigcup_{1\leq k\leq p^q}\mathcal{W}_q^k(y)
$$

where each multiset $\mathcal{W}_q^k(y) = \{w_{q,0}^k, w_{q,1}^k, \dots, w_{q,p-1}^k\}$ is such that, for all $0 \leq j, i \leq p-1$, $w_{q,j}^k$ belongs to $\mathcal{W}_q(y\xi_{p^{q+1}}^j)$, and

$$
0 \leq k < k' < p^q \implies w_{q,i}^k(y) \geq w_{q,j}^{k'}(y).
$$

Proof: Since the transition densities $W(y|x)$ are decreasing functions 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 $\{|y-\xi_{p^r}^x|, x \in p^{r-q-1}\mathbb{Z}_{p^r}\}$. Define $y = \rho e^{\theta i}$, $\varphi_j = j\frac{2\pi}{p^r}$ for $j \in p^{r-q-1}\mathbb{Z}_{p^r}$. Then

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

Let j^* be the closest input in $p^{r-q-1}\mathbb{Z}_{p^r}$ to the given output y, i.e., j^* is such that $|\theta-\varphi_{j^*}| \leq |\theta-\varphi_j|$ for all $j \in p^{r-q-1}\mathbb{Z}_{p^r}$. Then, either

$$
\varphi_{j^*} \leq \theta \leq \varphi_{j^*} + \frac{1}{2}\frac{2\pi}{p^{q+1}} \tag{36}
$$

or

$$
\varphi_{j^*} - \frac{1}{2}\frac{2\pi}{p^{q+1}} \leq \theta \leq \varphi_{j^*} \tag{37}
$$

hold true. Suppose that (36) holds true, and define $m := p^{r-q-1}$. Then

$$
\begin{aligned}
\cos(\theta-\varphi_{j^*}) &\geq \cos(\theta-\varphi_{j^*+1}) \geq \cos(\theta-\varphi_{j^*-1})\\
&\geq \cos(\theta-\varphi_{j^*+2}) \geq \cdots\\
&\geq \cos(\theta-\varphi_{j^*-\lfloor p^q/2\rfloor}).
\end{aligned}\tag{38}
$$

From (38) it follows that, for odd p

$$
\begin{aligned}
\mathcal{W}_q^0(y) = \Bigg\{ &W\left(y|j^*\right), W\left(y|j^*+m\right), W\left(y|j^*-m\right),\\
&\dots, W\left(y|j^*-\left\lfloor\frac{p}{2}\right\rfloor m\right)\Bigg\}
\end{aligned}
$$

$$\mathcal{W}_q^1(y) = \Big\{ W\left(y|j^* + \left\lceil \frac{p}{2} \right\rceil m\right), W\left(y|\left(j^* - \left\lceil \frac{p}{2} \right\rceil\right) m\right), \ldots, W\left(y|\left(j^* + p\right) m\right) \Big\}$$

$$\vdots$$

$$\mathcal{W}_q^{p^q-1}(y) = \Big\{ W\left(y|\left(j^* - \left(\left\lfloor \frac{p^q}{2} \right\rfloor + \left\lfloor \frac{p}{2} \right\rfloor\right) m\right)\right), \ldots, W\left(y|m\left(j^* - \left\lfloor \frac{p^q}{2} \right\rfloor\right)\right) \Big\}.$$

The claim follows, since for every k, $\mathcal{W}_q^k(y)$ contains exactly one $W(y|j)$ with j belonging to each coset of $p^{r-q}\mathbb{Z}_{p^r}$ in $p^{r-q-1}\mathbb{Z}_{p^r}$.

The case when (37) holds true instead of (36) is analogous, while the case $p = 2$ is much simpler. □

C. Proof of Lemma 14

For any subset $K \subseteq \mathbb{R}^n$, let $\mathrm{co}(K)$ denote the convex hull of K, i.e., the smallest convex subset of $\mathbb{R}^n$ containing K. A polytope is the convex hull of finite set $K \subset \mathbb{R}^n$. A general fundamental result (see [39]) states that $P \subset \mathbb{R}^n$ is a polytope if and only if it is bounded intersection of closed half-spaces. In the sequel, we shall deal with a special class of polytopes: given a point $\boldsymbol{x} \in \mathbb{R}^n$, we shall consider $co(S_n\boldsymbol{x})$, i.e., the convex hull of the set of all component permutations of $\boldsymbol{x}$. This is sometimes called the (generalized) permutahedron of $\boldsymbol{x}$. The next result explicitly characterizes $\mathrm{co}(S_n\boldsymbol{x})$ as the intersection of half-spaces.

Lemma 25: Let $\boldsymbol{w} \in \mathbb{R}^n$ be such that

$$w_1 \geq w_2 \geq \cdots \geq w_n. \tag{39}$$

Then $\mathrm{co}(S_n\boldsymbol{w}) = A$, where

$$A := \bigcap_J \left\{ \sum_{i \in J} x_i \leq \sum_{1 \leq i \leq |J|} w_i \right\} \bigcap \left\{ \sum_{1 \leq i \leq n} x_i = \sum_{1 \leq i \leq n} w_i \right\}.$$

Proof: In order to prove that $\mathrm{co}(S_n\boldsymbol{w}) \subseteq A$ it suffices to note that, for every $\sigma \in S_n$, $\sigma\boldsymbol{x} \in A$. In fact, it is easy to check that, due to (39), every constraint is satisfied. Since A is convex it immediately follows that $\mathrm{co}(S_n\boldsymbol{w}) \subseteq A$.

We now prove the converse inclusion, $A \subseteq \mathrm{co}(S_n\boldsymbol{w})$, by induction. The statement is trivially true for $n = 1$. Suppose that the claim is true for every $m \leq n$ for some $n \in \mathbb{N}$ and let $\boldsymbol{w} \in \mathbb{R}^{n+1}$ be such that $w_1 \geq \cdots \geq w_{n+1}$. Define

$$D := \left\{ \boldsymbol{x} \in \mathbb{R}^{n+1} : \sum_{1 \leq i \leq n+1} x_i = \sum_{1 \leq i \leq n+1} w_i \right\}.$$

For each $J \subset \{1, \ldots, n+1\}$, define $D_J, F_J \subseteq \mathbb{R}^{n+1}$ by $D_J := \left\{\boldsymbol{x} : \sum_{i \in J} x_i = \sum_{1 \leq i \leq |J|} w_i\right\}$ and, respectively, $F_J := \left\{\boldsymbol{x} : \sum_{i \in J} x_i = \sum_{n-|J| < i \leq n+1} w_i\right\}$. Consider the facet

$$A_J := D \bigcap D_J \bigcap_{I \neq J} \left\{ \boldsymbol{x} : \sum_{i \in I} x_i \leq \sum_{1 \leq i \leq |I|} w_i \right\}.$$

We observe that

$$\pi_J A_J \subseteq B_J, \quad \pi_{J^c} A_J \subseteq C_J \tag{40}$$

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

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

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

In fact, the former inclusion in (40) is trivial since B_J is defined as the intersection of a subset of the half-spaces whose intersection defines A_J, while for the latter it suffices to observe that, for each $I \subset J^c$, if $\boldsymbol{x}$ is in A_J, then

$$\sum_{i \in I \cup J} x_i \leq \sum_{i=1}^{|I|+|J|} x_i, \quad \sum_{i \in J} x_i = \sum_{i=1}^{|J|} x_i$$

so that

$$\sum_{i \in I} x_i = \sum_{i \in I \cup J} x_i - \sum_{i \in J} x_i \leq \sum_{i=|I|+1}^{|I|+|J|} x_i.$$

For $J \subseteq \{1, \ldots, n+1\}$, let $\Psi_J \in S_{n+1}$ be any permutation mapping the first $|J|$ elements onto J, i.e., such that $\Psi_J(\{1, \ldots, |J|\}) = J$. Define $S_J \subseteq S_{n+1}$ to be the set of permutations σ such that $\sigma\big|_{\{1,\ldots,|J|\}}$ is the identity. Notice that S_J commutes with S_{J^c} in the sense that $\sigma\rho = \rho\sigma$, for all $\sigma \in S_J$ and $\rho \in S_{J^c}$. Let $\phi_J : \pi_J\mathbb{R}^{n+1} \to \mathbb{R}^{|J|}$ and $\phi_{J^c} : \pi_{J^c}\mathbb{R}^{n+1} \to \mathbb{R}^{|J^c|}$ be the standard isomorphisms. By applying the inductive hypothesis to $\phi_J\pi_J\Psi_J\boldsymbol{w}$ and $\phi_{J^c}\pi_{J^c}\Psi_J\boldsymbol{w}$, respectively, and then immersing back the results in $\mathbb{R}^{n+1}$ by ϕ_J^{-1} and $\phi_{J^c}^{-1}$, respectively, we have that

$$B_J \subseteq \mathrm{co}(\pi_J\Psi_J S_J\boldsymbol{w}), \quad C_J \subseteq \mathrm{co}(\pi_{J^c}\Psi_J S_{J^c}\boldsymbol{w}). \tag{41}$$

For every $\boldsymbol{x} \in A_J$ we have $\pi_J\boldsymbol{x} \in B_J$ and $\pi_{J^c}\boldsymbol{x} \in C_J$ from (40). Then (41) implies that $\lambda' \in \mathcal{P}(S_J)$ and $\lambda'' \in \mathcal{P}(S_{J^c})$ exist such that

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

with $\lambda \in \mathcal{P}(\Psi_J S_J S_{J^c}) \subseteq P(S_{n+1})$ defined by $\lambda(\Psi_J\sigma\rho) := \lambda'(\sigma)\lambda''(\rho)$. Therefore, for every $J \subset \{1, \ldots, n+1\}$, we have $A_J \subseteq co(S_{n+1}\boldsymbol{w})$, and so $A = \mathrm{co}\left(\bigcup_J A_J\right) \subseteq \mathrm{co}(S_{n+1}\boldsymbol{w})$. □

Lemma 26: Suppose n^2 real numbers $\{a_i^k, 1 \leq i, k \leq n\}$ are given, such that

$$k < k' \Longrightarrow a_j^k \leq a_i^{k'}, \quad 1 \leq j, i \leq n. \tag{42}$$

Define $\boldsymbol{x}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$

$$\boldsymbol{x} := \left(\sum_{1 \le i \le n} a_i^1, \dots, \sum_{1 \le i \le n} a_i^n \right)$$
$$\boldsymbol{v} := \left(\sum_{1 \le k \le n} a_1^k, \dots, \sum_{1 \le k \le n} a_n^k \right).$$

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

Proof: Equation (42) implies that $x_1 \ge x_2 \ge \dots \ge x_n$, and, for every $J \subset \{1, \dots, n\}$, $\sum_{i \in J} v_i \le \sum_{1 \le i \le |J|} x_i$. Then, Lemma 25 can be applied to show that $\boldsymbol{v} \in co(S_n \boldsymbol{x})$. □

We can now prove Lemma 14.

Lemma: For every $1 \le \alpha < q < r$ and $y \in \mathbb{C}$, we have

$$H \left(\sum_{j \in \mathbb{Z}_p} \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} \boldsymbol{\delta}_q^\alpha(y, j) \right) \le H(\boldsymbol{\nu}_q(y)).$$

Proof: We shall show that

$$\boldsymbol{\nu}_q(y) \in \mathrm{co} \left(S_p \left(\sum_{j \in \mathbb{Z}_p} \frac{\lambda_q(y \xi_{p^{q+1}}^j)}{p \lambda_{q+1}(y)} \boldsymbol{\delta}_q^\alpha(y, j) \right) \right). \tag{43}$$

Then, the claim will simply follow from the concavity of the entropy function.

Let us consider the quantities $w_{q,j}^k$ introduced in Lemma 12. For all $0 \le j, s < p$, define

$$a_j^s := \sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_{q,j}^{sp^\alpha + h + \tilde{h} p^{\alpha+1}}.$$

From (24) it follows that

$$\sum_{0 \le j < p} p^q \lambda_q(y \xi_{p^{q+1}}^j) \left[\boldsymbol{\delta}_q^\alpha(y, j) \right](s) = \sum_{0 \le j < p} a_j^s$$

while, from (16), we have

$$p^{q+1} \lambda_{q+1}(y) \left[\boldsymbol{\nu}_q(y) \right](j) = \sum_{0 \le s < p} a_j^s.$$

Fix a pair $0 \le k < k \le p-1$: from (23) we have

$$w_{q,j}^{kp^\alpha + h + \tilde{h} p^{\alpha+1}} \ge w_{q,i}^{k' p^\alpha + h + \tilde{h} p^{\alpha+1}}$$

for every $0 \le j, i < p$, $0 \le h < p^{\alpha-1}$, $0 \le \tilde{h} < p^{q-\alpha-1}$. Thus

$$\begin{aligned} a_j^k &= \sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_{q,j}^{kp^\alpha + h + \tilde{h} p^{\alpha+1}}(y) \\ &\le \sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_{q,i}^{k' p^\alpha + h + \tilde{h} p^{\alpha+1}}(y) \\ &= a_i^{k'}. \end{aligned}$$

Therefore, the coefficients $\{a_j^k, 0 \le j, k < p\}$ satisfy (42) and then Lemma 26 can be applied to conclude that

$$p^{q+1} \lambda_{q+1}(y) \boldsymbol{\nu}_q(y) \in \mathrm{co} \left(S_p \left(p^q \sum_{j \in \mathbb{Z}_p} \lambda_q(y \xi_{p^{q+1}}^j) \boldsymbol{\delta}_q^\alpha(y, j) \right) \right)$$

which in turn implies (43). □

Appendix C
Proof of Proposition 18

Proposition: For every finite variance $\sigma^2 > 0$ and any integer $r \ge 2$, the family of $K_{2^r}^\beta$-AWGN channels satisfies

$$\lim_{\beta \longrightarrow \infty} C_{2^r}(\beta) = C(\infty) > 0, \qquad \lim_{\beta \longrightarrow \infty} C_{\mathbb{Z}_{2^r}}(\beta) = 0.$$

Proof: We start by observing that, for every $y \in \mathbb{R}^3$,

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

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

$$\frac{1}{2^r} \sum_{x \in K_{2^r}^\beta} \int_{\mathbb{R}^3} W(y|x) \frac{\log e}{2\sigma^2} \left(||y-x||^2 + 8 \right) \mathrm{d}y < +\infty.$$

Lebesgue's dominated convergence theorem can be applied (see [32]) in order to exchange the order of the limit and the integral in evaluating the expressions $\lim_{\beta \longrightarrow +\infty} C_{2^s(\beta)}$ for any $s \le r$. By this argument and the continuity of transition densities $W(y|x)$, we get, in the limit $\beta \longrightarrow +\infty$

$$\begin{aligned} \lim C_{2^r}(\beta) &= \lim \sum_{x \in K_{2^r}^\beta} \frac{1}{2^r} \int_{\mathbb{R}^3} W(y|x) \log \frac{W(y|x)}{\frac{1}{2^r} \sum_z W(y|z)} \mathrm{d}y \\ &= \int_{\mathbb{R}^3} \frac{1}{2} \sum_{x \in K^\infty} W(y|x) \log \frac{W(y|x)}{\frac{1}{2} \sum_z W(y|z)} \mathrm{d}y \\ &= C(\infty). \end{aligned}$$

Similarly, for every $1 \leq s < r$, define $\mathcal{K} := 2^{r-s}\mathbb{Z}_{2^r}$. In the limit $\beta \longrightarrow +\infty$ we get

$$\begin{aligned}\lim C_{2^s}(\beta) &= \lim \int_{\mathbb{R}^3} 2^{-s} \sum_{j \in \mathcal{K}} W(y|x_j) \\ &\quad \times \log \frac{W(y|x_j)}{2^{-s}\sum_{k\in\mathcal{K}} W(y|x_k)} \mathrm{d}y \\ &= \int_{\mathbb{R}^3} W(y|(0,1)) \log\left(\frac{W(y|(0,1))}{W(y|(0,1))}\right) \mathrm{d}y \\ &= 0. \qquad \square\end{aligned}$$

Acknowledgment

The authors wish to thank the Associate Editor and one of the referees for many detailed comments on the paper.

References

[1] P. Elias, "Coding for two noisy channels," in *Proc. 3rd London Symp. Information Theory*, London, U.K., 1955, pp. 61–76.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[3] A. J. Viterbi and J. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[4] R. L. Dobrushin, "Asymptotic optimality of group and systematic codes for some channels," *Theor. Probab. Appl.*, vol. 8, pp. 47–59, 1963.

[5] G. D. Forney , Jr, "Geometrically uniform codes," *IEEE Trans. Inf. Theory*, vol. 37, no. 5, pp. 1241–1260, Sep. 1991.

[6] H.-A. Loeliger, "Signal sets matched to groups," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1675–1679, Nov. 1991.

[7] D. Slepian, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, 1968.

[8] I. Ingemarsson, "Commutative group codes for the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 215–219, Mar. 1973.

[9] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 55–67, Jan. 1982.

[10] G. D. Forney , Jr and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams and canonical encoders," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1491–1513, Sep. 1993.

[11] S. Benedetto, R. Garello, M. Mondin, and G. Montorsi, "Geometrically uniform partitions of $L \times \mathrm{MPSK}$ constellations and related binary trellis codes," *IEEE Trans. Inf. Theory*, vol. 39, no. 6, pp. 1773–1798, Nov. 1993.

[12] S. Benedetto, R. Garello, M. Mondin, and G. Montorsi, "Geometrically uniform TCM codes based on $L \times \mathrm{MPSK}$ constellations," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 137–152, Jan. 1994.

[13] H.-A. Loeliger, G. D. Forney, T. Mittelholzer, and M. D. Trott, "Minimality and observability of group systems," *Linear Alg. its Applic.*, vol. 205-206, pp. 937–963, 1994.

[14] G. Caire and E. Biglieri, "Linear block codes over cyclic groups," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1246–1256, Sep. 1995.

[15] H.-A. Loeliger and T. Mittelholzer, "Convolutional codes over groups," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1660–1686, Nov. 1996.

[16] F. Fagnani and S. Zampieri, "Minimal syndrome formers for group codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 1–31, Jan. 1998.

[17] R. Johannesson, Z.-X. Wan, and E. Wittenmark, "Some structural properties of convolutional codes over rings," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 839–845, Mar. 1998.

[18] F. Fagnani and S. Zampieri, "System theoretic properties of convolutional codes over rings," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2256–2274, Sep. 2001.

[19] F. Fagnani and S. Zampieri, "Minimal and systematic convolutional codes over finite Abelian groups," *Linear Alg. its Applic.*, vol. 378, pp. 31–59, 2004.

[20] G. D. Forney , Jr and M. D. Trott, "The dynamics of group codes: Dual Abelian group codes and systems," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 2935–2965, Dec. 2004.

[21] R. Garello, G. Montorsi, S. Benedetto, D. Divsalar, and F. Pollara, "Labelings and encoders with the uniform bit error property with applications to serially concatenated trellis codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 123–136, Jan. 2002.

[22] F. Fagnani and F. Garin, "Analysis of serial concatenation schemes for non-binary constellations," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, SA, Australia, 2005, pp. 745–749.

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

[24] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 417–438, Mar. 2004.

[25] U. Erez and G. Miller, "The ML decoding performance of LDPC ensembles over $\mathbb{Z}_q$," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1871–1879, May 2005.

[26] D. Sridhara and T. E. Fuja, "LDPC codes over rings for PSK constellation," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3209–3220, Sep. 2005.

[27] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[28] G. Miller and D. Burshtein, "Bounds on the maximum likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2696–2710, Nov. 2001.

[29] G. Como and F. Fagnani, "Average type spectra and minimum distances of LDPC codes over Abelian groups," *SIAM J. Discr. Math.*, vol. 23, no. 1, pp. 19–53, Oct. 2008.

[30] G. Como, "Group codes outperform binary coset codes on non-binary symmetric memoryless channels," 2008 [Online]. Available: http://mit.edu/giacomo/www/material/gexp2, submitted for publication

[31] G. Como and F. Fagnani, "Ensembles of codes over Abelian groups," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, SA, Australia, Sep. 2005, pp. 1788–1792.

[32] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1966.

[33] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2101–2104, Sep. 1999.

[34] D. Slepian, "On neighbor distances and symmetry in group codes," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 5, pp. 630–632, Sep. 1971.

[35] A. Barg and G. D. Forney , Jr, "Random codes: Minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sep. 2002.

[36] R. G. Gallager, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 244–246, Mar. 1973.

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[38] T. W. Hungerford, *Algebra*. New York: Springer Verlag, 1974.

[39] M. Ziegler, *Lecture Notes on Polytopes*. New York: Springer, 1995.

Giacomo Como received the B.Sc., M.S., and Ph.D. degrees in applied mathematics from Politecnico di Torino, Torino, Italy, in 2002, 2004, and 2008, respectively.

During 2006-2007 he was Visiting Assistant in Research at the Department of Electrical Engineering, Yale University, New Haven, CT. He is currently a Postdoctoral Associate at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA. His current research interests include coding theory, information theory, distributed estimation, and control.

Fabio Fagnani received the Laurea degree in mathematics from the University of Pisa, Pisa, Italy, and Scuola Normale Superiore, Pisa, in 1986, and the Ph.D. degree in mathematics from the University of Groningen, Groningen, The Netherlands, in 1991.

During 1991–1998, he has been an Assistant Professor at the Scuola Normale Superiore. Since 1998, he has been with the Politecnico di Torino, Torino, Italy, where he is currently Full Professor of Mathematical Analysis. His research activities are on the fundamental mathematical aspects of systems and control theory and of coding theory on which he is author of about 30 papers on international journals. Specific themes of current interest are: control under communication constraints and coordinated control and their connection with graph theory and symbolic dynamics; inverse problems and recursive deconvolution techniques; codes over groups and their use in high-performance schemes.