POLITECNICO DI TORINO

SCUOLA DI DOTTORATO
Dottorato in Matematica per le Scienze dell'Ingegneria – XX ciclo
Settore scientifico-disciplinare: MAT/05 ANALISI MATEMATICA

Ph.D. Thesis

Ensembles of codes over Abelian groups

Giacomo Como

Advisor	Co-advisor	Director of the Ph.D. program
prof. F. Fagnani	prof. A. Bacciotti	prof. N. Bellomo

March 2008

Ensembles of codes over Abelian groups

Giacomo Como

March 10, 2008

Abstract

Group codes matched to geometrically uniform signal sets allow to transmit at higher spectral efficiency while inheriting many of the structural properties enjoyed by binary linear codes. In this thesis the information-theoretical limits of Abelian group codes are analyzed.

The capacity achievable by Abelian group codes over symmetric channels is characterized. For many important examples, like the Gaussian channel with the m-PSK modulation as input, Abelian group codes are shown to achieve Shannon capacity, as it is well-known to be the case for binary-linear codes over binary-symmetric channels. A counterexample is presented, based on a three dimensional signal set, for which instead, despite its group symmetry, the use of Abelian group codes leads to a loss in capacity.

The problem of characterizing the minimum Bhattacharyya distance of the typical Abelian group code is addressed. For the Gaussian channel with 8-PSK as input it is shown that the typical cyclic group code asymptotically meets the Gilbert-Varshamov bound, while the typical random code does not. This generalizes a result known for binary linear codes. It is also shown that a random binary affine code is bounded away from the Gilbert-Varshamov bound of this channel with probability one. Similar results can be inferred for the typical error exponent. This shows that not only group codes matching the symmetry of the channel cause no loss in capacity, but they can guarantee better performance.

Structural properties of low-density parity-check (LDPC) codes over finite Abelian groups are studied. Two ensembles of regular LDPC codes over the cyclic group $\mathbb{Z}_m$ are analyzed. In the first one the non-zero entries of the parity matrix are all equal to 1; in the second one they are randomly chosen, independently and uniformly, from the set of units of $\mathbb{Z}_m$. Precise combinatorial results are established for the exponential growth rate of their type-enumerating functions with respect to the code-length. Minimum Bhattacharyya distance properties are analyzed when such codes are employed over a $\mathbb{Z}_m$-symmetric transmission channel. In particular, in both cases minimum distances are shown to grow linearly in the code-length with probability one, and lower bounds are provided for the typical normalized minimum distance. Numerical results are presented indicating that the second ensemble definitely outperforms the first. Generalizations to LDPC codes over finite Abelian groups are also discussed.

The main topics left for future research consist in extending the theory to non-Abelian group codes and analyzing the performance of LDPC codes over Abelian groups with message-passing decoding.

Aknowledgements

I would like to thank my advisor Fabio Fagnani for all his encouragement, advisement and enthusiasm during the four years I have been working with him.
I would like to thank Andrea Bacciotti for the freedom he gave me in choosing the subject of this thesis, and Paolo Tilli for many stimulating discussions.
Special thanks go to my friends at Politecnico, and in particular to Federica Garin for the many exchanges of ideas we had throughout the PhD.
I thank Sekhar Tatikonda for inviting me as a visiting student at Yale University, where I had the great opportunity to work with him and Serdar Yuksel. Special thanks go to my officemates Ming Cao and Jian Ni.

Finally, thank you to Paola, Alberto, Irene, Aciredef, and to all my friends in Torino, L'Aquila and New Haven.

Contents

Chapter 1

Introduction

This dissertation deals with the analysis and design of transmission codes with Abelian group structure. The motivation comes from communication engineering, specifically the problem of reliable transmission of digital data over a bandwidth-limited noisy channel.

This is a classic problem of information theory dating back to Shannon's seminal work [60]. Shannon proved that for every transmission channel there exists a threshold C, called the capacity of the channel, such that arbitrarily reliable communication is possible at any rate below C, and conversely reliable transmission is not possible at rates above C. However, for almost fifty years Shannon's theoretical limits remained practically unreachable because of the unaffordable complexity of capacity-achieving coding schemes.

A major breakthrough in the discipline came in the '90s with the introduction of turbo codes [8] and the rediscovery of Gallager's low-density parity-check (LDPC) codes [30], which for the first time made it possible to achieve Shannon's theoretical limits in practice. Both these high-performance schemes are based on binary linear codes (i.e. linear subspaces over the binary field $\mathbb{Z}_2$) admitting a sparse graphical representation which allows them to be decoded using a low-complexity message-passing algorithm, known as belief propagation [52]. In particular LDPC codes are obtained as kernels of sparse binary matrices, i.e. binary matrices containing a limited amount of non-zero entries both in each row and in each column. To any LDPC matrix a sparse factor graph is associated, over which the message-passing decoding algorithm is implemented.

In this thesis, extensions of the theory of LDPC codes to the framework of group codes will be considered. Group codes are codes that have a group property under a componentwise group operation. They allow to use non-binary, highly spectral-efficient, geometrically uniform modulations, while inheriting many of the nice structural properties enjoyed by binary linear codes. The first part of the thesis is devoted to a fundamental investigation of the information-theoretical limits of Abelian group codes, with no

constraints on the density of their kernel (syndrome) representation. In the second part, structural properties of LDPC codes over arbitrary Abelian groups are investigated. Studying information-theoretical limits of Abelian group codes is propaedeutic to the analysis of LDPC codes over Abelian group. Indeed, this approach makes it possible to distinguish between the possible limitation in performance due to the group structure and the one due to the sparseness of their graphical representation.

1.1 Ensembles of codes over Abelian groups

Group codes were first introduced by Slepian [64] as extensions of binary linear codes [63]. A prototypical example comes from the m-PSK Gaussian channel. This is a channel accepting as possible input any element in the set m-PSK, consisting of all the m-ary complex roots of the unity; the received output is obtained by adding a homogeneous, zero-mean, two-dimensional Gaussian variable. By considering the natural labeling $\lambda : \mathbb{Z}_m \to m$-PSK, with $\lambda(l) = e^{\frac{l}{m}2\pi i}$, any subgroup $\mathcal{C} \leq \mathbb{Z}_m^N$ yields, through λ, a code over m-PSK. Such a code (as well as the associated subgroup) is called a $\mathbb{Z}_m$-code.

All this construction can be generalized to a broader family of memoryless transmission channels which are symmetric with respect to the action of a finite, possibly non-Abelian, group G. In this case a group code over G, briefly G-code, is any subgroup of the direct group product G^N. Group codes have complete symmetry, and as a consequence they have congruent Voronoi regions and invariant distance profiles, and they enjoy the uniform error property, i.e. independence of the error probability on the transmitted codeword. Structural properties of group codes have been extensively studied during the '80s and the '90s using the theory of behavioral group systems: minimality of state representation, existence of feedback-free encoders and syndrome formers -see [29] and references therein. In particular it is known that Abelian group codes admit both homomorphic encoders and syndrome-formers.

Recently, group codes have made their appearance also in the context of turbo codes [33, 20, 21, 22] and of LDPC codes [6, 66]. LDPC codes over a finite Abelian group G, briefly LDPC G-codes are subgroups of G^N admitting sparse syndrome representation. In the cyclic case $G \simeq \mathbb{Z}_m$ this means that we are considering codes

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{Z}_m^N \mid \Phi\boldsymbol{x} = 0\},$$

where Φ in $\mathbb{Z}_m^{N\times L}$ is a matrix containing only a linear (in N) amount of non-zero entries. LDPC codes over non-binary alphabets were first introduced in Gallager's seminal work [30, Sect.5], then more recently studied in [14, 6, 66, 7, 19]. However, almost all the results available in the literature are limited to LDPC codes over finite fields, while in this thesis a theory for LDPC codes over arbitrary finite Abelian groups will be developed.

A fundamental issue characterizing information theory since its beginning is the use of the probabilistic method [1]. In order to prove the existence of a coding scheme with certain properties, a probability space is constructed (code ensemble) and then it is shown that a randomly chosen code from this space satisfies the desired properties with positive probability. The probabilistic method was first used by Erdős [18] in graph theory. It was Shannon who introduced it in information theory in order to prove coding theorems [60]. He introduced the random coding ensemble, essentially consisting in the set of all codes of a given rate equipped with the unform probability, and showed that whenever the design rate is below capacity the average error probability vanishes in the limit of large block-lengths.

Rather than being an existence proof technique only (often misrepresented as 'non-constructive'), in modern coding theory [57] the probabilistic method is exploited as a fundamental design tool as well. When a code ensemble can be shown to attain some desired performance (asymptotically) almost surely, then a way to construct a coding scheme simply consists in randomly generating it accordingly to the code ensemble distribution. Different code ensembles can be compared in terms of their average or almost-sure performance, and design criteria can be optimized. In particular low-complexity code constructions are obtained randomly generating their sparse graphical representations. Most of the results of this thesis concern the performance of code ensembles with Abelian group structure.

1.2 Summary of the thesis

Chapter 2

In this chapter, all notation is introduced and Shannon classical coding theory for memoryless channel is summarized. The class of symmetric memoryless channels is introduced, the main example consisting in the AWGN channel with input constrained on a geometrically uniform constellation. The Gilbert-Varshamov bound for the minimum Bhattacharyya distance on symmetric memoryless channels is presented. Finally, group codes are introduced, as well as type-enumerating functions.

Chapter 3

In this chapter the theory of linear codes over binary symmetric channels is extended to group codes over non-binary symmetric channels. When a finite group G does admit Galois field structure (i.e. when it is isomorphic to $\mathbb{Z}_p^r$ for some prime p and positive integer r), it is a well known result of classical information theory [17, 31] that G-codes (and in fact linear codes over the Galois field $\mathbb{F}_{p^r}$) allow to achieve Shannon capacity and average error exponent of a symmetric memoryless channel. We address the question

whether the same holds true in the more general context of group codes: a result in this sense was conjectured by Loeliger in [44].

We solve this problem for a generic finite Abelian group G, showing that classical information-theoretic results generalize in a nontrivial way. A new concept of capacity is introduced, which we called G-capacity: it conveys information about Shannon capacities of subchannels associated to subgroups of G, and it is shown to be exactly the information theoretical limit achievable by G-codes. Examples are presented showing that in some cases G-capacity and Shannon capacity coincide while in other cases the former is strictly less than the latter. In particular, for the m-PSK Gaussian channel the $\mathbb{Z}_m$-capacity coincides with the Shannon capacity: therefore, in this case $\mathbb{Z}_m$-codes allow to achieve capacity. Average error exponents are obtained as well; it is shown that even when the use of Abelian group codes does not cause a loss in the achievable capacity it does lower the average error exponent at low rates. Extension of the theory to non-Abelian group codes has been left for future research: however, some results available in the literature for non-Abelian group codes [27, 49, 39] seem to indicate that the group product structure might not be the optimal choice for non-Abelian groups.

The material presented in this chapter is partially based on the following papers:

- G. Como, F. Fagnani, "Ensembles of Codes over Abelian Groups", in Proceedings of ISIT 2005 (Adelaide, SA, Australia), pp. 1788-1792, 5-9 Sept. 2005;
- G. Como F. Fagnani, "The capacity of Abelian group codes over symmetric channels", submitted to *IEEE Trans. Inform. Theory*, 2005, av. at http://calvino.polito.it/ricerca/2005/pdf/33_2005/art_33_2005.pdf.

Chapter 4

Beyond the capacity achievability problem, a fundamental question arising is whether, given a memoryless transmission channel exhibiting certain symmetries, designing codes matching these symmetries guarantees a gain with respect to designing codes without taking these symmetries into account. This question has been addressed in the information theory literature only for binary-input channels. In this case it is known that not only are binary linear codes appealing because of their nice algebraic structure and symmetries, but they also outperform nonlinear codes over binary symmetric channels. In fact random binary linear codes are known to meet with probability one the celebrated Gilbert-Varshamov (GV) [30] lower bound on the minimum distance and the so-called expurgated error exponent [4] . On the contrary, a binary codes sequence generated randomly with no linearity constraints can be shown not to achieve the GV bound with probability one. How this phenomenon generalizes to non-binary-input symmetric channels and group codes?

We focus on a special case, the 8-PSK AWGN channel, containing most of the key ingredients of the general situation. We analyze three different code ensembles all of which are capacity achieving: the random coding ensemble, i.e. the set of all possible codes (with no algebraic structure requirement), the $\mathbb{Z}_8$-code ensemble consisting in the set of all subgroups of $\mathbb{Z}_8^N$, and the binary affine code ensemble consisting in the set of all codes which are affine subspaces of $\mathbb{Z}_2^{3N}$. While, analogously to the binary case, the random coding ensemble does not asymptotically achieve the GV bound with probability one, we prove that almost surely a random $\mathbb{Z}_8$-group code sequence achieves the GV bound. We also show that almost surely a sequence of binary affine codes has minimum distance asymptotically bounded away from the GV distance. Similar results can be obtained for the error exponent which (at low rates) is larger for a typical $\mathbb{Z}_8$-code sequence than it is for a typical binary affine code sequence or for a typical code sequence sampled from the random coding ensemble. This stands in contrast with the results obtained for the average error exponent, which is larger for the random coding ensemble and for the binary affine ensemble than it is for the $\mathbb{Z}_8$-group code ensemble: hierarchies are reversed! The paradox can be explained by the fact that the average case analysis only gives a one side estimation of the performance of a typical code (thanks to Markov inequality). Ensemble performance may fail to concentrate around its expected value, and in this case the average case analysis ends up to be too conservative in estimating the error exponent.

The material presented in this chapter is partially based on the following papers:

- G. Como, F. Fagnani, "On the Gilbert-Varshamov distance of Abelian group codes", in Proceedings of ISIT 2007 (Nice, France), pp., 26-30 June 2007;
- G. Como, F. Fagnani, "The outperformance of group codes over non-binary symmetric channels: minimum distances", in preparation, 2008.

Chapter 5

The standard way to construct LDPC codes over a finite Abelian group G consists in generating a random regular hypergraph with N nodes of a given degree c and $L = Nc/d$ hyperedges of degree d, and to associate to each hyperedge a homomorphism from G^d to G. Sparseness is then enforced by considering the limit properties as N and L tend to infinity while c and d are kept constant. While the optimization of the degrees c and d has been widely studied in the literature of binary LDPC codes (more in general degree profiles for irregular ensembles are considered), the way to associate local homomorphisms to the hyperedges is a peculiar design parameter of non-binary LDPC codes.

We analyze structural properties of ensembles of regular LDPC codes over an Abelian group G. The non-zero entries of the parity matrix are randomly chosen, independently

and uniformly, from an arbitrary label group F of automorphisms of G. The two extreme cases are $F = \{1\}$ -called the unlabelled ensemble-, and $F = \mathrm{Aut}(G)$ -called the uniformly labeled ensemble. We study in full detail average type-spectra and minimum Bhattacharyya distances of the LDPC ensembles introduced above. We show that minimum distances grow linearly in N with probability one, and we obtain almost sure lower bounds on the asymptotic normalized minimum distance of the two LDPC ensembles. Finally, we present some numerical results for the average distance-spectra clearly indicating that the distance properties of the uniformly labelled ensemble are much better than those of the unlabelled ensembles. The material presented in this chapter is partially based on the following papers:

- G. Como, F. Fagnani, "Ensembles of Codes over Abelian Groups", in Proceedings of ISIT 2005 (Adelaide, SA, Australia), pp. 1788-1792, 5-9 Sept. 2005;
- G. Como, F. Fagnani, "Average spectra and minimum distances of low-density parity-check codes over cyclic groups", submitted to SIAM Journal on Discrete Mathematics, 2007;

Chapter 2

Memoryless symmetric channels and group codes

In this chapter, all notation is introduced. Shannon classical coding theory for memoryless channel is summarized in Sect.2.2. In Sect. 2.3 the class of symmetric memoryless channels is introduced, the main example consisting in the AWGN channel with input constrained on a geometrically uniform constellation. In Sect. 2.4 the Gilbert-Varshamov bound for the minimum Bhattacharyya distance on symmetric memoryless channels is presented. Finally, in Sect.2.5 group codes are introduced, as well as type-enumerating functions.

2.1 Notation

Throughout this dissertation $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ will denote the usual number sets. With $\mathbb{R}^+ := [0, +\infty)$ and $\mathbb{R}_+ := (0, +\infty)$ we will indicate the sets respectively of nonnegative and positive reals. If z is in $\mathbb{C}$, z^* is its conjugate. The functions log and exp are to be considered with respect to a fixed, arbitrarily chosen positive base, unless explicit mention to the contrary. Conventionally, $\exp(-\infty) = 0$, $\exp(+\infty) = +\infty$, $\inf(\emptyset) = +\infty$, $\sup(\emptyset) = -\infty$. For any subset $B \subseteq A$, $\overline{B} := A \setminus B$ will denote the complementary of B in A, while $\mathbb{1}_B : A \to \{0, 1\}$ will denote the indicator function of B, defined by $\mathbb{1}_B(a) = 1$ if a belongs to B, $\mathbb{1}_B(a) = 0$ otherwise.

Let $\mathcal{A} = (A, \mathcal{B}, \nu)$ be a σ-finite measure space [58]. As usual $L^1(\mathcal{A})$ will denote the space of (equivalence classes of) absolutely integrable functions $f : A \to \mathbb{R}$, and $\mathcal{P}(\mathcal{A}) \subseteq L^1(\mathcal{A})$ the subset of probability densities, namely real valued functions $f \in L^1(\mathcal{A})$ such that $f(a) \geq 0$ ν-almost everywhere, and such that $\int_A f(a)\, \mathrm{d}\nu(a) = 1$.

In the applications we have in mind there will basically be two possible situations. One case is when A is finite, the σ-algebra $\mathcal{B}$ consists of all the subsets of A and ν is the

counting measure on A. In this case $L^1(\mathcal{A}) = \mathbb{R}^A$, the space of all the possible functions from A to $\mathbb{R}$ and

$$\int_A f(a)\,\mathrm{d}\nu(a) = \sum_{a\in A} f(a)\,.$$

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

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

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

$$H(f) = -\int_{\{f>0\}} f(a)\log f(a)\,\mathrm{d}\nu(a)\,, \tag{2.1}$$

provided that the righthand side of (2.1) is well defined in $[-\infty, +\infty]$. Notice that the definition of entropy is thus dependent on the specific chosen measure space and in particular the integral in the righthand side of (2.1) is carried on with respect to the specific measure ν. In the finite case (2.1) reduces to the usual discrete entropy $\mathrm{H}(f) = -\sum_{a:f(a)>0} f(a)\log f(a)$ taking values in $[0, \log|A|]$. In the continuous case instead, it coincides with the so called differential entropy, effectively taking values in $[-\infty, +\infty]$ (see [3]). With a slight abuse of notation for any x in $[0,1]$ we will sometimes denote by $\mathrm{H}(x)$ the entropy of the binary probability density f in $\mathcal{P}(\{0,1\})$ defined by $f(1) = x$, $f(0) = 1 - x$. A few properties of the discrete entropy function are recalled in the Appendix.

For two $\mathbb{C}$-valued functions $\boldsymbol{f}, \boldsymbol{g}$ over a finite set A, we will use the notation $\langle \boldsymbol{f}, \boldsymbol{g}\rangle := \sum_{a\in A} \boldsymbol{f}(a)\boldsymbol{g}(a)^*$ for their scalar product, while $\boldsymbol{f}\cdot\boldsymbol{g} \in \mathbb{C}^A$ will denote their pointwise product. We shall indicate by $\mathrm{supp}(\boldsymbol{f}) := \{a \in A \big| \boldsymbol{f}(a) \neq 0\}$ the support of $\boldsymbol{f}$, while for $\mathbb{R}^+$-valued $\boldsymbol{f}$ and $\mathbb{C}$-valued $\boldsymbol{g}$ we define $\boldsymbol{f}^{\boldsymbol{g}}$ in $\mathbb{C}$ as $\boldsymbol{f}^{\boldsymbol{g}} := \prod_{a\in\mathrm{supp}(\boldsymbol{f})} \boldsymbol{f}(a)^{\boldsymbol{g}(a)}$.

2.2 Shannon theory for memoryless channels

In this section some notation and basic results from Shannon classical theory of memoryless channels is introduced.

A memoryless channel (MC) is described by

- a finite input set $\mathcal{X}$,

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

Such a channel will be identified by the triple $(\mathcal{X}, \mathcal{Y}, P)$.

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

A block encoder for the MC $(\mathcal{X}, \mathcal{Y}, P)$ consists of a finite set $\mathcal{U}$ and of a map $\phi : \mathcal{U} \to \mathcal{X}^N$. N is said to be the blocklength and

$$R := \frac{1}{N} \log |\mathcal{U}|$$

the transmission rate. The image of a block encoder

$$\mathcal{C} := \phi(\mathcal{U}) \subseteq \mathcal{X}^N$$

is a block code.

A decoder is any measurable mapping $\psi : \mathcal{Y}^N \to \mathcal{U}$. A coding scheme consists of a pair of an encoder and a decoder. Once a coding scheme has been fixed, its word error probability can be defined as follows. Assume $\boldsymbol{U}$ is a r.v. uniformly distributed on $\mathcal{U}$ and let $\boldsymbol{X} = \phi(\boldsymbol{U})$. Let moreover $\boldsymbol{Y}$ be the r.v. on Y^N whose probabilistic description is given by the conditional density $P_N(\boldsymbol{y}|\boldsymbol{x})$ and whose marginal density is thus given by

$$P_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} P_N(\boldsymbol{y}|\phi(u))$$

(in doing this we are automatically enforcing independence between $\boldsymbol{U}$ and the channel). Finally, let $\hat{\boldsymbol{U}} = \psi(\boldsymbol{Y})$ be the decoder's estimate of the transmitted message. The error probability is the probability of the event $\{\hat{\boldsymbol{U}} \neq \boldsymbol{U}\}$ and will be denoted by

$$p_e(\phi, \psi) := \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} p_e(\Phi, \psi|\, u)\,,$$

where

$$p_e(\Phi, \psi|\, u) := \int_{\mathcal{Y}^N} \mathbb{1}_{\{\psi^{-1}(\mathcal{U} \setminus \{u\})\}}(\boldsymbol{y}) \mathrm{d}\nu_N(\boldsymbol{y})$$

is the error probability conditioned to the transmission of the information word u.

It is well known that, given an encoder, the decoding scheme minimizing the error probability is the so called maximum likelihood (ML) decoding

$$\psi_{\mathrm{ML}}(\boldsymbol{y}) = \operatorname*{argmax}_{\boldsymbol{u}\in\mathcal{U}} P_N(\mathbf{y}|\phi(\mathbf{u}))\,,$$

From now on we will always assume that ML decoding is used and will use the simpler notation $p_e(\phi)$ and $p_e(\phi|\ u)$ for $p_e(\phi,\psi_{\mathrm{ML}})$ and $p_e(\phi,\psi_{\mathrm{ML}}|\ u)$.

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

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

$$p_e(\phi|\ u) = p_e(\phi\circ\sigma|\ \sigma^{-1}(u)) \tag{2.2}$$

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

$$\max_{1\le i\le r} p_e(\phi_i|\ u) \le P(e|\phi,u) \le \sum_{i=1}^{r} p_e(\phi_i|\ u) \tag{2.3}$$

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

$$p_e(\phi|\ u) \ge \frac{1}{2}\ . \tag{2.4}$$

It follows from (2.2) that, if ϕ is injective, $p_e(\phi)$ only depends on the encoder ϕ through its image, the code $\mathcal{C}=\phi(\mathcal{U})$. For this reason, we can speak of the error probability of a code $\mathcal{C}$, denoted by $p_e(\mathcal{C})$. The reason for considering (possibly non-injective) encoders, instead of codes only, is that sometimes they admit simpler parameterizations which are suitable for probabilistic averaging arguments.

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

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

$$C := \max_{p\in\mathcal{P}(\mathcal{X})} \sum_{x\in\mathcal{X}} p(x) \int_{\mathcal{Y}} P(y|x)\log\left(\frac{P(y|x)}{\sum_z p(z)P(y|z)}\right)\mathrm{d}\nu(y)\,. \tag{2.5}$$

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

$$E_0(\rho, p) := -\log \left(\int_{\mathcal{Y}} \left(\sum_{x \in \mathcal{X}} p(x) P(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \mathrm{d}\nu(y) \right) \tag{2.6}$$

and we define

$$E(R) := \max_{0 \le \rho \le 1} \max_{p \in \mathcal{P}(\mathcal{X})} E_0(\rho, p) - \rho R \,, \quad R \in [0, \log |\mathcal{X}|] \,. \tag{2.7}$$

A well known fact (see [31], [71]) is that

$$E(R) > 0 \;\Leftrightarrow\; R < C \,. \tag{2.8}$$

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

We can now state Shannon classical result:

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

(a)

$$\overline{p_e}^R \le \exp(-NE(R)) \,.$$

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

(b) For every $R > C$ there exists a constant $A_R > 0$ independent of N such that for any coding scheme (ϕ, ψ) having rate not smaller than R, we have that $p_e(\phi, \psi) \ge A_R$.

2.3 Symmetric channels and geometrically uniform constellations

In this thesis we will focus on channels exhibiting symmetries. Here we present fundamental definitions and examples.

We recall the concept of a group action. Given a finite group G with identity 1_G and a (finite) set A, we say that G acts on A if, for every $g \in G$, it is defined a map from A to A denoted by $a \mapsto ga$, such that $1_G a = a \,,\ \forall a \in A$; $h(ga) = (hg)a \,,\ \forall h, g \in G \,,\ \forall a \in A \,.$ The action of G over A is said to be (simply) transitive if for every $a, b \in A$ there exists

(just) one element g of G such that $ga = b$. If the action is simply transitive, G and A are clearly in bijection: $g \mapsto ga_0$ where a_0 is some fixed reference element in A.

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

$$\nu(gA) = \nu(A) \quad \forall A \in \mathcal{B}\,,\ \forall g \in G\,. \tag{2.9}$$

Notice that in the case when Y is a finite set, (2.9) is trivially always verified so that in this case all actions are isometric. Instead in the case when $Y = \mathbb{R}^n$, (2.9) is a real restriction and is verified if the maps $y \mapsto gy$ are isometries of $\mathbb{R}^n$.

Definition 2 *A MC $(\mathcal{X}, \mathcal{Y}, P)$ is said to be* G-symmetric *if*

(a) There exists a simply transitive action of G on $\mathcal{X}$,

(b) There exists an isometric action of G on $\mathcal{Y}$,

(c) $P(y|x) = P(gy|gx)$ for every $g \in G$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

It follows from (a) that $\mathcal{X}$ and G are in bijection: often we will tend to identify them.

A first important property of G-symmetric channels is that, for both their Shannon capacity C and their random coding exponent $E(R)$, the maximizing probability distribution $p \in \mathcal{P}(\mathcal{X})$ in the variational definitions (2.5) and (2.7) respectively can be chosen to be the uniform distribution over the input set $\mathcal{X}$. This easily follows from the convexity of the righthand side of (2.5), and the log-convexity of the righthand side of (2.6), as functions of the input distribution p, and their invariance with respect to the transitive action of G.

We now present a couple of simple examples.

Example 1 (Binary-input output-symmetric channels) *Consider the case when $G = \mathbb{Z}_2$. $\mathbb{Z}_2$-symmetric channels are known in the coding literature as binary-input output-symmetric (BIOS) channels. Typical examples are the binary symmetric channel (BSC), where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $P(1|0) = P(0|1)$, and the binary erasure channel (BEC), where $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, 2\}$, $P(1|0) = P(0|1) = 0$ and $P(2|0) = P(2|1)$.*

Example 2 (m-ary symmetric channel) *Consider a finite set $\mathcal{X}$ of cardinality $m \geq 2$ and some $\varepsilon \in [0, 1]$. The* m-ary symmetric channel *is described by the triple $(\mathcal{X}, \mathcal{X}, P)$, where $P(y|x) = 1 - \varepsilon$ 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 $m = 2$ is the well-known binary symmetric channel.*

The m-ary symmetric channel has 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. Another family of channels enjoying the same property is given by the additive Gaussian channel admitting m orthogonal equal-energy signals as input. In fact these were the channels considered by Gallager in [30, Sect.5]. 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 rich and important family of symmetric channels is provided by additive channels having geometrically uniform constellations as input. Consider the n-dimensional Euclidean space $\mathbb{R}^n$. An n-dimensional *constellation* is a finite subset $S \subset \mathbb{R}^n$ spanning $\mathbb{R}^n$, i.e. such that every $\mathbf{x} \in \mathbb{R}^n$ can be written as $\mathbf{x} = \sum_{s \in S} \alpha_s s$ with $\alpha_s \in \mathbb{R}$. We will restrict ourselves to the study of constellations $S \subset \mathbb{R}^n$ with barycenter $\mathbf{0}$, i.e. such that $\sum_{s \in S} s = \mathbf{0}$: they are the ones minimizing the average per symbol energy over the class of those constellations obtained one from the other by applying isometries.

We denote by $\mathrm{Iso}(S)$ its symmetry group, namely the set of all isometric permutations of S with the group structure endowed by the composition operation. Clearly $\mathrm{Iso}(S)$ acts on S. S is said to be *geometrically uniform (GU)* if this action is transitive; a subgroup $G \leq \Gamma(S)$ is a *generating group* for S if for every $s, r \in S$ a unique $g \in G$ exists such that $gs = r$, namely if G acts simply transitively on S. It is well known that not every GU constellation admits a generating group (see [65] for a counterexample). However in what follows we will always assume that the constellations we are dealing with, do admit generating groups, and, actually, Abelian ones.

Let S be an n-dimensional GU constellation equipped with a generating group G. Define the *S-AWGN channel* as the n-dimensional unquantized AWGN channel with input set S, output $\mathbb{R}^n$ with the usual measure structure, and transition probability densities given by $P(y|x) = N(y-x)$, where $N(x) = (2\pi\sigma^2)^{-n/2} e^{-||x||^2/2\sigma^2}$ is the density of an n-dimensional diagonal Gaussian random variable.

Now let S' be another GU constellation such that $S \subseteq S'$ and G is isomorphic to a subgroup of $\mathrm{Iso}(S')$. Let us introduce the quantization map over the Voronoi regions of S'

$$q : \mathbb{R}^n \to S', \qquad q(x) = \operatorname*{argmin}_{s \in S'} ||x - s||,$$

resolving non uniqueness cases by assigning to $q(x)$ a value arbitrarily chosen from the set of minima. We define the *(S, S')-AWGN channel* as the MC obtained by applying q to the output of the S-AWGN channel. Note that the special case $S = S'$ coincides with the so called hard decoding rule.

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

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

Example 3 *The unquantized isotropic Laplacian channel with input constrained on S is a S-AIN channel. Here $\mathcal{Y} = \mathbb{R}^n$ with the Lebesgue measure ν, while transition laws are given by*

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

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

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

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

In the following we present some examples of GU constellations admitting Abelian generating group.

Example 4 *The simplest, one-dimensional, GU constellation is the 2-PAM, 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.

Example 5 *For any integer $m \geq 2$, define $\xi_m := e^{i\frac{2\pi}{m}}$. Define the m-PSK constellation as*

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

Clearly S is two-dimensional for $m \geq 3$. It can be shown that $\Gamma(K_m) \simeq D_m$, where D_m is the dihedral group with $2m$ elements. K_m admits $\mathbb{Z}_m$, i.e. the Abelian group of integers modulo m, as generating group. When m is even there is another generating

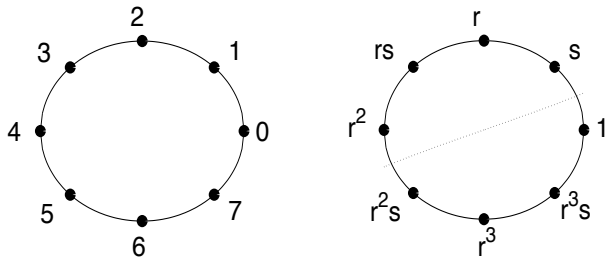

Figure 2.1: K_8-constellation with the two labelings $\mathbb{Z}_8$ and D_4

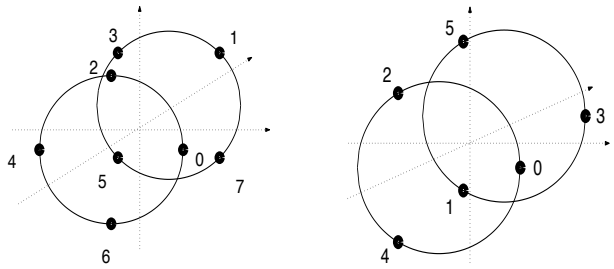

Figure 2.2: $\mathbb{Z}_8$-labelled K_8^β and $\mathbb{Z}_6$-labelled $K_{3\times 2}^\beta$

group (see [26], [44]): the dihedral group $D_{m/2}$, which is noncommutative for $m \geq 6$. Now, let $m'=am$ be an arbitrary multiple of m and define the quantization map over Voronoi regions of the m'-PSK constellation. The m-PSK-AWGN channel and the (m-PSK,m'-PSK)-AWGN channel are both $\mathbb{Z}_m$-symmetric and (for even m) $D_{m/2}$-symmetric. Constellation K_8 with the two possible labelings $\mathbb{Z}_8$ and D_4 is reported in Fig.2.1.

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

Example 6 *For any integer $m > 2$ consider the family of 3D GU constellations parametrized by $\beta \in (0, +\infty)$*

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

Fig.5.1 shows the special case $m = 3$. It's easy to show that $\mathbb{Z}_m \times \mathbb{Z}_2$ is a generating group for $K_{m\times 2}^\beta$; notice that, for odd m, $\mathbb{Z}_m \times \mathbb{Z}_2 \simeq \mathbb{Z}_{2m}$. Thus, for odd m, unquantized and quantized AWGN channels with input m-PSK×2-PAM are $\mathbb{Z}_{2m}$-symmetric. □

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

Example 7 *For even $m > 2$ we introduce the family of 3-dimensional GU constellations, parametrized by $\beta \in (0, +\infty)$*

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

An example with $m = 8$ is shown in Fig.5.1. It can be shown that, similarly to the constellations K_m, the the constellations K_m^β have two different generating groups, $\mathbb{Z}_m$ and $D_{m/2}$; so, in the standard way, we obtain channels that are both $\mathbb{Z}_m$-symmetric and $D_{m/2}$-symmetric. □

2.4 Bhattacharyya distance and the Gilbert-Varshamov bound for symmetric channels

The Gilbert-Varshamov (GV) bound is one of the most famous lower bounds on the achievable minimum Hamming distance of binary codes. Given a rate R in $(0, \log 2)$ and defined $\delta^{GV}(R)$ as the unique solution in $(0, 1/2)$ of the equation $\mathrm{H}(x) = \log 2 - R$, it states that there exist codes of length N and minimum distance at least $N\delta^{GV}(R)$, for every N.

The GV bound was introduced in early '50s [34, 70] and since then it has attracted a huge amount of attention from researchers. In particular the asymptotic tightness of the GV bound is one of the most important unproved conjectures in coding theory. This problem is closely related to the tightness of the expurgated error exponent at low rates [71]. A well known fact is that the Gilbert-Varshamov bound is asymptotically achieved with probability one by the binary linear coding ensemble [30], while this is not the case for the random coding ensemble. An analogous result holds for the expurgated error exponent on the BSC [4].

In this section we will present an extension of the GV bound to the non-binary case. There are many different notions of distance for non binary alphabets; the Hamming distance and the Lee distance for instance have been widely studied. However these distances have no direct application to the error exponents of channels usually considered. Here we will follow the approach of [9] considering the notion of Bhattacharyya distance of a memoryless channel and dealing with the corresponding Gilbert-Varshamov bound.

2.4.1 Bhattacharyya distance and weight

Consider a MC $(\mathcal{X}, \mathcal{Y}, P)$ and two input elements $x, x' \in \mathcal{X}$, we can consider the quantity $\int_{\mathcal{Y}} \sqrt{P(y|x)P(y|x')}\mathrm{d}\nu(y)$. Schwartz inequality gives

$$0 \leq \int_{\mathcal{Y}} \sqrt{P(y|x)P(y|x')}\mathrm{d}\nu(y) \leq \int_{\mathcal{Y}} P(y|x)\mathrm{d}\nu(y) \int_{\mathcal{Y}} P(y|x')\mathrm{d}\nu(y) = 1\,.$$

Moreover, the first inequality above is an equality iff the set $\{P(\,\cdot\,|x) > 0\} \cap \{P(\,\cdot\,|x') > 0\}$ has measure zero. Instead, the second inequality is equality iff $P(\cdot|x) = P(\cdot|x')$ ν-almost everywhere, which means that actually x and x' have indistinguishable outputs. In this paper we will assume that, for every $x \neq x'$, $0 < \int_{\mathcal{Y}} \sqrt{P(y|x)P(y|x')}\mathrm{d}\nu(y) < 1$. While there is no loss of generality in the latter part of this assumption, the former excludes from our analysis the class of channels whose 0-error capacity is strictly positive.

To any memoryless channel we can associate a function $\boldsymbol{\Delta} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ defined by

$$\boldsymbol{\Delta}(x, x') := -\log \int_{\mathcal{Y}} \sqrt{P(y|x)P(y|x')}\mathrm{d}\nu(y)\,.$$

This function is usually called the *Bhattacharyya distance* (or simply $\boldsymbol{\Delta}$-distance) of the channel and satisfies

$$\boldsymbol{\Delta}(x, x') = \boldsymbol{\Delta}(x', x)\,, \quad \forall\, x, x' \in \mathcal{X}\,; \qquad \boldsymbol{\Delta}(x, x') = 0 \;\Leftrightarrow\; x = x'\,.$$

If the MC $(\mathcal{X}, \mathcal{Y}, P)$ is G-symmetric, it is easy to verify that the Bhattacharyya distance function $\boldsymbol{\Delta}$ satisfies

$$\boldsymbol{\Delta}(gx, gx') = \boldsymbol{\Delta}(x, x') \qquad \forall x, x' \in \mathcal{X}\,,\ g \in G\,.$$

Identifying $\mathcal{X}$ with G as usual we can introduce the so-called *Bhattacharyya weight*:

$$\boldsymbol{\delta} : G \to [0, +\infty)\,, \qquad \boldsymbol{\delta}(x) = \boldsymbol{\Delta}(x, 1_G)\,, \qquad x \in G\,.$$

Notice that $\boldsymbol{\Delta}(x, x') = \boldsymbol{\delta}(x^{-1}x')$.

Bhattacharyya distance and weight can be extended to direct products in a natural way. Given $\boldsymbol{x}, \boldsymbol{x}'$ in $\mathcal{X}^N$, we put $\boldsymbol{\delta}(\boldsymbol{x}) = \sum_{i=1}^N \boldsymbol{\delta}(x_i)$ and $\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^N \boldsymbol{\Delta}(x_i, x'_i)$. The *minimum* $\boldsymbol{\Delta}$*-distance* of a code $\mathcal{C} \subseteq \mathcal{X}^N$ is defined as

$$\mathrm{d}_{\min}(\mathcal{C}) := \min\{\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{x}') \;\big|\; \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{C},\ \boldsymbol{x} \neq \boldsymbol{x}'\}\,.$$

In the case of a BIOS channel, we have that

$$\boldsymbol{\Delta}(\boldsymbol{x}, \boldsymbol{x}') = \textstyle\sum_{i=1}^N \boldsymbol{\delta}(x_i - x'_i) = \boldsymbol{\delta}(1)\,|\{1 \leq i \leq N :\, x_i \neq x'_i\}|\,, \qquad \forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^N\,,$$

i.e. the $\boldsymbol{\Delta}$-distance is proportional to the Hamming distance (the number of different entries of two strings). For the m-ary symmetric channel of Example 2 we obtain,

$$\boldsymbol{\Delta}(\boldsymbol{x},\boldsymbol{x}') = -\log\left(\varepsilon\tfrac{m-2}{m-1} + \sqrt{\tfrac{(1-\varepsilon)\varepsilon}{m-1}}\right)|\{1\le i\le N\,:\,x_i\neq x'_i\}|\,, \qquad \forall\,\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}^N\,,$$

so that once again the $\boldsymbol{\Delta}$-distance is proportional to the Hamming distance. For the S-AWGN channel instead, we obtain

$$\boldsymbol{\delta}(x) = -\log\int_{\mathbb{R}^n}\frac{1}{\sqrt{2\pi\sigma^2}^n}e^{-\left(||\boldsymbol{y}-\lambda_s(x)||^2+||\boldsymbol{y}-\lambda_s(1_G)||^2\right)/4\sigma^2}\mathrm{d}\boldsymbol{y} = ||\lambda_s(x)-\lambda_s(1_G)||^2\frac{\log e}{8\sigma^2}\,,$$

$$\boldsymbol{\Delta}(\boldsymbol{x},\boldsymbol{x}') = \boldsymbol{\delta}(\boldsymbol{x}-\boldsymbol{x}') = \textstyle\sum_{i=1}^N||\lambda_s(x_i-x'_i)-\lambda_s(1_G)||^2\frac{\log e}{8\sigma^2} = ||\lambda_s(\boldsymbol{x})-\lambda_s(\boldsymbol{x}')||^2\frac{\log e}{8\sigma^2}\,,$$

i.e. the Bhattacharyya distance is proportional to the squared Euclidean distance.

2.4.2 The Gilbert-Varshamov bound for symmetric channels

Suppose a G-symmetric MC is given, with G an arbitrary finite group, and let $\boldsymbol{\delta}$ the corresponding Bhattacharyya weight function. The Gilbert-Varshamov bound is a lower bound on the largest normalized minimum distance achievable by codes over G with rate greater than or equal to some value R. The result can be summarized as follows. For every R in $[0,\log|G|]$, define

$$\delta^{GV}(R) := \inf\left\{\langle\boldsymbol{\theta},\boldsymbol{\delta}\rangle\,\middle|\,\boldsymbol{\theta}\in\mathcal{P}(G)\,:\,\mathrm{H}(\boldsymbol{\theta})\ge\log|G|-R\right\}\,.$$

The following version of the GV bound, can be deduced from [9].

Theorem 5 *For every R in $(0,\log|G|)$ there exists a sequence of codes $(\mathcal{C}_N)$, with $\mathcal{C}_N\subseteq\mathcal{X}^N$, such that*

$$R(\mathcal{C}_N)\ge R\,,\qquad \mathrm{d}_{\min}(\mathcal{C}_N)\ge N\delta^{GV}(R)\,,\qquad \forall\,N\in\mathbb{N}\,.$$

The proof of Theorem 5 is constructive and based on the following greedy algorithm:

- initialize $A=\mathcal{X}^N$, $\mathcal{C}_N=\emptyset$;
- select an arbitrary point $\boldsymbol{x}$ from A; add $\boldsymbol{z}$ to $\mathcal{C}_N$ and erase from A the discrete $\boldsymbol{\Delta}$-ball of radius $r:=N\delta^{GV}(R)$ centered in $\boldsymbol{x}$

$$B_r(\boldsymbol{x}) := \left\{\boldsymbol{z}\in\mathcal{X}^N\,:\,\boldsymbol{\Delta}(\boldsymbol{z},\boldsymbol{x})<N\delta^{GV}(R)\right\}\,;$$

- iterate the previous point until $A=\emptyset$.

Clearly the algorithm described above constructs a code $\mathcal{C}_N \subseteq \mathcal{X}^N$ whose minimum distance satisfies $\mathrm{d}_{\min}(\mathcal{C}_N) \geq N\delta^{GV}(R)$. That it halts with a code whose rate $R(\mathcal{C}_N)$ is not smaller than R follows from the following estimate on the volume of discrete $\boldsymbol{\Delta}$-spheres $B_r(\boldsymbol{x}) := \{\boldsymbol{z} \in \mathcal{X}^N : ||\boldsymbol{x} - \boldsymbol{z}|| \leq r\}$ (see [44] for the Euclidean space case):

$$|B_r(\boldsymbol{x})| \leq \exp\left(N \max\left\{\mathrm{H}(\boldsymbol{\vartheta}) \,|\, \boldsymbol{\vartheta} \in \mathcal{P}(G) : \langle \boldsymbol{\vartheta}, \boldsymbol{\delta} \rangle \leq r/N\right\}\right) .$$

While Theorem 5 above guarantees the existence of a sequence of codes with rate not smaller than R and asymptotic normalized minimum distance above $\delta^{GV}(R)$, it is clear that the greedy algorithm its proof is based on does not guarantee that such a code sequence satisfies additional symmetry properties with respect to any algebraic structure.

2.5 Group codes and type-enumerating functions

When the MC is symmetric according to Definition 2, a natural class of codes to be considered is that of group codes. A *group code over* G, briefly *G-code*, of length N is any subgroup of the direct group product G^N. Group codes were first introduced by Slepian [65] as an extension of binary linear codes (the latter correspond to the case $G \simeq \mathbb{Z}_2$), and then studied by [44, 26]. In fact, G-codes enjoy many of the properties of binary-linear codes. In particular G-codes have complete symmetry, and as a consequence, when used on G-symmetric MC they enjoy the uniform error property, i.e. independence of the error probability on the transmitted codeword:

$$p_e(\mathcal{C}) = p_e(\mathcal{C}|\boldsymbol{x}) , \qquad \forall\, \boldsymbol{x} \in \mathcal{C} .$$

Structural properties of group codes have been extensively studied during the '80s and the '90s using the theory of behavioral group systems: see e.g. [29] and references therein.

For every G-code $\mathcal{C}$ of length N we now introduce some combinatorial quantities characterizing its performance. The *type-enumerating function* of a G-code $\mathcal{C}$ is defined as

$$W_{\mathcal{C}} : \mathcal{P}(G) \to \mathbb{Z}^+ , \qquad W_{\mathcal{C}}(\boldsymbol{\theta}) := \textstyle\sum_{\boldsymbol{x}\,:\,\boldsymbol{\theta}_G(\boldsymbol{x}) = \boldsymbol{\theta}} \mathbb{1}_{\mathcal{C}}(\boldsymbol{x}) \qquad \forall\, \boldsymbol{\theta} \in \mathcal{P}(G) .$$

Notice that since $\mathcal{C}$ is a subgroup of G^N, 1_{G^N} is always a codeword so that $W_{\mathcal{C}}(\delta_{1_G}) = 1$.

Assume we have fixed a G-symmetric MC channel $(\mathcal{X}, \mathcal{Y}, P)$ and let $\boldsymbol{\delta}$ be its associated Bhattacharyya weight. The minimum $\boldsymbol{\Delta}$-distance of a G-code $\mathcal{C}$ of length N is a function of its type-enumerating function:

$$\begin{aligned} \mathrm{d}_{\min}(\mathcal{C}) &= \min\{\boldsymbol{\delta}(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{C} \setminus \{\mathbf{0}\}\} \\ &= N \inf\left\{\langle \boldsymbol{\delta}, \boldsymbol{\theta} \rangle \,\middle|\, \boldsymbol{\theta} \in \mathcal{P}(G) \setminus \{\delta_0\} : W_{\mathcal{C}}(\boldsymbol{\theta}) > 0\right\} . \end{aligned} \tag{2.10}$$

Type-enumerating functions and the Bhattacharyya distance play an important role in the estimation of the maximum-likelihood decoding error probability of G-codes over memoryless G-symmetric channels. For instance, the so called union-Bhattacharyya bound, for the error probability of a G-code $\mathcal{C}$ of length N, can be written in the form

$$p_e(\mathcal{C}) \leq \sum_{\boldsymbol{\theta} \in \mathcal{P}(G)} W_{\mathcal{C}}(\boldsymbol{\theta}) \exp\left(-N\langle \boldsymbol{\delta}, \boldsymbol{\theta} \rangle\right) . \tag{2.11}$$

In fact, in Section 3.3.1 we will present a stronger result for the error probability of group codes.

We observe that (2.10) and (2.11) do not generally hold when a G-code is employed on MC which is not G-symmetric. While this is not an issue for the highly symmetric channels considered in Example 2, it does matter for the by far more common (and bandwidth efficient) AWGN chennel with a GU constellation as input. As a concrete example, one can think of the 8-PSK Gaussian channel. In this case, while both (2.10) and (2.11) are true for $\mathbb{Z}_8$-codes, for a $\mathbb{Z}_2^3$-code $\mathcal{C}$, and consequently for a $\mathbb{F}_8$-linear code, neither (2.10) nor (2.11) hold. In fact, the type-enumerating function of a $\mathbb{Z}_2^3$-code is not sufficient for characterizing its performance on the 8-PSK Gaussian channel.

Chapter 3

The capacity of Abelian group codes over symmetric channels

3.1 Introduction

In this chapter we address the problem of characterizing the capacity achievable by group codes over symmetric channels. It is a well-known fact that binary linear codes suffice to achieve capacity on binary input symmetric channels [31, 71]. The same is true for $\mathbb{Z}_p^r$-codes whenever p is a prime number; in fact in this case linear codes over the Galois field $\mathbb{F}_{p^r}$ have be shown capable to achieve the capacity of any $\mathbb{Z}_p^r$-symmetric channel [31]. Moreover, by averaging over the ensemble of linear codes, the same error exponent $E(R)$ is achieved as by averaging over random coding ensemble.

Here we investigate whether the same holds true for G-codes employed over G-symmetric channels. As a concrete example one might think of $\mathbb{Z}_m$-codes for the (m-PSK)-AWGN channel. In [44] it was conjectured that group codes should suffice in this case to achieve capacity exactly as in the binary case and, up to our knowledge, there has not been any progress towards this direction. On the other hand, interest in group codes has not decreased in these years: indeed they give the possibility to use more spectral efficient modulations while keeping many good qualities of the binary linear codes like the uniform error property and nice structure for the corresponding minimal encoders and minimal trellis representations. See [64, 38, 68, 28, 5, 45, 12, 46, 23, 41, 24, 25, 29] and references therein for an overview of the many research lines on group codes which have been developing during last years.

Our work focuses on the case when the group G is Abelian and consists of two parts. In the first part we determine a single-letter characterization for the capacity achievable using G-codes over this channel: this capacity is called the G-capacity. The result is contained in Theorem 6 which is a sort of inverse Shannon theorem and in Theorem 13

which exhibits an average result working in the ensemble of group encoders. Also the average error exponent is determined.

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

The chapter is organized as follows. In Section 3.2 we prove an inverse coding theorem for Abelian group codes, defining the G-capacity of a symmetric channel and showing that no reliable transmission is possible with G-codes at rates beyond this threshold value. The theorem is proved first for cyclic group codes, and the result is then extended to arbitrary Abelian groups. Section 3.3 contains the main result consisting in a channel coding theorem for Abelian group codes over symmetric channels, stating that reliable transmission is possible at any rate below the G-capacity. The result is obtained by using a probabilistic method: we introduce an ensemble of random group encoders and prove that its average word error probability goes to 0 as the blocklength is increased. More precisely we show that the average error probability goes to 0 exponentially fast in the blocklength and that the exponential rate of convergence is at least equal to a certain function $E_G(R)$ which we call the G-random coding exponent. Although we have no complete tightness result for $E_G(R)$ we show that even when there is no loss of capacity there is a loss in the error exponent at low rates. We also state a similar result holding for a different ensemble of group codes using the kernel representation instead of the encoder image one. Section 3.4 is devoted to the proof that for the AWGN channel with m-PSK constellation as input (and m the power of a prime) $\mathbb{Z}_m$-capacity and Shannon one do coincide, implying thus that $\mathbb{Z}_m$-codes employed over this channel achieve capacity. Finally, in Section 3.5 we provide an explicit counterexample consisting in a three-dimensional geometrically uniform constellation admitting $\mathbb{Z}_m$ as generating group: the AWGN channel with input restricted over this constellation the $\mathbb{Z}_m$-capacity is strictly less than Shannon capacity, implying thus that there is an algebraic obstruction to the use of $\mathbb{Z}_m$-codes in this case. It seems to be possible, but remains a completely open question, whether using non-Abelian group codes it allows to achieve capacity on this channel.

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

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

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

Whenever dealing with Abelian groups, we will use the additive notation to denote group operation, while 0 will always denote the identity element. We will use the symbol $\oplus$ to denote both external direct sum of groups, as well internal sum of subgroups when their intersection reduce to $\{0\}$. We will use the symbol $+$ and $\sum$ instead to denote general summation of subgroups. Some facts about the theory of Abelian groups will be recalled when needed, while we refer to [37] for further details.

3.2.1 The cyclic case

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

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

From now on we will identify $\mathcal{X}$ with $\mathbb{Z}_{p^r}$. For $l = 1, \ldots, r$, consider the channel obtained by restricting the input set from $\mathbb{Z}_{p^r}$ to its subgroup $p^{r-l}\mathbb{Z}_{p^r}$: call it the *l-th subchannel* and denote its capacity by C_l. The l-subchannel is easily seen to be $p^{r-l}\mathbb{Z}_{p^r}$-symmetric, so that C_l can be obtained, in the variational definition (2.5), with uniform distribution over the input set $p^{r-l}\mathbb{Z}_{p^r}$. As we will see soon, subchannels will play a fundamental role in our analysis. Let $\mathcal{U}$ be a finite Abelian group and $\phi : \mathcal{U} \to \mathbb{Z}_{p^r}^N$ a homomorphic encoder. It is not restrictive to assume that

$$\mathcal{U} = \mathbb{Z}_p^{k_1} \oplus \mathbb{Z}_{p^2}^{k_2} \oplus \ldots \oplus \mathbb{Z}_{p^r}^{k_r} \ . \tag{3.1}$$

for suitable positive integers $k_1, \ldots, k_r$. Indeed, in next subsection it will be shown that if $\mathcal{U}$ has not such a structure, than ϕ is surely noninjective so that $p_e(\phi) \geq \frac{1}{2}$ by property (2.4) of ML decoding. As a consequence of (3.1), there exist homomorphisms $\phi^j : \mathbb{Z}_{p^j}^{k_j} \to \mathbb{Z}_{p^r}^N$ such that, if we consider $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ with $\mathbf{u}_j \in \mathbb{Z}_{p^j}^{k_j}$ for every j, we have that $\phi(\mathbf{u}) = \sum_{j=1}^r \phi^j(\mathbf{u}_j)$.

ϕ's rate is given by

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

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

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

Note that

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

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

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

$$R^{(l)} \leq C_l \ . \tag{3.2}$$

Notice that,

$$\begin{aligned} R^{(l)} &= \tfrac{\log|\mathcal{U}_{(l)}|}{N} \\ &= \tfrac{\log p}{N}\left(\sum_{j=1}^{l} jk_j + l\sum_{j=l+1}^{r} k_j\right) \\ &\geq \tfrac{\log p}{N}\left(\tfrac{l}{r}\sum_{j=1}^{l} jk_j + l\sum_{j=l+1}^{r} \tfrac{j}{r}k_j\right) \\ &= \tfrac{\log p}{N}\tfrac{l}{r}\left(\sum_{j=1}^{l} k_j\right) \\ &= \tfrac{l}{r}R \ , \end{aligned} \tag{3.3}$$

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

By the property (2.3) of ML decoding,

$$p_e(\phi_l) \leq p_e(\phi)\,, \qquad l = 1, \ldots, r \ . \tag{3.4}$$

From (3.2), (3.3) and (3.4) it follows that necessary condition for $p_e(\phi)$ to be made arbitrarily small is that

$$R \leq \min_{l=1,\ldots,r} \frac{r}{l} C_l \,, \tag{3.5}$$

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

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

3.2.2 Arbitrary Abelian group

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

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

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

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

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

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

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

$$M = M_{(p_1^{r_1})} \oplus \cdots \oplus M_{(p_s^{r_s})}\,. \tag{3.6}$$

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

$$M_{(p_i^{r_i})} = \mathbb{Z}_{p_i}^{k_{i,1}} \oplus \mathbb{Z}_{p_i^2}^{k_{i,2}} \oplus \cdots \oplus \mathbb{Z}_{p_i^{r_i}}^{k_{i,r_i}}\,. \tag{3.7}$$

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

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

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 will say that M is σ-adapted if σ^M is a subsequence of σ. Notice that, once the sequence of primes σ has been fixed, all σ-adapted Abelian groups are completely determined by their type (which includes the multiplicities r_i^M with the agreement that some of them could be equal to 0). We will denote by $M_{\mathbf{k}}$ the finite Abelian group having type $\mathbf{k}$.

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

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

Suppose now we have fixed, once for all, a finite Abelian group G having spectrum $\sigma^G = (p_1, \ldots, p_s)$, multiplicity $\mathbf{r}^G = (r_1^G, \ldots, r_s^G)$ and type $\mathbf{k}^G$. We will consider G-encoders $\phi \in \mathrm{Hom}(\mathcal{U}, G^N)$ with domain consisting of a finite Abelian group $\mathcal{U}$ which is σ^G-adapted and is such that, $\mathbf{r}^{\mathcal{U}} \leq \mathbf{r}^G$ (in the sense that $r_i^{\mathcal{U}} \leq r_i^G$ for each i). In fact if $\mathcal{U}$ does not fulfil these requirements then ϕ is surely noninjective for our previous considerations, and thus its ML woed error probability is bounded from below by the constant $1/2$. The group $\mathcal{U}$ admits a decomposition as illustrated above in (3.6) and (3.7). Let us fix now a matrix

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

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

$$\mathcal{U}(\mathbf{l}) = \bigoplus_{i=1}^{s} \mathcal{U}_{(p_i^{r_i^G})}(\mathbf{l}_i)\,. \tag{3.8}$$

$$\mathcal{U}_{(p_i^{r_i^G})}(\mathbf{l}_i) = \bigoplus_{j=1}^{r_i^G} p_i^{j-l_{i,j}} \mathbb{Z}_{p_i^j}^{k_{i,j}}\,. \tag{3.9}$$

An immediate consequence of previous considerations is that

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

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

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

Then,

$$R_{\mathbf{l}} \leq C_{\mathbf{l}} \ \text{ for every } \mathbf{r}^G - \text{compatible } \mathbf{l} \tag{3.10}$$

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

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

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

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

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

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

Hence,

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

Consequently, (3.10) can be equivalently expressed as

$$R \leq \min_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{comp.}}} \frac{C_{\mathbf{l}}}{\sum_{i=1}^{s}\sum_{j=1}^{r_i^G}\frac{l_{i,j}}{j}\alpha_{i,j}}\,, \tag{3.11}$$

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

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

$$C_G = \max_{\boldsymbol{\alpha}\in\mathcal{P}(\mathbf{r}^G)}\ \min_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{comp.}}} \frac{C_{\mathbf{l}}}{\sum_{i=1}^{s}\sum_{j=1}^{r_i^G}\frac{l_{i,j}}{j}\alpha_{i,j}}\,. \tag{3.12}$$

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

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

is a continuous map from $\mathcal{P}(\boldsymbol{r}^G)$ to $\mathbb{R}^+$, definition (3.12) is well posed in the sense that f has a maximum point in $\mathcal{P}(\boldsymbol{r}^G)$. Such a maximum point could be not unique in principle: nevertheless we will call G-optimal splitting and denote by $\boldsymbol{\alpha}_G$ any element of $\mathcal{P}(\boldsymbol{r}^G)$ such that

$$\min_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{comp.}}} \frac{C_{\mathbf{l}}}{\sum_{i=1}^{s}\sum_{j=1}^{r_i^G}\frac{l_{i,j}}{j}\alpha_{i,j}^G} = C_G\,. \tag{3.13}$$

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

Theorem 6 *Consider a G-symmetric channel and let C_G be its G-capacity. Then, for every $R > C_G$ there exists $A_R > 0$ depending on R but not on N, such that, for every G-encoder ϕ of rate R and length N, with any decoding rule, the corresponding word error probability satisfies*

$$p_e(\phi) \geq A_R \ .$$

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

Example 8 *Suppose the group G admits Galois field structure. In this case we necessarily have $G \simeq \mathbb{Z}_p^k$ for some prime p and positive integer k. Thus*

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

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

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

$$n \geq \begin{cases} k, & \text{if } p = 2 \ ; \\ 2k, & \text{if } p \geq 2 \ . \end{cases} \tag{3.14}$$

□

In next example we would like to show that in the special case when $G = \mathbb{Z}_{p^r}$ condition (3.11) coincides with condition (3.5) obtained in the previous subsection.

Example 9 *Let $G = \mathbb{Z}_{p^r}$. We want to show that*

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

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

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

where

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

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

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

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

Now,

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

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

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

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

□

Example 10 *Now consider the* $K_{2\times 3}^{\beta}$ *constellation introduced in Example 6. Consider a* $K_{2\times 3}^{\beta}$*-AWGN channel. It is easy to show that the independence of orthogonal components of the Gaussian noise imply that the capacity* $C_6(\beta)$ *of such a channel is equal to the sum of the capacities of its two subchannels,* $C_2(\beta)$ *and* $C_3(\beta)$. *This fact allows us to explicitly write down the optimal splitting, i.e. the* $\boldsymbol{\alpha} \in \mathcal{P}(r^G)$ *solution of the variational problem (3.12) 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* $\boldsymbol{r}^G = \left(r_1^G, r_2^G\right) = (1, 1)$. *(3.12) reduces to*

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

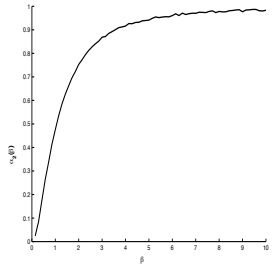

Figure 3.1: The optimal splitting for $K^{\beta}_{2\times 3}$ as a function of β

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

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

Indeed we have that

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

In Figure 3.1 $\alpha_2^{\mathbb{Z}_6}(\beta)$ *is plotted: notice how the optimal splitting follows the geometry of the constellation as* $\alpha_2(\beta)$ *is monotonically increasing in* β *with* $\lim_{\beta\to 0} \boldsymbol{\alpha}^{\mathbb{Z}_6}(\beta) = (0,1)$ *(as* β *goes to* 0 $K_{2\times 3}(\beta)$ *collapses onto constellation* K_3*) and* $\lim_{\beta\to+\infty} \boldsymbol{\alpha}^{\mathbb{Z}_6}(\beta) = (1,0)$ *(as* β *goes to* $+\infty$ $K_{2\times 3}(\beta)$ *collapses onto constellation 2-PAM).* □

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

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

3.3 Classical ensembles of G-codes

We now present a result which completes Theorem 6 by stating that at every rate $R < C_G$ reliable transmission over a G-symmetric channel is possible using G-encoders.

Following the classical technique originally proposed by Shannon we will use a probabilistic method, introducing ensembles of G-encoders and analyzing their average performances. This will then allow us to obtain our result. This technique had already been used to study performances of linear codes in [31]: this covers the case when $G \simeq \mathbb{Z}_p^k$ for some prime p. For general Abelian group however the derivation is more complicate.

We consider G-code ensembles, defined as sequences of Abelian groups $\mathcal{U}_N$ and of independent uniformly distributed random variables $\Phi_N \in \mathrm{Hom}(\mathcal{U}_N, G^N)$. We will see later that different choices of ensembles are possible and give similar results.

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

$$(k_N)_{i,j} = \left\lfloor \frac{RN\alpha_{i,j}}{j \log p_i} \right\rfloor . \tag{3.15}$$

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

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

3.3.1 Gallager Bound for codes over groups

In this subsection we state a convenient version of the Gallager bound (see [31]) for the special case of G-symmetric channels; it is based on the techniques presented in [62].

We start by recalling the classical Gallager bound.

Lemma 7 (Gallager bound) *Given a MC $(\mathcal{X}, \mathcal{Y}, P)$, suppose we have a block encoder*

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

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

$$p_e(\phi|\, u) \leq \int_{\mathcal{Y}^N} P_N(\mathbf{y}|\phi(u))^{\frac{1}{1+\rho}} \left(\sum_{v \neq u} P_N(\mathbf{y}|\phi(v))^{\frac{1}{1+\rho}} \right)^{\rho} \mathrm{d}\mu^N(\mathbf{y}) . \tag{3.16}$$

We now want to rewrite the Gallager bound in the special case when the channel is G-symmetric for an Abelian group G. It is not restrictive to assume that $\mathcal{X} = G$. Recall that, for any $\boldsymbol{x}$ in G^N,$\boldsymbol{\theta}_G(\mathbf{x})$ in $\mathcal{P}(G)$ denotes the type or empirical frequency of $\boldsymbol{x}$. The subset of $\mathcal{P}(G)$ containing all types of vectors $\mathbf{x} \in G^N$ is denoted by $\mathcal{P}_N(G)$. For $\boldsymbol{\theta} \in \mathcal{P}_N(G)$ we define $G^N_{\boldsymbol{\theta}}$ as the subset of G^N containing all vectors of type $\boldsymbol{\theta}$. Clearly $G^N = \cup_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} G^N_{\boldsymbol{\theta}}$. We introduce type-spectra of an encoder $\phi : \mathcal{U} \to G^N$. For each $u \in \mathcal{U}$ and $\boldsymbol{\theta} \in \mathcal{P}_N(G)$ we define $W_\phi(\boldsymbol{\theta}|\ u)$ as the cardinality of the subset of $\mathcal{U} \setminus \{u\}$ consisting of those v such that the difference $\phi(v) - \phi(u)$ has type $\boldsymbol{\theta}$, i.e.

$$W_\phi(\boldsymbol{\theta}|\ u) = \sum_{v \in \mathcal{U} \setminus \{u\}} \mathbb{1}_{\mathcal{G}^N_{\boldsymbol{\theta}}} \left(\phi(v) - \phi(u)\right) \ . \tag{3.17}$$

Lemma 8 *Given a G-symmetric MC $(G, \mathcal{Y}, P)$, suppose we have a block encoder*

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

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

$$p_e(\phi|\ u) \leq \frac{1}{|G|^N} \sum_{\mathbf{z} \in G^N} \int_{\mathcal{Y}^N} P_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}} \left(\sum_{\boldsymbol{\theta} \in \mathcal{P}_N(G)} \frac{W_\phi(\boldsymbol{\theta}|\ u)}{\binom{N}{N\boldsymbol{\theta}}} \sum_{\mathbf{x} \in G^N_{\boldsymbol{\theta}}} \left(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x})\right)^{\frac{1}{1+\rho}} \right)^{\rho} \mathrm{d}\mu^N(\mathbf{y}) \, . \tag{3.18}$$

Proof: We generate the following random encoder from ϕ:

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

where:

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

Throughout the proof we will denote by $\mathbb{E}[\cdot]$ the average operator with respect to such a probabilistic structure. The crucial point here is that the average word error probability

of the random encoder Φ is equal to the word error probability of ϕ. In fact for any realization π of Π

$$p_e(\phi\pi|\ u) = p_e(\phi\pi|\ u) = p_e(\phi|\ u)\ .$$

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

$$\begin{aligned}
p_e(\omega\phi|\ u) &= 1 - \int\limits_{\Lambda_{\omega\phi}(u)} P_N(\mathbf{y}|\omega\phi u)\mathrm{d}\mu^N(\mathbf{y}) \\
&= 1 - \int\limits_{\omega\Lambda_\phi(u)} P_N(\mathbf{y}|\omega\phi u)\mathrm{d}\mu^N(\mathbf{y}) \\
&= 1 - \int\limits_{\Lambda_\phi(u)} P_N(\omega\mathbf{y}|\omega\phi u)\mathrm{d}\mu^N(\mathbf{y}) \\
&= 1 - \int\limits_{\Lambda_\phi(u)} P_N(\mathbf{y}|\phi u)\mathrm{d}\mu^N(\mathbf{y}) \\
&= p_e(\phi|\ u)\ .
\end{aligned}$$

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

$$p_e(\mathbf{g} + \phi|\ u) = p_e(\phi|\ u)\ .$$

Thus we have

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

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

$$\begin{aligned}
P(\Phi(u) = \mathbf{x}) &= P(\mathbf{G} + \Omega\phi(\Pi u) = \mathbf{x}) \\
&= \sum_{\mathbf{z}\in G^N} P(\mathbf{G} = \mathbf{x} - \mathbf{z}|\Omega\phi(\Pi u) = \mathbf{z})P(\Omega\phi(\Pi u) = \mathbf{z}) \\
&= \sum_{\mathbf{z}\in G^N} P(\mathbf{G} = \mathbf{x} - \mathbf{z})P(\Omega\phi(\Pi u) = \mathbf{z}) \qquad (3.19) \\
&= \sum_{\mathbf{z}\in G^N} \frac{1}{|G|^N} P(\Omega\phi(\Pi u) = \mathbf{z}) = \frac{1}{|G|^N}\ ;
\end{aligned}$$

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

$$P\big(\phi(\Pi v) = \mathbf{x}\big) = \frac{1}{|\mathcal{U}| - 1} \sum_{w\in\mathcal{U}\setminus\{u\}} \mathbb{1}_{\{\mathbf{x}\}}(\phi(w))\ . \qquad (3.20)$$

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

$$\begin{aligned} P(\Phi(v)=\mathbf{z}+\mathbf{x}|\Phi(u)=\mathbf{z}) &= P(\Phi(v)-\Phi(u)=\mathbf{x}|\Phi(u)=\mathbf{z}) \\ &= P(\Omega\phi(\Pi v)+\mathbf{G}-\Omega\phi(\Pi u)-\mathbf{G}=\mathbf{x}|\mathbf{G}+\Omega\phi(u)=\mathbf{z}) \\ &= P(\Omega\phi(\Pi v)-\Omega\phi(u)=\mathbf{x}|\mathbf{G}+\Omega\phi(u)=\mathbf{z}) \\ &= P(\Omega\phi(\Pi v)-\Omega\phi(u)=\mathbf{x})\ . \end{aligned} \tag{3.21}$$

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

$$(N\boldsymbol{\theta}(\mathbf{x}))!:=\prod_{g\in G}(N\theta_g(\mathbf{x}))!\ . \tag{3.22}$$

By successively applying (3.21), (3.22), (3.20) and (3.17) we get

$$\begin{aligned} P(\Phi(v)=\mathbf{z}+\mathbf{x}|\Phi(u)=\mathbf{z}) &= P(\Omega(\phi(\Pi v)-\phi(u))=\mathbf{x}) \\ &= \sum_{\omega\in S_N}\frac{1}{N!}P(\phi(\Pi v)-\phi(u)=\omega\mathbf{x}) \\ &= \tfrac{1}{N!}\sum_{\mathbf{y}\in G^N_{\boldsymbol{\theta}(\mathbf{x})}}\ \sum_{\omega\in Stab(\mathbf{y})}P(\phi(\Pi v)=\phi(u)+\mathbf{y}) \\ &= \tfrac{1}{N!}\sum_{\mathbf{y}\in G^N_{\boldsymbol{\theta}(\mathbf{x})}}(N\boldsymbol{\theta}(\mathbf{x}))!\frac{1}{|\mathcal{U}|-1}\sum_{v\in\mathcal{U}\setminus\{u\}}\mathbb{1}_{\{\phi(u)+\mathbf{y}\}}(\phi(v)) \\ &= \tbinom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1}\sum_{\mathbf{y}\in G^N_{\boldsymbol{\theta}(\mathbf{x})}}\frac{1}{|\mathcal{U}|-1}\sum_{v\in\mathcal{U}\setminus\{u\}}\mathbb{1}_{\{\mathbf{y}\}}(\phi(v)-\phi(u)) \\ &= \tbinom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1}\tfrac{1}{|\mathcal{U}|-1}\sum_{v\in\mathcal{U}\setminus\{u\}}\mathbb{1}_{G^N_{\boldsymbol{\theta}(\mathbf{x})}}(\phi(v)-\phi(u)) \\ &= \tfrac{1}{|\mathcal{U}|-1}\tbinom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1}W_\phi(\boldsymbol{\theta}(\mathbf{x})|\ u)\ . \end{aligned} \tag{3.23}$$

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

get

$$
\begin{aligned}
p_e(\phi|\ u) &= \mathbb{E}\left[p_e(\Phi|\ u)\right] \\
&\leq \mathbb{E}\left[\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\Phi(u))\left(\sum_{v\in\mathcal{U}\setminus\{u\}}\left(\frac{P_N(\mathbf{y}|\Phi(v))}{P_N(\mathbf{y}|\Phi(u))}\right)^{\frac{1}{1+\rho}}\right)^{\rho} \mathrm{d}\mu^N(\mathbf{y})\right] \\
&= \mathbb{E}\left[\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\Phi(u))^{\frac{1}{1+\rho}}\left(\sum_{v\in\mathcal{U}\setminus\{u\}}\left(P_N(\mathbf{y}|\Phi(v))\right)^{\frac{1}{1+\rho}}\right)^{\rho} \mathrm{d}\mu^N(\mathbf{y})\right] \\
&= \tfrac{1}{|G|^N}\sum_{\mathbf{z}\in G^N}\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}}\mathbb{E}\left[\left(\sum_{v\in\mathcal{U}\setminus\{u\}}\left(P_N(\mathbf{y}|\Phi(v))\right)^{\frac{1}{1+\rho}}\right)^{\rho}\middle|\ \Phi(u)=\mathbf{z}\right]\mathrm{d}\mu^N(\mathbf{y})\ ,
\end{aligned}
\tag{3.24}
$$

last equality following from (3.19). The conditional expectation in the last term of (3.24) can be upperbounded by the Jensen inequality, yielding

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sum_{v\in\mathcal{U}\setminus\{u\}}\left(P_N(\mathbf{y}|\Phi(v))\right)^{\frac{1}{1+\rho}}\right)^{\rho}\middle|\ \Phi(u)=\mathbf{z}\right] \\
&\leq \left(\mathbb{E}\left[\sum_{v\in\mathcal{U}\setminus\{u\}}\left(P_N(\mathbf{y}|\Phi(v))\right)^{\frac{1}{1+\rho}}\middle|\ \Phi(u)=\mathbf{z}\right]\right)^{\rho} \\
&= \left(\sum_{\mathbf{x}\in G^N}\sum_{v\in\mathcal{U}\setminus\{u\}} P(\Phi(v)=\mathbf{z}+\mathbf{x}|\Phi(u)=\mathbf{z})\left(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x})\right)^{\frac{1}{1+\rho}}\right)^{\rho} \\
&= \left(\sum_{\boldsymbol{\theta}\in\mathcal{P}_N(G)}\sum_{\mathbf{x}\in G^N_{\boldsymbol{\theta}}} W_\phi(\boldsymbol{\theta}(\mathbf{x})|\ u)\binom{N}{N\boldsymbol{\theta}(\mathbf{x})}^{-1}\left(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x})\right)^{\frac{1}{1+\rho}}\right)^{\rho} \\
&= \left(\sum_{\boldsymbol{\theta}\in\mathcal{P}_N(G)} W_\phi(\boldsymbol{\theta}|\ u)\binom{N}{N\boldsymbol{\theta}}^{-1}\sum_{\mathbf{x}\in G^N_{\boldsymbol{\theta}}}\left(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x})\right)^{\frac{1}{1+\rho}}\right)^{\rho}
\end{aligned}
\tag{3.25}
$$

where the second equality follows from (3.23) and from $G^N = \bigcup_{\boldsymbol{\theta}\in\mathcal{P}_N(G)} G^N_{\boldsymbol{\theta}}$. Substituting (3.25) into (3.24) yields (3.18). ■

We would like to emphasize the fact that both Definition 3.17 and Lemma 8 do not need ϕ to be a G-encoder; in what follows we will make use of this generality. When ϕ

is a G-encoder it is easy to show that $W_\phi(\boldsymbol{\theta}|\ u)$ does not depend on u, so in this case we will use the notation $W_\phi(\boldsymbol{\theta})$. Generalizations of both Definition 3.17 and Lemma 8 to non Abelian groups are straightforward: the only difference is that one has to define left and right distance spectra (the two notions coincide for group encoders but they generally do not for arbitrary encoders).

3.3.2 Averaged estimations

The idea is to average estimation (3.18) on our ensembles. In the field case this would lead us to the classical direct Shannon theorem for linear codes. However, in this context, we need to be more careful since averaging distance spectra becomes more delicate. For this we first develop some further considerations on random variables taking values over Abelian groups.

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

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

Lemma 9 *$\theta \circ X$ is a r.v. uniformly distributed over H.*

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

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

■

Let us go back to our setting with the Abelian groups M and L. Given any $m \in M$ we can consider the valuation homomorphism $\psi_{M,L,m} : \mathrm{Hom}(M, L) \to L$ given by $\psi_{M,L,m}(\phi) = \phi(m)$. Using Lemma 9 we thus obtain that the r.v. $\Phi(m)$ is uniformly distributed on $\mathrm{Im}(\psi_{M,L,m})$. The problem is therefore to characterize the image of $\psi_{M,L,m}$: this depends on the choice of the element m.

We gather a few simple properties of the valuation homomorphism:

Lemma 10 *$\psi_{M,L,m}$ satisfies the following properties:*

(1) If $M = \mathbb{Z}_{p^r}$ and $m \in M$ is invertible, we have that $\mathrm{Im}(\psi_{M,L,m}) = L_{(p^r)}$.

(2) Assume that $M = M_1 \oplus M_2$ and let $m = (m_1, m_2) \in M$. Then, $\mathrm{Im}(\psi_{M,L,m}) = \mathrm{Im}(\psi_{M_1,L,m_1}) + \mathrm{Im}(\psi_{M_2,L,m_2})$.

Assume that M has the structure given by (3.6) and (3.7). Each $m \in M$ can be decomposed accordingly

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

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

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

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

$$H_{\mathbf{l}} = \{m \in M \mid \mathbf{l}(m) = \mathbf{l}\}\,. \tag{3.26}$$

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

Proposition 11 *Let $m \in H_{\mathbf{l}}$. Then,*

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

Proof Immediate consequence of Lemma 10. ■

Corollary 12 *Let $m \in H_{\mathbf{l}}$. Then, the r.v. $\Phi(m)$ is uniformly distributed over the set*

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

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

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

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

$$(k_N)_{i,j} = \left\lfloor \frac{RN\alpha_{i,j}}{j \log p_i} \right\rfloor\,. \tag{3.27}$$

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

We are now ready to prove our first fundamental result:

Theorem 13 *Let $(G,\mathcal{Y},P)$ be a G-symmetric MC. For every $R\in[0,\log|G|[$, $\boldsymbol{\alpha}\in\mathcal{P}(\mathbf{r}^G)$, the following estimation holds true:*

$$\overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha})}\leq \sum_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{compatible}}} \exp\left(-NE_{\mathbf{l}}\left(R_{\mathbf{l}}\right)\right)\,, \tag{3.28}$$

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

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

Proof Let $H_{\mathbf{k_N},\mathbf{l}}$ be the set defined by (3.26) for the group $\mathcal{U}_{\mathbf{k}_N}$. We can thus decompose

$$\mathcal{U}_{\mathbf{k}_N}=\bigcup_{\substack{\mathbf{l}\\ \mathbf{r}^G-\text{comp.}}} H_{\mathbf{k_N},\mathbf{l}}\,. \tag{3.29}$$

It follows from Corollary 12 that, if $\mathbf{u}\in H_{\mathbf{k_N},\mathbf{l}}$, $\Phi_N(\mathbf{u})$ is a r.v. uniformly distributed over $G_{\mathbf{l}}^N$.

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

$$p_e(\phi)=p_e(\phi,\mathbf{0})$$

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

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

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

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

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

A union bound yields

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

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

Consider the r.v. $\Phi_{N,\mathbf{l}}$ obtained by restricting Φ_N to the subset $H_{\mathbf{k_N},\mathbf{l}}$. Now, given an $\mathbf{r}^G$-compatible $\mathbf{l}$, apply the bound of Lemma 8 (which, as we already remarked, does not need the encoder to be an homomorphism) to each realization of $p_e(\Phi_{N,\mathbf{l}},\mathbf{0})$, and then average with respect to Φ_N. For any $\rho\in[0,1]$ we obtain

$$\begin{aligned}
&\overline{p_e(\Phi_{N,\mathbf{l}},\mathbf{0})}^{(R,\boldsymbol{\alpha})} \leq \\
&\leq \overline{\frac{1}{|G|^N}\sum_{\mathbf{z}\in G^N}\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}}\left(\sum_{\boldsymbol{\theta}\in\mathcal{P}_N(G)}\frac{W_{\Phi_{N,\mathbf{l}}}(\boldsymbol{\theta}|\,0)}{\binom{N}{N\boldsymbol{\theta}}}\sum_{\mathbf{x}\in G^N_{\boldsymbol{\theta}}}(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{\rho}\mathrm{d}\mu^N(\mathbf{y})}^{(R,\boldsymbol{\alpha})} \\
&\leq \frac{1}{|G|^N}\sum_{\mathbf{z}\in G^N}\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}}\left(\sum_{\boldsymbol{\theta}\in\mathcal{P}_N(G)}\frac{\overline{W_{\Phi_{N,\mathbf{l}}}(\boldsymbol{\theta}|\,0)}^{(R,\boldsymbol{\alpha})}}{\binom{N}{N\boldsymbol{\theta}}}\sum_{\mathbf{x}\in G^N_{\boldsymbol{\theta}}}(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{\rho}\mathrm{d}\mu^N(\mathbf{y}) \ ,
\end{aligned} \tag{3.30}$$

where the last inequality follows from Jensen inequality.

It remains to calculate the average distance spectra of $\Phi_{N,\mathbf{l}}$. Using the fact (see Corollary 12) that for any $\mathbf{u}\in H_{\mathbf{k}_N,\mathbf{l}}$ we have that $\Phi_N(\mathbf{u})$ is uniformly distributed over $G^N_{\mathbf{l}}$, we obtain

$$\begin{aligned}
\overline{W_{\Phi_{N,\mathbf{l}}}(\boldsymbol{\theta}|\mathbf{0})}^{(R,\boldsymbol{\alpha})} &= \overline{\sum_{\mathbf{u}\in H_{\mathbf{k}_N,\mathbf{l}}}\mathbb{1}_{G^N_{\boldsymbol{\theta}}}(\Phi_{\mathbf{l}}\mathbf{u})}^{(R,\boldsymbol{\alpha})} = \sum_{\mathbf{u}\in H_{\mathbf{k}_N,\mathbf{l}}}\overline{\mathbb{1}_{G^N_{\boldsymbol{\theta}}}(\Phi\mathbf{u})}^{(R,\boldsymbol{\alpha})} \\
&= \sum_{\mathbf{u}\in H_{\mathbf{k}_N,\mathbf{l}}} P\left(\Phi_{N,\mathbf{l}}(\mathbf{u})\in G^N_{\boldsymbol{\theta}}\right) = |H_{\mathbf{k}_N,\mathbf{l}}|\frac{\binom{N}{N\boldsymbol{\theta}}\mathbb{1}_{\mathcal{P}_N(G_{\mathbf{l}})}(\boldsymbol{\theta})}{|G_{\mathbf{l}}|^N}
\end{aligned} \tag{3.31}$$

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

exactly one element for each coset of $G_{\mathbf{l}}^N$. By substituting (3.31) into (3.30) we obtain

$$\begin{aligned}
&\overline{p_e(\Phi_{N,\mathbf{l}},\mathbf{0})}^{(R,\boldsymbol{\alpha})} \\
&\leq \frac{1}{|G|^N}\sum_{\mathbf{z}\in G^N}\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}}\left(\sum_{\boldsymbol{\theta}\in\mathcal{P}_N(G)}|H_{\mathbf{k}_N,\mathbf{l}}|\frac{1}{|G_{\mathbf{l}}|^N}\mathbb{1}_{\mathcal{P}_N(G_{\mathbf{l}})}(\boldsymbol{\theta})\sum_{\mathbf{x}\in G_{\boldsymbol{\theta}}^N}(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{\rho}\mathrm{d}\mu^N(\mathbf{y}) \\
&= \frac{1}{|G|^N}\sum_{\mathbf{z}\in G^N}\int_{\mathcal{Y}^N} P_N(\mathbf{y}|\mathbf{z})^{\frac{1}{1+\rho}}\left(|H_{\mathbf{k}_N,\mathbf{l}}|\frac{1}{|G_{\mathbf{l}}|^N}\sum_{\mathbf{x}\in G_{\mathbf{l}}^N}(P_N(\mathbf{y}|\mathbf{z}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{\rho}\mathrm{d}\mu^N(\mathbf{y}) \\
&= |H_{\mathbf{k}_N,\mathbf{l}}|^{\rho}\int_{\mathcal{Y}^N}\sum_{\mathbf{v}\in\Omega_{\mathbf{l}}}\frac{|G_{\mathbf{l}}|^N}{|G|^N}\sum_{\mathbf{w}\in G_{\mathbf{l}}^N}\frac{1}{|G_{\mathbf{l}}|^N}P_N(\mathbf{y}|\mathbf{v}+\mathbf{w})^{\frac{1}{1+\rho}}\left(\frac{1}{|G_{\mathbf{l}}|^N}\sum_{\mathbf{x}\in G_{\mathbf{l}}^N}(P_N(\mathbf{y}|\mathbf{v}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{\rho}\mathrm{d}\mu^N(\mathbf{y}) \\
&= |H_{\mathbf{k}_N,\mathbf{l}}|^{\rho}\sum_{\mathbf{v}\in\Omega_{\mathbf{l}}}\frac{|G_{\mathbf{l}}|^N}{|G|^N}\int_{\mathcal{Y}^N}\left(\frac{1}{|G_{\mathbf{l}}|^N}\sum_{\mathbf{x}\in G_{\mathbf{l}}^N}(P_N(\mathbf{y}|\mathbf{v}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{1+\rho}\mathrm{d}\mu^N(\mathbf{y})\ .
\end{aligned} \tag{3.32}$$

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

$$\begin{aligned}
&\int_{\mathcal{Y}^N}\left(\frac{1}{|G_{\mathbf{l}}|^N}\sum_{\mathbf{x}\in G_{\mathbf{l}}^N}(P_N(\mathbf{y}|\mathbf{v}+\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{1+\rho}\mathrm{d}\mu^N(\mathbf{y}) \\
&= \int_{\mathcal{Y}^N}\left(\frac{1}{|G_{\mathbf{l}}|^N}\sum_{\mathbf{x}\in G_{\mathbf{l}}^N}(P_N((-\mathbf{v})\mathbf{y}|\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{1+\rho}\mathrm{d}\mu^N(\mathbf{y}) \\
&= \int_{\mathcal{Y}^N}\left(\frac{1}{|G_{\mathbf{l}}|^N}\sum_{\mathbf{x}\in G_{\mathbf{l}}^N}(P_N(\mathbf{y}|\mathbf{x}))^{\frac{1}{1+\rho}}\right)^{1+\rho}\mathrm{d}\mu^N(\mathbf{y}) \\
&= \left(\int_{\mathcal{Y}}\left(\frac{1}{|G_{\mathbf{l}}|}\sum_{x\in G_{\mathbf{l}}}(P_N(y|x))^{\frac{1}{1+\rho}}\right)^{1+\rho}\mathrm{d}\mu(y)\right)^N
\end{aligned} \tag{3.33}$$

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

$$\overline{p_e(\Phi_{N,\mathbf{l}},\mathbf{0})}^{(R,\boldsymbol{\alpha})} \leq |H_{\mathbf{K},\mathbf{l}}|^{\rho}\left(\int_{\mathcal{Y}}\left(\frac{1}{|G_{\mathbf{l}}|}\sum_{x\in G_{\mathbf{l}}}(P_N(y|x))^{\frac{1}{1+\rho}}\right)^{1+\rho}\mathrm{d}\mu(y)\right)^N\ .$$

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

$$\overline{p_e(\Phi_{\mathbf{l}}, \mathbf{0})}^{(R,\boldsymbol{\alpha})} \leq \exp(-NE_{\mathbf{l}}(R_{\mathbf{l}}))\ .$$

Now (3.28) follows because $E_{\mathbf{l}}$ is a non increasing function. ■

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

$$E_G(R) = \max_{\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{r}^G)} \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G - \text{comp.}}} E_{\mathbf{l}} \left(R \sum_{i=1}^{s} \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} \right)\ , \tag{3.34}$$

while the G-optimal splitting rate function is defined by letting, for every $R \in [0 \log |G|]$, $\boldsymbol{\alpha}^G(R)$ be one of the elements of $\mathcal{P}(\boldsymbol{r}^G)$ for which the maximum in (3.34) is achieved, i.e.

$$\min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G - \text{comp.}}} E_{\mathbf{l}} \left(R \sum_{i=1}^{s} \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j}^G(R) \right) = \max_{\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{r}^G)} \min_{\substack{\mathbf{l} \neq \mathbf{0} \\ \mathbf{r}^G - \text{comp.}}} E_{\mathbf{l}} \left(R \sum_{i=1}^{s} \sum_{j=1}^{r_i^G} \frac{l_{i,j}}{j} \alpha_{i,j} \right)\ . \tag{3.35}$$

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

We can now state the following result which is an easy consequence of Theorem 13.

Corollary 14 *Consider a G-symmetric memoryless channel of G-capacity C_G, G-random coding exponent $E_G(R)$, G-optimal splitting rate function $\boldsymbol{\alpha}^G(R)$. Then, $E_G(R) > 0$ if and only if $R < C_G$ and*

$$\overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha}^G(R))} \leq A_G \exp(-NE_G(R))\,, \tag{3.36}$$

where

$$A_G = \left| \left\{ \mathbf{l} \neq \mathbf{0}\,, \boldsymbol{r}^G - \text{compatible} \right\} \right| = \sum_{i=1}^{s} \frac{r_i^G (r_i^G + 3)}{2} - 1\,. \tag{3.37}$$

Proof

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

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

Hence,

$$\min_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{comp.}}} E_{\mathbf{l}}\left(R\sum_{i=1}^{s}\sum_{j=1}^{r_i}\frac{l_{i,j}}{j}\alpha_{i,j}\right) > 0 \Leftrightarrow R < \min_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{comp.}}} \frac{C_{\mathbf{l}}}{\sum_{i=1}^{s}\sum_{j=1}^{r_i}\frac{l_{i,j}}{j}\alpha_{i,j}}\,. \tag{3.38}$$

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

$$\min_{\substack{\mathbf{l}\neq\mathbf{0}\\ \mathbf{r}^G-\text{comp.}}} E_{\mathbf{l}}\left(R\sum_{i=1}^{s}\sum_{j=1}^{r_i}\frac{l_{i,j}}{j}\alpha^G_{i,j}\right) > 0 \Leftrightarrow R < C_G\,. \tag{3.39}$$

This clearly implies that if $R < C_G$ then $E_G(R) > 0$. Actually, Theorem 6 immediately implies that

$$E_G(R) > 0 \Leftrightarrow R < C_G\,. \tag{3.40}$$

Using now Theorem 13 we obtain the result.

■

Remark: It follows from the proof of Corollary 14 that using input groups corresponding to the G-optimal splitting $\boldsymbol{\alpha}^G$, we can reach C_G-capacity. However, in order to obtain best mean rate of convergence one has to use input groups corresponding to the splitting $\boldsymbol{\alpha}_G(R)$ which in general is a function of the rate R. Straightforward continuity arguments allow to show that $\boldsymbol{\alpha}^G(R)$ can always be chosen in such a way that $\boldsymbol{\alpha}^G(C_G) = \boldsymbol{\alpha}^G$. Notice that in the cyclic example $G = \mathbb{Z}_{p^r}$ $(s = 1)$ it was already proven that $\alpha^G(R)_r = 1$ and $\alpha^G(R)_j = 0$ if $j < r$. This corresponds to take $\mathcal{U} = \mathbb{Z}_{p^r}^{\lfloor RN \rfloor}$. In other words free input groups over $\mathbb{Z}_{p^r}$ in this case suffice to achieve $\mathbb{Z}_{p^r}$-capacity.

Standard probabilistic considerations allow us to state the following.

Corollary 15 *Consider a G-symmetric memoryless channel of G-capacity C_G, G-random coding exponent $E_G(R)$, and G-optimal splitting rate function $\boldsymbol{\alpha}^G(R)$. The ensemble $\mathcal{E}(R, \boldsymbol{\alpha}^G(R))$ satisfies*

$$P_G\left(\liminf_N \frac{-\log p_e(\Phi_N)}{N} \geq E_G(R)\right) = 1\ , \tag{3.41}$$

where P_G denotes the probability on the ensemble.

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

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

where A_G is defined in (3.37). By applying (3.36) and the Markov inequality to each r.v. $p_e(\Phi_N)$ we obtain

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

Then

$$\sum_{N=1}^{+\infty} P(A_N^\varepsilon) \leq \sum_{N=1}^{+\infty} \exp(-N\varepsilon) < +\infty \ . \tag{3.42}$$

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

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

By Borel Cantelli theorem, (3.42) implies that $P_G(A_N^\varepsilon \text{ i.o.}) = 0$ for every $\varepsilon > 0$. But clearly

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

By the σ-additivity of P_G, this implies

$$P_G\left(\liminf_N \frac{-\log p_e(\Phi_N)}{N} \geq E_G(R)\right) = 1 \ .$$

■

Corollary 16 *Consider a G-symmetric channel whose G-capacity is C_G. Then, for every $R < C_G$ and for every $\varepsilon > 0$, there exists a G-encoder ϕ_G, of rate greater than or equal to R, whose ML decoding word error probability satisfies*

$$p_e(\phi_G) < \varepsilon \ . \tag{3.43}$$

Proof: Trivial consequence of Corollary 15. ■

3.3.3 On tightness of the error exponent

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

First we want to specify what we actually mean by 'tight'. Consider a memoryless channel $(\mathcal{X}, \mathcal{Y}, P)$. It is a well known fact (see [31], [71], [4]) that the random coding exponent in (2.7) is given by

$$E(R) = \begin{cases} R_0 - R\,, & 0 \le R \le R_{cr} \\ E_{sp}(R)\,, & R_{cr} \le R \le C\,. \end{cases} \tag{3.44}$$

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

For the classical Shannon random coding ensemble the error exponent is tight for any deterministic sequence of codes only for $R \ge R_{cr}$, while this is not the case for low rates $R < R_{cr}$: in fact in this case expurgation techniques lead to the existence of sequences of codes guaranteeing higher error exponents. There are conjectures ([71], [26]) about the actual achievable error exponent (the so called reliability function of the channel) at any rate $R \in [0, C]$, but still no completely proved results.

Nevertheless it was proved in [32] that $E(R)$ is tight for the average code from the classical random coding ensemble at any rate, i.e.

$$\lim_{N\to\infty} -\frac{\log \overline{p_e(\Phi_N)}}{N} = E(R)\,, \qquad \forall 0 \le R \le C\,. \tag{3.45}$$

Moreover, when dealing with a channel which is symmetric with respect to the action of a Galois field $\mathbb{F}_q$ (as for instance a binary-input symmetric-output channel), it is well known that (3.45) holds true for the $\mathbb{F}_q$-linear coding ensemble. The proof of this fact, although probably never explicitly published yet ([26]), can be obtained with a slight modification of Gallager's proof in [32]. Indeed, a closer look at [32] shows that the fundamental ingredients of that proof in the special case of $\mathbb{F}_q$-symmetrical channels are uniform distribution of the codewords over $\mathbb{F}_q^N$ and their pairwise independence. As these two properties are preserved when moving from the random coding ensemble to the $\mathbb{F}_q$-linear one, almost the same proof of [32] can be carried on showing that (3.45) continues to hold true in this case.

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

$$\lim_{N\to+\infty} -\frac{\log \overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha}_G(R))}}{N} = E_G(R)\,. \tag{3.46}$$

We have not yet a complete proof of (3.46), but only some partial results, and we want to explain them in the simple special case when $G = \mathbb{Z}_4$.

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

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

where, as usual, $E_2(R)$ and $E_1(R)$ are, respectively, the random coding exponent of the $\mathbb{Z}_4$-symmetric channel, and of its $2\mathbb{Z}_4$-symmetric subchannel. In this case partition (3.26) reduces to

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

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

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

We have, by successively applying the UEP, property (2.3) of ML decoding, and Jensen inequality (notice that function $\mathbb{R}^d \ni \mathbf{x} \mapsto \max_{1\le i\le d} x_i \in \mathbb{R}$ is convex),

$$\begin{aligned}
\overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha}_G(R))} &= \overline{p_e(\Phi_N,\mathbf{0})}^{(R,\boldsymbol{\alpha}_G(R))} \\
&\ge \overline{\max\{p_e(\Phi_{N,1},\mathbf{0}), p_e(\Phi_{N,2},\mathbf{0})\}}^{(R,\boldsymbol{\alpha}_G(R))} \\
&\ge \max\left\{\overline{p_e(\Phi_{N,1},\mathbf{0})}^{(R,\boldsymbol{\alpha}_G(R))}, \overline{p_e(\Phi_{N,2},\mathbf{0})}^{(R,\boldsymbol{\alpha}_G(R))}\right\}.
\end{aligned} \tag{3.47}$$

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

$$\lim_{N\to\infty} \frac{\log \overline{p_e(\Phi_{N,1},\mathbf{0})}^{(R,\boldsymbol{\alpha}_G(R))}}{N} = E_1(R/2)\,, \quad 0 \le R \le C_1\,. \tag{3.48}$$

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

$$\lim_{N\to\infty} \frac{\log \overline{p_e(\Phi_{N,1},\mathbf{0})}^{(R,\boldsymbol{\alpha}_G(R))}}{N} = E_2(R)\,, \quad R_{cr,2} \le R \le C_2\,, \tag{3.49}$$

where $R_{cr,2}$ denotes the critical rate of the $\mathbb{Z}_4$-symmetrical channel By combining (3.47) with (3.48) and (3.49), we obtain that

$$\lim_{N\to\infty} \frac{\log \overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha}_G(R))}}{N} = E_{\mathbb{Z}_4}(R)\,, \quad \text{whenever } E_{\mathbb{Z}_4}(R) = E_1(R) \text{ or } R_{cr,2} \le R \le C_{\mathbb{Z}_4}\,. \tag{3.50}$$

We observe that the first condition in (3.50) is surely holding at very low rates: indeed

$$\lim_{R\to 0} E_1(R/2) = E_1(0) \leq E_2(0) = \lim_{R\to 0} E_2(R)\,, \tag{3.51}$$

with strict inequality holding true in (3.51) for nontrivial channels. So we can conclude that, even for $\mathbb{Z}_4$-symmetric channels for which $C_{\mathbb{Z}_4} = C_4$ so that there is no loss of capacity, there is a loss in the average error exponent at low rates when restricting from Shannon's random coding ensemble to the $\mathbb{Z}_4$-code ensemble.

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

3.3.4 The parity check ensemble

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

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

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

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

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

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

Let $\overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha},')}$ denote the word error probability averaged over this ensemble. Using techniques very similar to those used to upperbound $\overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha})}$ it is possible to prove the following estimation, which constitutes an analogous of Theorem 13.

Theorem 17 *Let $(G, \mathcal{Y}, P)$ a G-symmetric MC. For every $R \in [0, \log|G|[$, $\boldsymbol{\alpha} \in \mathcal{P}(\mathbf{r}^G)$,*

$$\overline{p_e(\Phi_N)}^{(R,\boldsymbol{\alpha},')} \leq \sum_{\substack{\mathbf{l}\neq\mathbf{0} \\ \mathbf{r}^G-\text{compatible}}} \exp\left(-NE_{\mathbf{l}}\left(R_{\boldsymbol{l}}\right)\right)$$

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

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

3.4 $\mathbb{Z}_{p^r}$-codes for p^r-PSK do achieve capacity of the AWGN channel!

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

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

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

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

Definition 18 *A K_m* additive isotropic decreasing noise *(K_m-AIDN) channel is a memoryless channel $(K_m, \mathcal{Y}, P)$ where*

- *$\mathcal{Y} = (Y, \mathcal{B}, \mu)$ where Y is a closed subgroup of $\mathbb{C}^*$ such that $K_m \leq Y$, $\mathcal{B}$ is the corresponding Borel σ-algebra , and μ is a measure over $\mathcal{B}$;*
- *the transition laws $P(y|x)$ only depend on the distance $|x-y|$ and such dependence is monotonically decreasing, i.e. there exists a decreasing function $\zeta : [0, +\infty) \to [0, +\infty)$ such that $P(y|x) = \zeta(|y-x|)$.*

This rather abstract definition allows us to treat at once many different widely used symmetric channels with input K_m and either continuous or discrete output. Notice that from Def.18 it follows that any K_m-AIDN is $\mathbb{Z}_m$-symmetric, since $\mathbb{Z}_m$ is a generating group for K_m, $\mathbb{Z}_m$ isometrically acts on $\mathcal{Y}$ (by rotations), and $P(gy|gx) = \zeta(|gy - gx|) = \zeta(|x-y|) = P(y|x)$. We show some examples of K_m-AIDN channels.

Example 11 *Both the unquantized K_m-AWGN channel and the unquantized K_m-Laplacian channel are AIDN channels.* □

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

Example 12 *Suppose an unquantized K_m-AIDN channel (a K_m-AWGN for instance)*

$$(K_m, \mathbb{C}^*, P) \tag{3.52}$$

is given, and let $\zeta : [0,+\infty) \to [0,+\infty)$ *the decreasing function such that* $P(y|x) = \zeta(|y-x|)$ *(whose existence is guaranteed by the Def.18). For every positive integer* m' *such that* $m \mid m'$*, we can introduce a new, discrete output,* K_m*-AIDN channel by quantizing the output of (3.52) over Voronoi regions of the* $K_{m'}$ *constellation. Explicitly such channel is given by*

$$\left(K_m, K_{m'}, P'\right) \ , \tag{3.53}$$

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

$$P'(\xi^k_{m'}|\xi^j_m) := \int_V P(\xi^k_{m'}t|\xi^j_m)\mathrm{d}\mu(t) = \int_V \zeta\left(|\xi^k_{m'}t - \xi^j_m|\right)\mathrm{d}\mu(t) = \int_V \zeta\left(\left|\xi^{k-j\frac{m'}{m}}_{m'}t - 1\right|\right)\mathrm{d}\mu(t)\ ,$$

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

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

Clearly $K_m \leq K_{m'} \leq \mathbb{C}$*, so that, in order to see that channel (3.53) fulfil Def.18, it remains to show that the second requirement is fulfilled. We start by noticing that for every* $\rho > 0$*,* $\theta \in \mathbb{R}$*,*

$$\begin{aligned}\left|\xi^k_{m'}\rho e^{i\theta} - \xi^j_m\right|^2 &= \left(\rho\cos(\theta + \tfrac{2\pi}{m'}k) - \cos(j\tfrac{2\pi}{m})\right)^2 + \left(\rho\sin(\theta + \tfrac{2\pi}{m'}k) - \sin(j\tfrac{2\pi}{m})\right)^2 \\ &= \rho^2 + 1 - 2\rho\left(\cos(\theta + \tfrac{2\pi}{m'}k)\cos(j\tfrac{2\pi}{m}) + \sin(\theta + \tfrac{2\pi}{m'}k)\sin(j\tfrac{2\pi}{m})\right) \\ &= \rho^2 + 1 + 2\rho\cos\left(\theta + \tfrac{2\pi}{m'}k - j\tfrac{2\pi}{m}\right) \ .\end{aligned} \tag{3.54}$$

From (3.54), and from the fact that $\cos(x) = cos(y)$ *if and only if* $x = \pm y$ *(mod* 2π*), it immediately follows that for every couple of values* $k, k' \in \mathbb{Z}_{m'}$

$$\left|\xi^k_{m'} - \xi^j_m\right| = \left|\xi^{k'}_{m'} - \xi^j_m\right| \iff k - \frac{m'}{m}j = -k' + \frac{m'}{m}j \ \ or \ \ k = k' \ . \tag{3.55}$$

Then for $k \neq k' \in \mathbb{Z}_{m'}$ *such that* $\left|\xi^k_{m'} - \xi^j_m\right| = \left|\xi^{k'}_{m'} - \xi^j_m\right|$*, we have*

$$\begin{aligned}P'(\xi^k_{m'}|\xi^j_m) &= \int_V \zeta\left(\left|\xi^{k-j\frac{m'}{m}}_{m'}t - 1\right|\right)\mathrm{d}\mu(t) \\ &= \int_V \zeta\left(\left|\xi^{-k'+j\frac{m'}{m}}_{m'}t - 1\right|\right)\mathrm{d}\mu(t) \\ &= \int_V \zeta\left(\left|\xi^{-k'+j\frac{m'}{m}}_{m'}t^{-1} - 1\right|\right)\mathrm{d}\mu(t) \\ &= \int_V \zeta\left(\left|\xi^{k'-j\frac{m'}{m}}_{m'}t - 1\right|\right)\mathrm{d}\mu(t) = P'(\xi^{k'}_{m'}|\xi^j_m)\end{aligned} \tag{3.56}$$

where we exploited (3.55), the symmetry property of Voronoi region $V = V^{-1}$, and finally (3.54). Equality (3.56) clearly implies that transition laws $P'(y|x)$ are function of the distance $|x-y|$, i.e. we can define a function $\zeta' : [0,+\infty) \to [0,+\infty)$ such that

$$\zeta'(|\xi^k_{m'} - \xi^j_m|) := \int_V \zeta\left(\left|\xi^{k-j\frac{m'}{m}}_{m'} t - 1\right|\right) \mathrm{d}\mu(t) = P'(\xi^k_{m'}|\xi^j_m) \ ,$$

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

$$|\xi^k_{m'} - \xi^j_m| < |\xi^{k'}_{m'} - \xi^j_m| \implies \zeta'(|\xi^k_{m'} - \xi^j_m|) \geq \zeta'(|\xi^{k'}_{m'} - \xi^j_m|) \ . \tag{3.57}$$

Due to (3.55) it is sufficient to show (3.57) for value of k and k' such that $j\frac{m'}{m} \leq k, k' \leq j\frac{m'}{m} + \frac{m'}{2}$. Notice that if

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

then

$$\begin{aligned} \left|\xi^k_{m'}\rho e^{i\theta} - \xi^j_m\right|^2 &= \rho^2 + 1 + 2\rho\cos\left(\theta + \tfrac{2\pi}{m'}\left(k - j\tfrac{m'}{m}\right)\right) \\ &\geq \rho^2 + 1 + 2\rho\cos\left(\theta + \tfrac{2\pi}{m'}\left(k' - j\tfrac{m'}{m}\right)\right) \\ &= \left|\xi^{k'}_{m'}\rho e^{i\theta} - \xi^j_m\right|^2 \ , \end{aligned}$$

from which it follows that

$$\begin{aligned} P'(\xi^k_{m'}|\xi^j_m) &= \int_V \zeta\left(\left|\xi^k_{m'} t - \xi^j_m\right|\right) \mathrm{d}\mu(t) \\ &\geq \int_V \zeta\left(\left|\xi^{k'}_{m'} t - \xi^j_m\right|\right) \mathrm{d}\mu(t) \\ &= P'(\xi^k_{m'}|\xi^j_m) \end{aligned}$$

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

We conclude our series of examples with the following one.

Example 13 *Let $S^1 = \{e^{i\theta}\} \subset \mathbb{C}$ be the complex unitary circumference. Consider an unquantized K_m-AIDN channel of type (3.52). We define a new channel by projecting the output $\mathbb{C}^*$ onto S^1. Explicitly we consider the channel*

$$\left(K_m, \mathcal{Y} = (S^1, \mathcal{B}', \mu'), \{P(\cdot|x)\}_{x \in K_m} \in \mathcal{P}(\mathcal{Y})\right) \tag{3.58}$$

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

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

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

The verification that (3.58) is an AIDN channel is almost the same as that of Example 12. □

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

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

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

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

Hence, our result is equivalent to say that

$$rC_{p^l} \geq lC_{p^r} \ , \quad \forall l, r : 1 \leq sl \leq r \ . \tag{3.59}$$

Notice that a simple inductive argument shows that (3.59) is equivalent to

$$qC_{p^q+1} \leq (q+1)C_{p^q} \ , \quad \forall q = 1, \dots, r-1 \tag{3.60}$$

The rest of section will be devoted to the proof of (3.60). The result will be achieved through a series of technical intermediate steps. We start with some notation.

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

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

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

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

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

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

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

$$\boldsymbol{\omega}_q(y) := \frac{1}{p^q \lambda_q(y)} \left(P(y|\xi^0_{p^q}), P(y|\xi^1_{p^q}), \ldots, P(y|\xi^{p^q-1}_{p^q}) \right) \ .$$

Notice that:

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

$$\lambda_{q+1}(y) = \frac{1}{p} \sum_{k=1}^{p} \lambda_q \left(y \xi^k_{p^{q+1}} \right) \ . \tag{3.61}$$

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

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

So

$$C_{p^q} = H(\lambda_q) - H\left(P(\cdot|1)\right) \ . \tag{3.62}$$

Therefore (3.60) is equivalent to

$$H\left(P(\cdot|1)\right) + qH(\lambda_{q+1}) \leq (q+1)H(\lambda_q) \ , \quad q = 1, \ldots, r-1 \ . \tag{3.63}$$

Next lemma shows how the entropies of the families of probability laws $\boldsymbol{\omega}_q(y)$ and $\boldsymbol{\nu}_q(y)$ come out in (3.63).

Lemma 19 *For every $q = 1, \ldots, r-1$,*

- $$H(P(\cdot|1)) = H(\lambda_q) - q + \int\limits_{\{\lambda_q > 0\}} \lambda_q(x) H(\boldsymbol{\omega}_q(x)) \mathrm{d}\mu(x) \ ; \tag{3.64}$$

- $$H(\lambda_q) = H(\lambda_{q+1}) - 1 + \int\limits_{\{\lambda_{q+1}>0\}} \lambda_{q+1}(x) H(\boldsymbol{\nu}_q(x)) \mathrm{d}\mu(x) \ . \tag{3.65}$$

Proof:
We have

$$\begin{aligned}
&H\left(P(\cdot|1)\right) = -\int\limits_{\mathcal{Y}} P(y|1)\log P(y|1)\mathrm{d}\mu(y) \\
&= -\tfrac{1}{p^q}\sum_{k=0}^{p^q-1}\int\limits_{\mathcal{Y}} P(y\xi_{p^q}^k|1)\log P(y\xi_{p^q}^k|1)\mathrm{d}\mu(y) = -\frac{1}{p^q}\sum_{k=0}^{p^q-1}\int\limits_{\{\lambda_q>0\}} P(y|\xi_{p^q}^k 1)\log P(y|\xi_{p^q}^k)\mathrm{d}\mu(y) \\
&= -\int\limits_{\{\lambda_q>0\}}\left[\frac{1}{p^q}\sum_{k=0}^{p^q-1} P(y|\xi_{p^q}^k)\right]\log\lambda_q(x)\mathrm{d}\mu(y) - \int\limits_{\{\lambda_q>0\}}\lambda_q(y)\sum_{k=0}^{p^q-1}\frac{P(y|\xi_{p^q}^k)}{p^q\lambda_q(y)}\log\frac{P(y|\xi_{p^q}^k)}{\lambda_q(y)}\mathrm{d}\mu(y) \\
&= -\int\limits_{\{\lambda_q>0\}}\lambda_q(y)\log\lambda_q(y)\mathrm{d}\mu(y) - \int\limits_{\{\lambda_q>0\}}\lambda_q(y)\sum_{k=0}^{p^q-1}(\boldsymbol{\omega}_q(y))_k\log(p^q(\boldsymbol{\omega}_q(y))_k)\mathrm{d}\mu(y) \\
&= H(\lambda_q) - q + \int\limits_{\{\lambda_q>0\}}\lambda_q(y) H\left(\boldsymbol{\omega}_q(y)\right)\mathrm{d}\mu(y) \ ,
\end{aligned} \tag{3.66}$$

and

$$\begin{aligned}
&H(\lambda_q) = -\int\limits_{\mathcal{Y}}\lambda_q(y)\log\lambda_q(y)\mathrm{d}\mu(y) \\
&= -\tfrac{1}{p}\sum_{k=0}^{p-1}\int\limits_{\{\lambda_{q+1}>0\}}\lambda_q(y\xi_{p^{q+1}}^k)\log\lambda_q(y\xi_{p^{q+1}}^k)\mathrm{d}\mu(y) \\
&= -\int\limits_{\{\lambda_{q+1}>0\}}\frac{1}{p}\sum_{k=0}^{p-1}\lambda_q(y\xi_{p^{q+1}}^k)\log\lambda_{q+1}(y)\mathrm{d}\mu(y) - \int\limits_{\{\lambda_{q+1}>0\}}\lambda_{q+1}(y)\sum_{k=0}^{p-1}\frac{\lambda_q(y\xi_{p^{q+1}}^k)}{p\lambda_{q+1}(y)}\log\frac{\lambda_q(y\xi_{p^{q+1}}^k)}{\lambda_{q+1}(y)}\mathrm{d}\mu(y) \\
&= -\int\limits_{\{\lambda_{q+1}>0\}}\lambda_{q+1}(y)\log\lambda_{q+1}(y) - \int\limits_{\{\lambda_{q+1}>0\}}\lambda_{q+1}(y)\sum_{k=0}^{p-1}(\boldsymbol{\nu}_q(y))_k\log(p(\boldsymbol{\nu}_q(y))_k)\mathrm{d}\mu(y) \\
&= H(\lambda_{q+1}) - 1 + \int\limits_{\{\lambda_{q+1}>0\}}\lambda_{q+1}(y) H\left(\boldsymbol{\nu}_q(y)\right)\mathrm{d}\mu(y) \ .
\end{aligned} \tag{3.67}$$

■

Lemma 19 shows that (3.63) is equivalent to

$$q \int\limits_{\{\lambda_{q+1}>0\}} \lambda_{q+1}(y) H\left(\boldsymbol{\nu}_q(y)\right) \mathrm{d}\mu(y) \geq \int\limits_{\{\lambda_q>0\}} \lambda_q(y) H\left(\boldsymbol{\omega}_q(y)\right) \mathrm{d}\mu(y) \ , \quad \forall q = 1, \ldots, r-1 \ . \tag{3.68}$$

We will prove (3.68) by estimating the two entropies appearing in the integrals.

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

$$P_q(y) := \left\{ P(y|\xi^0_{p^q}), P(y|\xi^1_{p^q}), \ldots, P(y|\xi^{p^q-1}_{p^q}) \right\} = \left\{ P(y\xi^0_{p^q}|1), P(y\xi^{p^q-1}_{p^q}|1), \ldots, P(y\xi^1_{p^q}|1) \right\} \ .$$

Notice that

$$P_{q+1}(y) = \left\{ P(y\xi^0_{p^{q+1}}|1), P(y\xi^1_{p^{q+1}}|1), \ldots, P(y\xi^{p^{q+1}-1}_{p^{q+1}}|1) \right\} = \bigcup_{j=0}^{p-1} P_q(y\xi^j_{p^{q+1}}) \ .$$

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

Lemma 20 *For every $1 \leq q < r$ and $y \in \mathcal{Y}$, there is a partition*

$$P_{q+1}(y) = \bigcup_{k=1}^{p^q} P^k_q(y) \ , \qquad P^k_q(y) = \{w^k_{q,0}(y), w^k_{q,1}(y), \ldots, w^k_{q,p-1}(y)\}$$

such that:

- $$w^k_{q,j}(y) \in P_q(\xi^j_{p^q} y) \ , \quad \forall k = 0, \ldots p^q - 1 \ , \ \forall j = 0, \ldots, p-1 \ ;$$
- $$0 \leq k < k' < p^q \implies w^k_{q,i}(y) \geq w^{k'}_{q,j}(y) \ , \ \forall \ i, j = 0, \ldots, p-1 \ . \tag{3.69}$$

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

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

Then

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

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

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

Then

$$\varphi_{j^*} \leq \theta \leq \varphi_{j^*} + \frac{1}{2}\frac{2\pi}{p^{q+1}} \tag{3.70}$$

or

$$\varphi_{j^*} - \frac{1}{2}\frac{2\pi}{p^{q+1}} \leq \theta \leq \varphi_{j^*} \ . \tag{3.71}$$

Suppose that (3.70) holds true. Then

$$\cos(\theta-\varphi_{j^*}) \geq \cos(\theta-\varphi_{j^*+1}) \geq \cos(\theta-\varphi_{j^*-1}) \geq \cos(\theta-\varphi_{j^*+2}) \geq \ldots \geq \cos(\theta-\varphi_{j^*-\lfloor\frac{p^q}{2}\rfloor}) \ . \tag{3.72}$$

From (3.72) it follows that, for odd p,

$$\begin{aligned}
&P_q^0(y) = \left\{P(y|\xi_{p^{q+1}}^{j^*}), P(y|\xi_{p^{q+1}}^{j^*+1}), P(y|\xi_{p^{q+1}}^{j^*-1}), \ldots, P(y|\xi_{p^{q+1}}^{j^*-\lfloor\frac{p}{2}\rfloor})\right\} \\
&P_q^1(y) = \left\{P(y|\xi_{p^{q+1}}^{j^*+\lceil\frac{p}{2}\rceil}), P(y|\xi_{p^{q+1}}^{j^*-\lceil\frac{p}{2}\rceil}), \ldots, P(y|\xi_{p^{q+1}}^{j^*+p})\right\} \\
&\vdots \\
&P_q^{p^q-2}(y) = \left\{P(y|\xi_{p^{q+1}}^{j^*+\lfloor\frac{p^q}{2}\rfloor-p}), P(y|\xi_{p^{q+1}}^{j^*-\lfloor\frac{p^q}{2}\rfloor+p}), \ldots, P(y|\xi_{p^{q+1}}^{j^*+\lfloor\frac{p^q}{2}\rfloor-\lfloor\frac{p}{2}\rfloor})\right\} \\
&P_q^{p^q-1}(y) = \left\{P(y|\xi_{p^{q+1}}^{j^*-\lfloor\frac{p^q}{2}\rfloor+\lfloor\frac{p}{2}\rfloor}), P(y|\xi_{p^{q+1}}^{j^*+\lfloor\frac{p^q}{2}\rfloor-\lceil\frac{p}{2}\rceil}), \ldots, P(y|\xi_{p^{q+1}}^{j^*-\lfloor\frac{p^q}{2}\rfloor})\right\} \ .
\end{aligned} \tag{3.73}$$

But (3.73) implies the desired result since for every k the set of ξ's exponents of the elements of $P_q^k(y)$ contains exactly one element from each equivalence class of integers modulo p.

The cases when (3.71) holds true instead of (3.70), and $p = 2$ are analogous. ■

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

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

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

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

we clearly have

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

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

Consider now the p-adic expansion map

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

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

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

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

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

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

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

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

$$\boldsymbol{\delta}^\alpha_q(y,j) = \frac{1}{p^q \lambda_q(y\xi^j_{p^{q+1}})} \left(\sum_{h=0}^{p^{\alpha-1}} \sum_{\tilde{h}=0}^{p^{q-\alpha-1}-1} w_j^{\tilde{h}p^{\alpha+1}+sp^\alpha+h}(y) \Big| s = 0,\dots,p-1 \right) \ . \tag{3.74}$$

Lemma 21 *For every $1 \le \alpha \le q$,*

$$H\left(\boldsymbol{\omega}_q(y\xi^j_{p^{q+1}})\right) \le \sum_{\alpha=1}^{q} H\left(\boldsymbol{\delta}^\alpha_q(y,j)\right) \tag{3.75}$$

Proof:
We have

$$\begin{aligned} H\left(\boldsymbol{\omega}_q(y\xi^j_{p^{q+1}})\right) &= H\left(\overline{\boldsymbol{\omega}_q}(y,j)\right) = H\left(Z(y,j)\right) = H\left(\theta \circ Z(y,j)\right) = H\left(Y(y,j)\right) \\ &\le \sum_{\alpha=1}^{q} H\left(Y_\alpha(y,j)\right) = \sum_{\alpha=0}^{q-1} H\left(\boldsymbol{\delta}^\alpha_q(y,j)\right) \ , \end{aligned}$$

where we first used the fact that θ is a bijection, then apply chain rule for entropy, and finally the conditional entropy bound (see [13] for instance). ■

Next step of our proof consists in upperbounding the entropies $H(\boldsymbol{\delta}^\alpha_q(x))$ with the entropy $H(\boldsymbol{\nu}_q(x))$, for every $1 \le \alpha \le q < r$ and for every $j \in \{0,\dots,p-1\}$ and $y \in \mathcal{Y}$ such that $\lambda_q(y\xi^j_{p^{q+1}}) > 0$.

We start by stating a simple result characterizing the so called (generalized) 'permutahedron' of a given point in the n-dimensional Euclidean space. Let us introduce some notation. For any $K \subseteq \mathbb{R}^n$ the convex hull of K is defined as the smallest convex subset of $\mathbb{R}^n$ containing K: it will be denoted by $\mathrm{co}(K)$. The set $\mathrm{co}(K)$ can be characterized as the intersection of all the convex sets K' such that $\mathbb{R}^n \supseteq K' \supset K$. A $\mathcal{V}$-polytope in $\mathbb{R}^n$ is the convex hull of finite set $K \subset \mathbb{R}^n$. A $\mathcal{H}$-polytope in $\mathbb{R}^n$ is a bounded intersection of closed halfspaces ($\{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n a_i x_i \leq a_0\}$, $a_i \in \mathbb{R}$ for $0 \leq i \leq n$). Notice that, since every halfspace is convex, then every $\mathcal{H}$-polytope is covex too; moreover it can be easily proved that every $\mathcal{H}$-polythope is the convex hull of its boundary. There is a general fundamental result (see [74] for instance) stating that an arbitrary set $P \subset \mathbb{R}^n$ is a $\mathcal{V}$-polytope if and only if it is an $\mathcal{H}$-polytope: we will therefore simply call it polythope.

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

Lemma 22 *Let $\mathbf{w} \in \mathbb{R}^n$ be such that*

$$w_1 \geq w_2 \geq \ldots \geq w_n \ . \tag{3.76}$$

Then

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

where

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

Proof:
To prove $\mathrm{co}(S_n\mathbf{w}) \subseteq A$ it suffices to note that, for every $\sigma \in S_n$, $\sigma\mathbf{x} \in A$: in fact, it is easy to check that, due to (3.76), every constraint is satisfied. Since A is convex and because of the definition of $\mathrm{co}(S_n\mathbf{w})$ it immediately follows that $\mathrm{co}(S_n\mathbf{w}) \subseteq A$.

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

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

We observe that

$$\pi_J A_J \subseteq B_J \ , \quad \pi_{J^c} A_J \subseteq C_J \ , \tag{3.77}$$

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

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

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

In fact, the former inclusion in (3.77) is trivial since B_J is defined as the intersection of a subset of the halfspaces whose intersection defines A_J, while for the latter it suffices to observe that, for each $I \subset J^c$,

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

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

$$\theta_J \left(\{1, \ldots, |J|\} \right) = J \ ,$$

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

$$B_J \subseteq \mathrm{co}(\pi_J \theta_J S_J \mathbf{w}), \quad C_J \subseteq \mathrm{co}(\pi_{J^c} \theta_J S_{J^c} \mathbf{w}) \ . \tag{3.78}$$

For every $\mathbf{x} \in A_J$ we have $\pi_J \mathbf{x} \in B_J$ and $\pi_{J^c} \mathbf{x} \in C_J$ from (3.77) and then (3.78) implies that $\lambda' \in \mathcal{P}(S_J)$ and $\lambda'' \in \mathcal{P}(S_{J^c})$ exist such that

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

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

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

but then

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

■

Lemma 23 *Suppose n^2 real numbers $\{a_i^k, i,k = 0,\dots,n-1\}$ are given, such that*

$$k < k' \Longrightarrow a_j^k \le a_i^{k'} \ , \qquad j,l = 0,\dots,n-1 \ . \tag{3.79}$$

Define the two vectors

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

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

Proof: (3.79) implies that

$$x_0 \ge x_1 \ge \ldots \ge x_{n-1}$$

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

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

So Lemma 22 can be applied to show that $\mathbf{v} \in \mathrm{co}(S_n\mathbf{x})$. ■

We can now prove the following inequality.

Lemma 24 *For every $0 \le \alpha < q < r$, and $y \in \mathcal{Y}$ such that $\lambda_{q+1}(y) > 0$,*

$$H\left(\sum_{0\le j<p:\lambda_q(y\xi^j_{p^{q+1}})>0} \frac{\lambda_q(y\xi^j_{p^{q+1}})}{p\lambda_{q+1}(y)} \boldsymbol{\delta}_q^\alpha(y,j)\right) \le H\left(\boldsymbol{\nu}_q(y)\right) \tag{3.80}$$

Proof: We will show that

$$\boldsymbol{\nu}_q(y) \in \mathrm{co}\left(S_p\left(\sum_{0\le j<p:\lambda_q(y\xi^j_{p^{q+1}})>0} \frac{\lambda_q(y\xi^j_{p^{q+1}})}{p\lambda_{q+1}(y)} \boldsymbol{\delta}^\alpha_q(y,j) \right)\right) \ . \tag{3.81}$$

Then (3.80) simply follows by the concavity of the entropy function.

From Lemma 20 and from (3.74) it follows that

$$\sum_{0\le j<p:\lambda_q(y\xi^j_{p^{q+1}})>0} \lambda_q(y\xi^j_{p^{q+1}})\boldsymbol{\delta}^\alpha_q(y,j) = \left(\sum_{j=0}^{p-1}\sum_{h=0}^{p^{\alpha-1}}\sum_{\tilde h=0}^{p^{q-\alpha-1}-1} w_j^{sp^\alpha+h+\tilde h p^{\alpha+1}}(y) \ , \ s=0,\dots,p-1 \right)$$

while, by definition,

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

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

$$a^s_j := \sum_{h=0}^{p^{\alpha-1}}\sum_{\tilde h=0}^{p^{q-\alpha-1}-1} w^{sp^\alpha+h+\tilde h p^{\alpha+1}}_{q,j}(y) \ ,$$

then

$$p^q \sum_{0\le j<p:\lambda_q(y\xi^j_{p^{q+1}})>0} \lambda_q(y\xi^j_{p^{q+1}})\boldsymbol{\delta}^\alpha_q(y,j) = \left(\sum_{j=0}^{p-1} a^0_j \sum_{j=0}^{p-1} a^1_j, \dots, \sum_{j=0}^{p-1} a^{p-1}_j \right) \ ,$$

while

$$p^{q+1}\lambda_{q+1}(y)\boldsymbol{\nu}_q(y) = \left(\sum_{s=0}^{p-1} a^s_0, \sum_{s=0}^{p-1} a^s_1, \dots, \sum_{s=0}^{p-1} a^s_{p-1}, \right)$$

Fix a couple $(k,k') \in \{0,\dots,p-1\}$ such that $k<k'$: from (3.69) we have

$$w^{kp^\alpha+h+\tilde h p^{\alpha+1}}_{q,j}(y) \ge w^{k'p^\alpha+h+\tilde h p^{\alpha+1}}_{q,i}(y) \ ,$$

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

$$\begin{aligned} a^k_j &= \sum_{h=0}^{p^{\alpha-1}}\sum_{\tilde h=0}^{p^{q-\alpha-1}-1} w^{kp^\alpha+h+\tilde h p^{\alpha+1}}_j(y) \\ &\le \sum_{h=0}^{p^{\alpha-1}}\sum_{\tilde h=0}^{p^{q-\alpha-1}-1} w^{k'p^\alpha+h+\tilde h p^{\alpha+1}}_i(y) \ = \ a^{k'}_i \ . \end{aligned}$$

So the coefficients $\{a_j^k,\ j,k=0,\dots,p-1\}$ satisfy (3.79) and then Lemma 23 can be applied to conclude that

$$p^{q+1}\lambda_{q+1}(y)\boldsymbol{\nu}_q(y) \in \mathrm{co}\left(S_p\left(p^q \sum_{0\le j<p:\lambda_q(y\xi^j_{p^{q+1}})>0} \lambda_q(y\xi^j_{p^{q+1}})\boldsymbol{\delta}_q^\alpha(y,j)\right)\right)\ , \qquad (3.82)$$

which in turn implies (3.81), since we have supposed $\lambda_{q+1}(y)>0$. ■

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

Theorem 25 *For every $1\le q<r$,*

$$qC_{p^{q+1}} \le (q+1)C_{p^q}\ . \qquad (3.83)$$

Proof: We already noticed that (3.83) is equivalent to

$$q\int\limits_{\{\lambda_{q+1}>0\}} \lambda_{q+1}(y)H\left(\boldsymbol{\nu}_q(y)\right)\mathrm{d}\mu(y) \ge \int\limits_{\{\lambda_q>0\}} \lambda_q(y)H\left(\boldsymbol{\omega}_q(y)\right)\mathrm{d}\mu(y)\ . \qquad (3.84)$$

Fix an arbitrary $y\in\{\lambda_{q+1}(y)>0\}$. Successively applying (3.75), the concavity of the entropy function H and (3.80), we obtain

$$\begin{aligned}
\sum_{\substack{0\le j<p:\\ \lambda_q(y\xi^j_{p^{q+1}})>0}} \frac{\lambda_q(y\xi^j_{p^{q+1}})}{p\lambda_{q+1}(y)}H\left(\boldsymbol{\omega}_q(y\xi^j_{p^{q+1}})\right) &\le \sum_{\alpha=1}^q\left[\sum_j \frac{\lambda_q(y\xi^j_{p^{q+1}})}{p\lambda_{q+1}(y)}H\left(\boldsymbol{\delta}_q^\alpha(y,j)\right)\right]\\
&\le \sum_{\alpha=1}^q H\left(\sum_j \frac{\lambda_q(y\xi^j_{p^{q+1}})}{p\lambda_{q+1}(y)}\boldsymbol{\delta}_q^\alpha(y,j)\right) \qquad (3.85)\\
&\le \sum_{\alpha=1}^q H\left(\boldsymbol{\nu}_q(y)\right)\\
&= qH\left(\boldsymbol{\nu}_q(y)\right)\ .
\end{aligned}$$

Thus

$$\frac{1}{p}\sum_{\substack{0\le j<p:\\ \lambda_q(y\xi^j_{p^{q+1}})>0}} \lambda_q(y\xi^j_{p^{q+1}})H\left(\boldsymbol{\omega}_q(y\xi^j_{p^{q+1}})\right) \le q\lambda_{q+1}(y)H\left(\boldsymbol{\nu}_q(y)\right) \qquad \forall y\in\{\lambda_{q+1}>0\}\ , \qquad (3.86)$$

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

$$\begin{aligned}
\int\limits_{\{\lambda_q>0\}} H\left(\boldsymbol{\omega}_q(y)\right) \mathrm{d}\mu(y) &= \int\limits_{\mathcal{Y}} \lambda_q(y) H\left(\boldsymbol{\omega}_q(y)\right) \mathbb{1}_{\{\lambda_q>0\}}(y) \mathrm{d}\mu(y) \\
&= \tfrac{1}{p} \sum_{j=0}^{p-1} \int\limits_{\mathcal{Y}} \lambda_q(y\xi_{p^{q+1}}^j) H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^k)\right) \mathbb{1}_{\{\lambda_q>0\}}(y\xi_{p^{q+1}}^j) \mathrm{d}\mu(y) \\
&= \int\limits_{\mathcal{Y}} \frac{1}{p} \sum_{\substack{0\leq j<p: \\ \lambda_q(y\xi_{p^{q+1}}^j)>0}} \lambda_q(y\xi_{p^{q+1}}^j) H\left(\boldsymbol{\omega}_q(y\xi_{p^{q+1}}^j)\right) \mathbb{1}_{\{\lambda_{q+1}>0\}}(y) \mathrm{d}\mu(y) \\
&\leq q \int\limits_{\mathcal{Y}} \lambda_{q+1}(y) H\left(\boldsymbol{\nu}_q(y)\right) \mathbb{1}_{\{\lambda_{q+1}>0\}}(y) \mathrm{d}\mu(y) \\
&= q \int\limits_{\{\lambda_{q+1}>0\}} \lambda_{q+1}(y) H\left(\boldsymbol{\nu}_q(y)\right) \mathrm{d}\mu(y) \quad .
\end{aligned} \tag{3.87}$$

■

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

Corollary 26 *For any prime p and positive integer r, every K_{p^r}-AIDN channel is such that*

$$\hat{C}_{\mathbb{Z}_{p^r}} = C_{p^r} \ . \tag{3.88}$$

Combining Corollary 26 with Corollary 16, we can finally state a result first conjectured by Loeliger in [44].

Corollary 27 *$\mathbb{Z}_{p^r}$(-free) codes achieve capacity of the p^r-PSK AWGN channel.*

3.5 An example when $C_G < C$

In the previous section we have shown that for a wide class of $\mathbb{Z}_{p^r}$-symmetric channels with p^r-PSK as input $\mathbb{Z}_{p^r}$-capacity and Shannon capacity do coincide, thus implying by Corollary 16 that $\mathbb{Z}_{p^r}$-codes do suffice to achieve Shannon capacity of such channels. At this point the question arising is whether it is the case for any higher dimensional GU constellation admitting generating group isomorphic to $\mathbb{Z}_{p^r}$. The answer is negative as we will show in this section. In fact we will provide a whole family of counterexamples based on the three-dimensional constellations introduced in Example 7 of Section 2. We will prove that $\mathbb{Z}_{2^r}$-capacity of the AWGN channel with input constrained on some of these constellations is strictly less than the corresponding Shannon capacity, thus

leading to an effective algebraic obstruction to the use of $\mathbb{Z}_{2^r}$-codes. This motivates the investigation of non Abelian group codes for such constellations.

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

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

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

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

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

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

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

$$C_{\mathbb{Z}_{2^r}}(0) = C_{2^r}(0) \ . \tag{3.89}$$

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

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

which coincides with a 3-dimensional embedding of the constellation $\sqrt{\frac{1}{1+\beta^2}} K_{2^{r-1}}$, i.e. the 2^{r-1}-PSK rescaled by the homotopy $x \mapsto \sqrt{\frac{1}{1+\beta^2}} x$. This observation, combined with

the equivalence of AWGN-channels with the same signal to noise ratio, and again the independence of orthogonal components of the Gaussian noise, allows us to apply the results of the previous section to state that

$$(r-1)C_{2^s}(\beta) \geq sC_{2^{r-1}}(\beta)\ ,\quad 1 \leq s \leq r-1\,. \tag{3.90}$$

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

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

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

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

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

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

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

$$\begin{aligned}
\sum_{x\in K^\beta_{2^r}} \tfrac{1}{2^r} P(y|x) \log\left(\frac{P(y|x)}{\frac{1}{2^r}\sum\limits_{z\in K^\beta_{2^r}} P(y|z)}\right) &\leq \sum_{x\in K^\beta_{2^r}} \tfrac{1}{2^r} \sum_{z\in K^\beta_{2^r}} \tfrac{1}{2^r} P(y|x)\log\left(\frac{P(y|x)}{P(y|z)}\right)\\
&= \tfrac{1}{2^{2r}} \sum_{x,z\in K^\beta_{2^r}} P(y|x)\log e\left(-\tfrac{||y-x||^2}{2\sigma^2}+\tfrac{||y-z||^2}{2\sigma^2}\right)\\
&\leq \tfrac{1}{2^{2r}} \sum_{x,z\in K^\beta_{2^r}} P(y|x)\tfrac{\log e}{2\sigma^2}\left(-||y-x||^2+(||y-x||+||z-x||)^2\right)\\
&\leq \tfrac{1}{2^{2r}} \sum_{x,z\in K^\beta_{2^r}} P(y|x)\tfrac{\log e}{2\sigma^2}\left(||y-x||^2+2||x-z||^2\right)\\
&\leq \tfrac{1}{2^{r}} \sum_{x\in K^\beta_{2^r}} P(y|x)\tfrac{\log e}{2\sigma^2}\left(||y-x||^2+8\right)
\end{aligned}$$

where the first inequality is due to the convexity of the function $x \to \log\frac{1}{x}$, the second one to the triangular inequality, the third one comes from the fact that $2ab \leq a^2+b^2$

for every $a, b \in \mathbb{R}$, and the last one from the fact that x and z both lie on a sphere of radius 1, so that $||x - z|| \leq ||x|| + ||z|| \leq 2$. Since

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

we can apply Lebesgue's dominated convergence theorem (see [58]) in order to exchange the limit and the integral signs in evaluating the expressions $\lim_{\beta \to +\infty} C_{2^s(\beta)}$ for any $s \leq r$. By this argument and the continuity of transition densities $P(y|x) = \frac{1}{2\pi\sigma^2} e^{-\frac{||y-x||^2}{2\sigma^2}}$, we get

$$\begin{aligned}
\lim_{\beta \to +\infty} C_{2^r}(\beta) &= \lim_{\beta \to +\infty} \sum_{x \in K_{2^r}^{\beta}} \frac{1}{2^r} \int_{\mathcal{Y}} P(y|x) \log \left(\frac{P(y|x)}{\frac{1}{2^r} \sum\limits_{z \in K_{2^r}^{\beta}} P(y|z)} \right) \mathrm{d}\mu(y) \\
&= \int_{\mathcal{Y}} \frac{1}{2^r} \lim_{\beta \to +\infty} \sum_{x \in K_{2^r}^{\beta}} P(y|x) \log \left(\frac{P(y|x)}{\frac{1}{2^r} \sum\limits_{z \in K_{2^r}^{\beta}} P(y|z)} \right) \mathrm{d}\mu(y) \qquad (3.91) \\
&= \int_{\mathcal{Y}} \frac{1}{2} \sum_{x \in K^{\infty}} P(y|x) \log \left(\frac{P(y|x)}{\frac{1}{2} \sum\limits_{z \in K^{\infty}} P(y|z)} \right) \mathrm{d}\mu(y) = C(\infty)
\end{aligned}$$

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

$$\begin{aligned}
\lim_{\beta \to +\infty} C_{2^s}(\beta) &= \lim_{\beta \to +\infty} \int_{\mathcal{Y}} \frac{1}{2^s} \sum_{j=0}^{2^s-1} P(y|x_{2^{r-s}j}) \log \left(\frac{P(y|x_{2^{r-s}j})}{\frac{1}{2^s} \sum\limits_{k=0}^{2^s-1} P(y|x_{2^{r-s}k})} \right) \mathrm{d}\mu(y) \\
&= \int_{\mathcal{Y}} \frac{1}{2^s} \sum_{j=0}^{2^s-1} \lim_{\beta \to +\infty} P(y|x_{2^{r-s}j}) \log \left(\frac{P(y|x_{2^{r-s}j})}{\frac{1}{2^s} \sum\limits_{k=0}^{2^s-1} P(y|x_{2^{r-s}k})} \right) \mathrm{d}\mu(y) \\
&= \int_{\mathcal{Y}} P(y|(0,0,1)) \log \left(\frac{P(y|(0,0,1))}{P(y|(0,0,1))} \right) \mathrm{d}\mu(y) = 0 \,.
\end{aligned} \qquad (3.92)$$

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

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

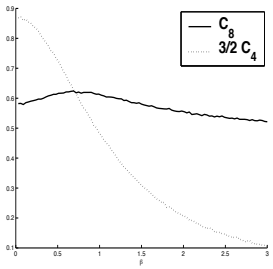

Figure 3.2: Shannon capacity and $\mathbb{Z}_8$-capacity of K_8^β as a function of β

As a consequence we have that

$$C_{\mathbb{Z}_{2^r}}(\beta) = C_{2^r}(\beta) \Longleftrightarrow \beta \leq \overline{\beta} \ . \tag{3.93}$$

As an immediate consequence of Theorem 6, (3.93) tell us that for every $\sigma > 0$ algebraic obstructions surely occur for $\beta > \overline{\beta}(\sigma)$. We can conclude that *for sufficiently high -but finite- values of β $\mathbb{Z}_{2^r}$-codes do not achieve Shannon capacity of the $K_{2^r}^\beta$-AWGN channel of any arbitrary given signal to noise ratio.* We observe that it can be proved that, for $r > 2$,

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

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

Figure 3.2 reports the behaviour of $C_8(\beta)$ and $C_{\mathbb{Z}_8}(\beta)$ as a function of the parameter β (Montecarlo simulations).

Summarizing, in this section we have provided an example of Abelian G-symmetric channel -the $K_{2^r}^\beta$-AWGN channel for $\beta > \overline{\beta}$- for which G-codes are not sufficient to achieve Shannon capacity. It remains an open question whether or not for high values of β the capacity of the $K_{2^R}^\beta$ AWGN-channels can still be achieved by $D_{2^{r-1}}$-codes, i.e. codes which are subgroups of $D_{2^{r-1}}^N$. Our feeling is that it could be possible: it seems to us that, roughly speaking, the structure of the dihedral group is more suitable to be adapted to $K_{2^r}^\beta$ when β goes to infinity, since $D_{2^{r-1}}$ contains a binary subgroup with corresponding subconstellation $\left\{\left(\sqrt{\frac{1}{1+\beta^2}}, \sqrt{\frac{\beta^2}{1+\beta^2}}\right), \left(\sqrt{\frac{1}{1+\beta^2}}e^{\frac{2\pi}{2^r}i}, -\sqrt{\frac{\beta^2}{1+\beta^2}}\right)\right\}$ approaching $K^\infty = \{(0,1)(0,-1)\}$ as β goes to infinity. More in general, one can ask which geometrically uniform constellations S admit eventually non-Abelian generating groups G such that Shannon capacity of the S-AWGN channels can be achieved by G-codes.

3.6 Conclusions

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

Among the still open problems we recall:

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

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

Chapter 4

Typical minimum distances of Abelian group codes

4.1 Introduction

In this chapter, we investigate the minimum Bhattacharyya-distance properties of Abelian group codes on symmetric channels. It is well-known [30] that the typical binary linear code achieves the Gilbert-Varshamov bound on the minimum Hamming distance. It is also known that the typical random code does not achieve the GV bound [4]. Analogous results are known for the typical error exponent: the linear-coding ensemble achieves the so-called expurgated exponent [71] with probability one, while the random coding ensemble is bounded away from it. Therefore, not only is the linear-coding ensemble capacity achieving on any BIOS channel, but in fact it is superior to the random coding ensemble in terms of the performance of the typical code.

We have seen in Chapter 3 that Abelian group codes allow to achieve capacity on a large family of symmetric channels, even if not for all of them. Here we will analyze the typical performance of Abelian group code ensembles on symmetric channels, and in particular study the behaviour of the minimum Bhattacharyya distance. Rather than presenting a general theory, we will focus on specific example, the 8-PSK AWGN channel, containing most of the key ingredients of the general situation. Three different code ensembles will be analyzed: the random coding ensemble, i.e. the set of all possible codes (with no algebraic structure requirement), the $\mathbb{Z}_8$-code ensemble consisting in the set of all subgroups of $\mathbb{Z}_8^n$, and the binary affine code ensemble consisting in the set of all codes which are affine subspaces of $\mathbb{Z}_2^{3n}$. All the three ensembles achieve the Shannon capacity of the channel. While, analogously to the binary case, the random coding ensemble does not asymptotically achieve the GV bound with probability one, we will prove that almost surely a random $\mathbb{Z}_8$-group code sequence achieves the GV bound. We

will also show that almost surely a sequence of binary affine codes has minimum distance asymptotically bounded away from the GV distance.

Analogous results could be obtained for the error exponent which (at low rates) is larger for a typical $\mathbb{Z}_8$-code sequence than it is for a typical binary affine code sequence or for a typical code sequence sampled from the random coding ensemble. This stands in contrast with the results obtained for the average error exponent, which is larger for the random coding ensemble and for the binary affine ensemble than it is for the $\mathbb{Z}_8$-group code ensemble: hierarchies are reversed! The paradox can be easily explained by the fact that the average case analysis only gives a one side estimation of the performance of a typical code (thanks to Markov inequality). However ensemble performance may fail to concentrate around its expected value, and in this case the average case analysis ends up to be too conservative in estimating the error exponent.

The rest of this chapter is organized as follows. In Sect. 4.2 three capacity-achieving ensembles for the 8-PSK AWGN channel are introduced: the random coding ensemble, the $\mathbb{Z}_8$-code ensemble, and the binary affine ensemble. In Sect. 4.3 we analyze the typical asymptotics of normalized minimum distance of the random coding ensemble: these results are standard generalizations of the binary case. In Sect. 4.4 we study the normalized minimum distance of the $\mathbb{Z}_8$-code ensemble: although the results obtained generalize the ones known for the binary linear ensemble, this generalization is non-trivial. In Sect. 4.5 we characterize the asymptotics of the normalized minimum distance of the binary-affine ensemble: the results presented are new, at to our knowledge.

4.2 Three capacity-achieving ensembles for the 8-PSK-AWGN channel

Throughout this chapter we will restrict ourself to considering transmission on the 8-PSK AWGN channel. As usual, since the channel is $\mathbb{Z}_8$-symmetric, we will identify its input $\mathbb{Z}_8$ with $\mathbb{Z}_8$ itself. For every design rate R in $(0, \log 8)$ and N in $\mathbb{N}$, define

$$\overline{R} := \log 8 - R\,, \qquad\qquad L := \left\lfloor \frac{\overline{R}}{\log 8} N \right\rfloor .$$

We will analyze three code ensembles, defined as follows:

1. let $\mathcal{S}_N^R$ be a random subset of $\mathbb{Z}_8^N$, such that the events $\{\boldsymbol{x} \in \mathcal{S}_N^R\}$, for $\boldsymbol{x}$ in $\mathbb{Z}_8^N$, are independent each having probability 8^{-L}; the *random coding ensemble* of rate R is the sequence of random codes $\left(\mathcal{S}_N^R\right)$; [1]

[1] In Chapter 2 the random coding ensemble was constructed in a different way. However, the two constructions are very close to each other and in fact they share all the fundamental properties studied in this chapter.

2. let Φ_N^R be a random variable uniformly distributed over $\hom\left(\mathbb{Z}_8^N, \mathbb{Z}_8^L\right)$, the set of all group homomorphisms from $\mathbb{Z}_8^N$ to $\mathbb{Z}_8^L$. The *cyclic group code ensemble* of rate R is the sequence of random codes $\left(\mathcal{T}_N^R := \ker \Phi_N^R\right)$.

3. let $\eta : \mathbb{Z}_2^3 \to \mathbb{Z}_8$ be an arbitrary bijection; for every N let $\eta_N : \mathbb{Z}_2^{3N} \to \mathcal{X}^N$ denote its componentwise extension. Consider a random variable Ψ_N^R uniformly distributed over $\mathrm{Hom}\left(\mathbb{Z}_2^{3N}, \mathbb{Z}_2^{3L}\right)$, the set of $\mathbb{Z}_2$-linear maps from $\mathbb{Z}_2^{3N}$ to $\mathbb{Z}_2^{3L}$. Moreover, let $\boldsymbol{Z}_N$ be a random variable uniformly distributed over $\mathbb{Z}_2^{3L}$, independent of Ψ_N^R. The *binary affine code ensemble* of rate R is the sequence of random codes $\left(\mathcal{U}_N^R\right)$, where $\mathcal{U}_N^R$ is defined as the image through η_N of the preimage of $\boldsymbol{Z}_N$ through Ψ_N^R: [2]

$$\mathcal{U}_N^R := \eta_N \left(\left(\Psi_N^R\right)^{-1} \boldsymbol{Z}_N \right) . \tag{4.1}$$

The fact that the random coding ensemble as defined above achieves capacity follows from the mutual independence of the events $\left\{\boldsymbol{x} \in \mathcal{S}_N^R\right\}$ for $\boldsymbol{x} \in \mathbb{Z}_8^N$, by an adaptation of the standard averaging random coding arguments. More precisely such an argument allows to show that

$$\mathbb{E}\left[p_e(\mathcal{S}_N^R)\right] \leq \exp(-NE(R)) , \tag{4.2}$$

so that, once fixed a design rate R below the capacity C, the average error probability of the random coding ensemble $\left(\mathcal{S}_N^R\right)$ approaches zero exponentially fast in the blocklength N. Moreover from the independence of the events $S_{\boldsymbol{x}}$ it follows that the arguments of [32] can be applied showing that (4.2) is exponentially tight for the average code, i.e.

$$\lim_{N\in\mathbb{N}} -\frac{1}{N} \log \mathbb{E}\left[p_e(\mathcal{S}_N^R)\right] = E(R) . \tag{4.3}$$

Similar arguments apply to the binary affine ensemble defined by (4.1). Here, for any $\boldsymbol{x}$ in $\mathbb{Z}_8^N$, the event $A_{\boldsymbol{x}} := \left\{\boldsymbol{x} \in \mathcal{U}_N^R\right\}$ has probability 8^{-L}. Moreover, arbitrarily choosing three distinct N-tuples $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ in $\mathbb{Z}_8^N$, it is possible to show that the events $A_{\boldsymbol{x}_i}$ for $i = 1, 2, 3$ are mutually independent. Then, all the considerations made above can be repeated concluding that

$$\mathbb{E}\left[p_e(\mathcal{U}_N^R)\right] \leq \exp(-NE(R)) , \qquad \lim_{N\in\mathbb{N}} -\frac{1}{N} \log \mathbb{E}\left[p_e \mathcal{U}_N^R)\right] = E(R) . \tag{4.4}$$

Finally, the fact that $\mathbb{Z}_8$-code ensemble achieves capacity follows from Theorem 16 of Chapter 2. In particular we have seen that the average error probability can be upperbounded by a term exponentially decreasing in the blocklength

$$p_e(\mathcal{T}_N^R) \leq \exp(-NE_{\mathbb{Z}_8}(R)) .$$

[2] An alternative way to sample the binary affine ensemble consists in considering a r.v. $\boldsymbol{V}_N$ uniformly distributed over $\mathbb{Z}_2^{3N}$ and independent of Ψ_N^R. Then, $\ker \Psi_N^R + \boldsymbol{V}_N$ and $\left(\Psi_N^R\right)^{-1} \boldsymbol{Z}_N$ can be shown to be identically distributed.

The exponent appearing in the righthand side of the above inequality is given by

$$E_{\mathbb{Z}_8}(R) := \min\left\{E_8(R), E_4(\tfrac{2}{3}R), E_2(\tfrac{1}{R})\right\}$$

with $E_4(\frac{2}{3}R)$ and $E_2(\frac{1}{R})$ respectively denoting the random coding error exponents of the AWGN channels with input restricted over the 4-PSK and the 2-PSK modulation.

Thus, the average error exponents of respectively the random kernel ensemble and the binary affine ensemble both equal the random coding ensemble $E_8(R)$, while the error exponent of the $\mathbb{Z}_8$-linear ensemble is given by $E_{\mathbb{Z}_8}(R)$. We have observed in Chapter 3 that, while $E_{\mathbb{Z}_8}(R)$ and $E_8(R)$ do coincide at rates close to capacity, the former is strictly less than the latter at lower rates; in fact it coincides with the exponent $E_2(\frac{1}{3}R)$ of the binary subchannel. In other words, even if algebraic constraints do not affect the capacity achievable by $\mathbb{Z}_8$-codes over the 8-PSK AWGN channel, they do lower the average error exponent achievable by the $\mathbb{Z}_8$-linear ensemble of codes.

If fact, as a consequence of the results presented in the sequel, we will see that the above claim is misleading. Indeed, it refers to the performance of the average code rather than to the performance of the typical code sampled from the three ensembles. While a first order method can always be used to guarantee that the error exponent of a typical code sequence is always bounded from below by the average case error exponent, this bound can fail to be tight. Indeed average performance of code ensembles can be affected by asymptotically vanishing fractions of bad codes. This generally observed phenomenon can cause the average-case analysis to be conservative in estimating error exponents. In this case so called expurgation techniques can be used to obtain the exact error exponent of a typical code sequence.

The main results of this chapter concern the typical asymptotic behaviour of the normalized minimum distance of the three ensembles. In particular we will characterize the almost sure limit of the three random sequences. It will be shown in the following sections that:

- for the random coding ensemble with probability one the sequence $\frac{1}{N}d_{\min}(\mathcal{S}_N^R)$ converges to $\delta_8(2R)$, the Gilbert-Varshamov distance corresponding to twice the design rate;
- for the cyclic group code ensemble with probability one the sequence $\frac{1}{N}d_{\min}(\mathcal{T}_N^R)$ converges to the Gilbert-Varshamov distance $\delta_8(R)$;
- for the binary affine code ensemble (with an arbitrary labeling $\eta : \mathbb{Z}_2^3 \to \mathbb{Z}_8$) with probability one the sequence $\frac{1}{N}d_{\min}(\mathcal{U}_N^R)$ has minimum distance asymptotically strictly between $\delta^{GV}(2R)$ and $\delta^{GV}(R)$ (see Sect.4.5 for a precise characterization).

Therefore hierarchies for typical normalized minimum distances are reversed with respect to those for the average error probability. The typical $\mathbb{Z}_8$-linear code has larger minimum

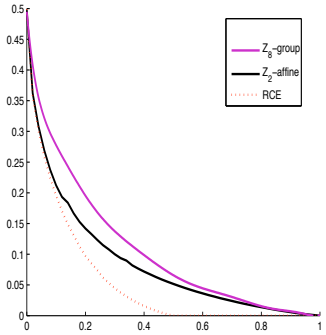

Figure 4.1:

distance than the typical $\mathbb{Z}_2$-affine code, which in turn has larger minimum distance than the typical random code.

4.3 The minimum distance of the typical random code

In this section we will prove that with probability one the normalized minimum Bhattacharyya-distance of the random coding ensemble converges to $\delta^{GV}(2R)$, the GV bound corresponding to twice the rate. Although this is an extension of known results for binary codes [4], we present here a proof, since it allows us to present some techniques which will be used in the following sections.

For any $\boldsymbol{\vartheta}$ in the set of joint types $\mathcal{P}(\mathbb{Z}_8 \times \mathbb{Z}_8)$, we denote the number of ordered pairs of elements of the random code $\mathcal{S}_N^R$ of joint type $\boldsymbol{\vartheta}$ by

$$S_N^R(\boldsymbol{\vartheta}) := \left| (\mathbb{Z}_8 \times \mathbb{Z}_8)_{\boldsymbol{\vartheta}}^N \cap \left(\mathcal{S}_N^R \times \mathcal{S}_N^R\right) \right| .$$

Observe that the minimum distance of the random code $\mathcal{S}_N^R$ is a function of its joint-type enumerating function S_N^R:

$$\begin{aligned} \mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) &= \min\left\{ \boldsymbol{\Delta}(\boldsymbol{x},\boldsymbol{z}) \,\middle|\, \boldsymbol{x} \neq \boldsymbol{z} \in \mathcal{S}_N^R \right\} \\ &= N \min\left\{ \langle \boldsymbol{\vartheta}, \boldsymbol{\Delta} \rangle \,\middle|\, \boldsymbol{\vartheta} \in \mathcal{P}^*(\mathbb{Z}_8 \times \mathbb{Z}_8) \,:\, S_N^R(\boldsymbol{\vartheta}) > 0 \right\} , \end{aligned}$$

where

$$\mathcal{P}^*(\mathbb{Z}_8 \times \mathbb{Z}_8) := \{ \boldsymbol{\vartheta} \in \mathcal{P}(\mathbb{Z}_8 \times \mathbb{Z}_8) \,|\, \exists x \neq z \,:\, \boldsymbol{\vartheta}(x,z) > 0 \}$$

is the set of joint measures whose support is not contained in the diagonal set $\{(x,x)\}$.

It turns out that the typical asymptotic spectrum of the RCE can be characterized as explained in the following series of results. The first of them concerns the average value and variance of the joint-type enumerating function S_N^R. We will use the notation $\pi^1_\# \boldsymbol{\vartheta}(x) = \sum_y \boldsymbol{\vartheta}(x,y)$, $\pi^2_\# \boldsymbol{\vartheta}(y) = \sum_x \boldsymbol{\vartheta}(x,y)$ in $\mathcal{P}(\mathbb{Z}_8)$ for the marginals of a joint measure $\boldsymbol{\vartheta}$ in $\mathcal{P}(\mathbb{Z}_8 \times \mathbb{Z}_8)$.

Lemma 28 *For every $\boldsymbol{\vartheta}$ in $\mathcal{P}_N^* (\mathbb{Z}_8 \times \mathbb{Z}_8)$ we have*

$$\mathbb{E}\left[S_N^R(\boldsymbol{\vartheta})\right] = \binom{N}{N\boldsymbol{\vartheta}} \frac{1}{8^{2L}}\,, \tag{4.5}$$

$$\mathrm{Var}\left[S_N^R(\boldsymbol{\vartheta})\right] \leq \binom{N}{N\boldsymbol{\vartheta}} \frac{2}{8^{2L}} + 2\binom{N}{N\boldsymbol{\vartheta}}^2 \frac{1}{8^{3L}} \left(\binom{N}{N\pi^1_\#\boldsymbol{\vartheta}}^{-1} + \binom{N}{N\pi^2_\#\boldsymbol{\vartheta}}^{-1} \right). \tag{4.6}$$

Proof Fix a pair $(\boldsymbol{x}, \boldsymbol{z})$ in $(\mathbb{Z}_8 \times \mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}$. Since $\boldsymbol{\vartheta}$ is in $\mathcal{P}^* (\mathbb{Z}_8 \times \mathbb{Z}_8)$, it follows that necessarily $\boldsymbol{x} \neq \boldsymbol{z}$. Then, the events $\{\boldsymbol{x} \in \mathcal{S}_N^R\}$ and $\{\boldsymbol{z} \in \mathcal{S}_N^R\}$ are independent so that

$$\mathbb{P}\left(\boldsymbol{x} \in \mathcal{S}_N^R, \boldsymbol{z} \in \mathcal{S}_N^R\right) = \mathbb{P}\left(\boldsymbol{x} \in \mathcal{S}_N^R\right) \mathbb{P}\left(\boldsymbol{z} \in \mathcal{S}_N^R\right) = \frac{1}{8^{2L}}\,.$$

It thus follows that

$$\begin{aligned} \mathbb{E}[S_N^R(\boldsymbol{\vartheta})] &= \mathbb{E}\left[\textstyle\sum_{(\boldsymbol{x},\boldsymbol{z}) \in (\mathbb{Z}_8 \times \mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}} \mathbb{1}_{\{\boldsymbol{x} \in \mathcal{S}_N^R\}} \mathbb{1}_{\{\boldsymbol{z} \in \mathcal{S}_N^R\}}\right] \\ &= \textstyle\sum_{(\boldsymbol{x},\boldsymbol{z}) \in (\mathbb{Z}_8 \times \mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}} \mathbb{P}\left(\boldsymbol{x} \in \mathcal{S}_N^R, \boldsymbol{z} \in \mathcal{S}_N^R\right) \quad = \quad \binom{N}{N\boldsymbol{\vartheta}} \frac{1}{8^{2L}}\,, \end{aligned}$$

showing (4.5). Let us turn to estimate the variance of $S_N^R(\boldsymbol{\vartheta})$. We have

$$\begin{aligned} \mathrm{Var}\left[S_N^R(\boldsymbol{\vartheta})\right] &= \mathrm{Var}\left[\textstyle\sum_{(\boldsymbol{x},\boldsymbol{z}) \in (\mathbb{Z}_8 \times \mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}} \mathbb{1}_{\{\boldsymbol{x} \in \mathcal{S}_N^R, \boldsymbol{z} \in \mathcal{S}_N^R\}}\right] \\ &= \textstyle\sum_{(\boldsymbol{x},\boldsymbol{z}),(\boldsymbol{x}',\boldsymbol{z}') \in (\mathbb{Z}_8 \times \mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}} \mathrm{Cov}\left[\mathbb{1}_{\{\boldsymbol{x} \in \mathcal{S}_N^R, \boldsymbol{z} \in \mathcal{S}_N^R\}}, \mathbb{1}_{\{\boldsymbol{x}' \in \mathcal{S}_N^R, \boldsymbol{z}' \in \mathcal{S}_N^R\}}\right]. \end{aligned}$$

Consider two pairs $(\boldsymbol{x}, \boldsymbol{z})$ and $(\boldsymbol{x}', \boldsymbol{z}')$ in $(\mathbb{Z}_8 \times \mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}$. Since, as already observed $\boldsymbol{x} \neq \boldsymbol{z}$, then

$$2 \leq \left|\{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{x}', \boldsymbol{z}'\}\right| \leq 4\,.$$

If $|\{\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{x}', \boldsymbol{z}'\}| = 4$, then the events $\left\{\boldsymbol{x} \in \mathcal{S}_N^R, \boldsymbol{z} \in \mathcal{S}_N^R\right\}$ and $\left\{\boldsymbol{x}' \in \mathcal{S}_N^R, \boldsymbol{z}' \in \mathcal{S}_N^R\right\}$ are independent and therefore

$$\mathrm{Cov}\left[\mathbb{1}_{\{\boldsymbol{x} \in \mathcal{S}_N^R, \boldsymbol{z} \in \mathcal{S}_N^R\}}, \mathbb{1}_{\{\boldsymbol{x}' \in \mathcal{S}_N^R, \boldsymbol{z}' \in \mathcal{S}_N^R\}}\right] = 0\,.$$

If instead $|\{\boldsymbol{x},\boldsymbol{z},\boldsymbol{x}',\boldsymbol{z}'\}|=2$, then necessarily $\boldsymbol{x}'=\boldsymbol{x}$ and $\boldsymbol{z}=\boldsymbol{z}'$ or $\boldsymbol{x}'=\boldsymbol{z}$ and $\boldsymbol{z}=\boldsymbol{x}'$. In both cases

$$\mathrm{Cov}\left[\mathbb{1}_{\{\boldsymbol{x}\in\mathcal{S}_N^R,\boldsymbol{z}\in\mathcal{S}_N^R\}},\mathbb{1}_{\{\boldsymbol{x}'\in\mathcal{S}_N^R,\boldsymbol{z}'\in\mathcal{S}_N^R\}}\right]=\mathrm{Var}\left[\mathbb{1}_{\{\boldsymbol{x}\in\mathcal{S}_N^R,\boldsymbol{z}\in\mathcal{S}_N^R\}}\right]=\frac{1}{8^{2L}}\left(1-\frac{1}{8^{2L}}\right).$$

As there are at most $2\binom{N}{N\boldsymbol{\vartheta}}$ such choices of pairs $(\boldsymbol{x},\boldsymbol{z})$ and $(\boldsymbol{x}',\boldsymbol{z}')$ in $(\mathbb{Z}_8\times\mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}$, their contribution is taken into account by the first summand in the righthand side of (4.6). Finally, we consider pairs $(\boldsymbol{x},\boldsymbol{z})$ and $(\boldsymbol{x}',\boldsymbol{z}')$ such that $|\{\boldsymbol{x},\boldsymbol{z},\boldsymbol{x}',\boldsymbol{z}'\}|=3$. In this case

$$\mathrm{Cov}\left[\mathbb{1}_{\{\boldsymbol{x},\boldsymbol{z}\in\mathcal{S}_N^R\}},\mathbb{1}_{\{\boldsymbol{x}',\boldsymbol{z}'\in\mathcal{S}_N^R\}}\right]\leq\mathbb{P}\left(\boldsymbol{x},\boldsymbol{z},\boldsymbol{x}',\boldsymbol{z}'\in\mathcal{S}_N^R\right)=\frac{1}{8^{3L}}.$$

We claim that there are at most $2\binom{N}{N\boldsymbol{\vartheta}}^2\left(\binom{N}{N\pi^1_\#\boldsymbol{\vartheta}}^{-1}+\binom{N}{N\pi^2_\#\boldsymbol{\vartheta}}^{-1}\right)$ such choices of pairs $(\boldsymbol{x},\boldsymbol{z})$ and $(\boldsymbol{x}',\boldsymbol{z}')$ in $(\mathbb{Z}_8\times\mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}$. Indeed, once fixed $(\boldsymbol{x},\boldsymbol{z})$ (there are $\binom{N}{N\boldsymbol{\vartheta}}$ different possibilities), for $|\{\boldsymbol{x},\boldsymbol{z},\boldsymbol{x}',\boldsymbol{z}'\}|=3$ it is necessary that either $\boldsymbol{x}'=\boldsymbol{x}$, or $\boldsymbol{x}'=\boldsymbol{z}$, or $\boldsymbol{z}'=\boldsymbol{x}$, or $\boldsymbol{z}'=\boldsymbol{z}$. There are $\binom{N}{N\boldsymbol{\vartheta}}\binom{N}{N\pi^1_\#\boldsymbol{\vartheta}}^{-1}$ different choices of $(\boldsymbol{x}',\boldsymbol{z}')$ in $(\mathbb{Z}_8\times\mathbb{Z}_8)^N_{\boldsymbol{\vartheta}}$ with $\boldsymbol{x}'=\boldsymbol{x}$, $\binom{N}{N\boldsymbol{\vartheta}}\binom{N}{N\pi^2_\#\boldsymbol{\vartheta}}^{-1}$ choices with $\boldsymbol{z}'=\boldsymbol{z}$, at most $\binom{N}{N\boldsymbol{\vartheta}}\binom{N}{N\pi^2_\#\boldsymbol{\vartheta}}^{-1}$ with $\boldsymbol{z}'=\boldsymbol{x}$ and at most $\binom{N}{N\boldsymbol{\vartheta}}\binom{N}{N\pi^1_\#\boldsymbol{\vartheta}}^{-1}$ with $\boldsymbol{x}'=\boldsymbol{z}$. This justifies the second summand in the righthand side of (4.6). ■

Now, a first-order method allows us to prove that, almost surely, pairs of codewords of any joint type $\boldsymbol{\theta}$ whose entropy is below $2\overline{R}$ exist only for finitely many values of N.

Proposition 29 *For every rate R in $(0,\log 8)$ and $\varepsilon>0$ we have that with probability one there exists some N_0 in $\mathbb{N}$ such that*

$$S_N^R(\boldsymbol{\vartheta})=0\,,\qquad\forall\,\boldsymbol{\vartheta}\in\mathcal{P}^*\left(\mathbb{Z}_8\times\mathbb{Z}_8\right):\,\mathrm{H}(\boldsymbol{\vartheta})\leq 2\overline{R}-\varepsilon\,,\qquad\forall\,N\geq N_0\,.\tag{4.7}$$

Proof Consider a type $\boldsymbol{\vartheta}$ in $\mathcal{P}^*\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)$ such that $\mathrm{H}(\boldsymbol{\vartheta})\leq 2\overline{R}$. From (4.5) it follows that for every N

$$\mathbb{E}\left[S_N^R(\boldsymbol{\vartheta})\right]\leq\binom{N}{N\boldsymbol{\vartheta}}\frac{1}{8^{2L}}\leq\exp\left(N\left(\mathrm{H}(\boldsymbol{\vartheta})-2\overline{R}\right)\right)\leq\exp(-\varepsilon N)\,,$$

with the first inequality above holding as an equality iff $\boldsymbol{\vartheta}$ belongs to $\mathcal{P}_N\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)$. For every N define the events

$$A_N^{\boldsymbol{\vartheta}}:=\left\{S_N^R(\boldsymbol{\vartheta})\geq 1\right\},\qquad A_N^{\varepsilon}:=\bigcup_{\mathrm{H}(\boldsymbol{\vartheta})\leq 2\overline{R}-\varepsilon}A_N^{\boldsymbol{\vartheta}}\,.$$

We have

$$\mathbb{P}\left(A_N^\varepsilon\right) \le \sum_{\mathrm{H}(\boldsymbol{\vartheta})\le 2\overline{R}-\varepsilon} \mathbb{P}\left(A_N^{\boldsymbol{\vartheta}}\right) \le \sum_{\mathrm{H}(\boldsymbol{\vartheta})\le 2\overline{R}-\varepsilon} \mathbb{E}\left[S_N^R(\boldsymbol{\vartheta})\right] \le \left|\mathcal{P}_N\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)\right| \exp\left(-\varepsilon N\right) ,$$

the first inequality above following by a union bound estimation, the second one from Markov inequality, the third one from (4.5). Since $\left|\mathcal{P}_N\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)\right|$ grows polynomially fast in N, we have

$$\textstyle\sum_N \mathbb{P}\left(A_N^\varepsilon\right) \le \sum_N \left|\mathcal{P}_N\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)\right| \exp(-\varepsilon N) < +\infty\,.$$

Then, an application of Borel-Cantelli lemma implies that with probability one the event A_N^ε occurs only for finitely many values of N in $\mathbb{N}$. ■

The following can be considered as a partial converse to Proposition 29. Its proof is based on an application of the second-moment method.

Proposition 30 *For a design rate R in $(0,\log 8)$, let $\boldsymbol{\vartheta}$ in $\mathcal{P}\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)$ satisfy*

$$\mathrm{H}(\boldsymbol{\vartheta}) > 2\overline{R}\,, \qquad \mathrm{H}(\pi^1_\#\boldsymbol{\vartheta}) > \overline{R}\,, \qquad \mathrm{H}(\pi^2_\#\boldsymbol{\vartheta}) > \overline{R}\,. \tag{4.8}$$

Then for every sequence $(\boldsymbol{\vartheta}_N)$ in $\mathcal{P}\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)$ converging to $\boldsymbol{\vartheta}$, with $\boldsymbol{\vartheta}_N$ in $\mathcal{P}_N\left(\mathbb{Z}_8\times\mathbb{Z}_8\right)$ for every N, with probability one

$$\exists\, N_0\in\mathbb{N}: \qquad S_N^R(\boldsymbol{\vartheta}_N)\ge 1\,, \qquad \forall\, N\ge N_0\,. \tag{4.9}$$

Proof From Chebyshev inequality and Lemma 28 it follows that

$$\mathbb{P}\left(S_N^R(\boldsymbol{\vartheta}_N)=0\right) \le \frac{\mathrm{Var}\left[S_N^R(\boldsymbol{\vartheta}_N)\right]}{\mathbb{E}\left[S_N^R(\boldsymbol{\vartheta}_N)\right]^2} \le 2\frac{8^{2L}}{\binom{N}{N\boldsymbol{\vartheta}}} + 2\frac{8^{2L}}{\binom{N}{N\pi^1_\#\boldsymbol{\vartheta}}} + 2\frac{8^{2L}}{\binom{N}{N\pi^2_\#\boldsymbol{\vartheta}}}\,.$$

Then

$$\begin{aligned}
\limsup_{N\in\mathbb{N}} \frac{1}{N}\log\mathbb{P}\left(S_N^R(\boldsymbol{\vartheta}_N)=0\right) &\le \limsup_{N\in\mathbb{N}} \frac{1}{N}\log\left(\frac{8^{2L}}{\binom{N}{N\boldsymbol{\vartheta}_N}} + \frac{8^{L}}{\binom{N}{N\pi^1_\#\boldsymbol{\vartheta}_N}} + \frac{8^{L}}{\binom{N}{N\pi^2_\#\boldsymbol{\vartheta}_N}}\right)\\
&= \max\left\{2\overline{R}-\mathrm{H}(\boldsymbol{\vartheta})\,,\ \overline{R}-\mathrm{H}\left(\pi^1_\#\boldsymbol{\vartheta}\right)\,,\ \overline{R}-\mathrm{H}\left(\pi^2_\#\boldsymbol{\vartheta}\right)\right\}\\
&< 0\,.
\end{aligned}$$

Therefore $\sum_N \mathbb{P}(S_N^R(\boldsymbol{\vartheta}_N)=0) < +\infty$, and Borel-Cantelli lemma implies that (4.9) holds with probability one. ■

We are now ready to prove the main result of this section.

Theorem 31 *For every design rate R in $(0,\log 8)$*

$$\mathbb{P}\left(\lim_{N\in\mathbb{N}} \frac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) = \delta^{GV}(2R)\right) = 1 \tag{4.10}$$

Proof For every $\varepsilon > 0$, it follows from Proposition 29 that with probability one there exists N_0 such that

$$\mathrm{d}_{\min}(\mathcal{S}_N^R) \geq N \min\left\{\boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_8 \times \mathbb{Z}_8\right) : \mathrm{H}(\boldsymbol{\vartheta}) \geq 2(\overline{R} - \varepsilon)\right\}, \qquad \forall\, N \geq N_0\,.$$

Therefore,

$$\mathbb{P}\left(\liminf_{N\in\mathbb{N}} \frac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) \geq \delta^{GV}(2R+\varepsilon)\right) = 1\,, \qquad \forall\, \varepsilon > 0\,.$$

It thus follows from the monotonicity and continuity properties of $\delta^{GV}(R)$ that

$$\begin{aligned}
\mathbb{P}\left(\liminf_{N\in\mathbb{N}} \tfrac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) \geq \delta^{GV}(2R)\right) &= \mathbb{P}\left(\liminf_{N\in\mathbb{N}} \tfrac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) \geq \lim_{k\in\mathbb{N}} \delta^{GV}(2R+\tfrac{1}{k})\right) \\
&= \mathbb{P}\left(\bigcap_{k\in\mathbb{N}} \left\{\liminf_{N\in\mathbb{N}} \tfrac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) \geq \delta^{GV}(2R+\tfrac{1}{k})\right\}\right) \\
&= \lim_{k\in\mathbb{N}} \mathbb{P}\left(\liminf_{N\in\mathbb{N}} \tfrac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) \geq \delta^{GV}(2R+\tfrac{1}{k})\right) \\
&= 1\,.
\end{aligned}$$

Let us now turn our attention to proving that

$$\mathbb{P}\left(\limsup_{N\in\mathbb{N}} \frac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{S}_N^R\right) \leq \delta^{GV}(2R)\right) = 1\,. \tag{4.11}$$

Fix an arbitrary $\varepsilon > 0$, and let $\boldsymbol{\theta}_\varepsilon$ in $\mathcal{P}\left(\mathbb{Z}_8\right)$ be such that

$$\delta^{GV}(2R-\varepsilon) = \langle \boldsymbol{\theta}_\varepsilon, \boldsymbol{\delta}\rangle\,, \qquad \mathrm{H}(\boldsymbol{\theta}_\varepsilon) \geq \log 8 - 2R + \varepsilon\,.$$

Then, define $\boldsymbol{\vartheta}_\varepsilon$ in $\mathcal{P}\left(\mathbb{Z}_8 \times \mathbb{Z}_8\right)$ by

$$\boldsymbol{\vartheta}_\varepsilon(x,w) := \frac{1}{8}\,\boldsymbol{\theta}_\varepsilon(w-x)\,, \qquad \forall\, w,x \in \mathbb{Z}_8\,.$$

It is easy to verify that both the marginals $\pi^1_\#\boldsymbol{\vartheta}_\varepsilon$ and $\pi^2_\#\boldsymbol{\vartheta}_\varepsilon$ coincide with the uniform distribution over $\mathbb{Z}_8$, while the conditioned measures are shifted versions of $\boldsymbol{\theta}_\varepsilon$:

$$\boldsymbol{\vartheta}_\varepsilon|_{\{x\}\times\mathbb{Z}_2^3}(\cdot) = \boldsymbol{\theta}_\varepsilon\left(\cdot - x\right)\,.$$

Hence,

$$\mathrm{H}\left(\pi^1_\#\boldsymbol{\vartheta}_\varepsilon\right) = \mathrm{H}\left(\pi^2_\#\boldsymbol{\vartheta}_\varepsilon\right) = \log 8 \geq \overline{R} + \varepsilon\,,$$

and from (7.3) we have

$$\begin{aligned} \mathrm{H}\left(\boldsymbol{\vartheta}_\varepsilon\right) &= \mathrm{H}\left(\pi^1_\#\boldsymbol{\vartheta}_\varepsilon\right) + \textstyle\sum_{x\in\mathbb{Z}_8}[\pi^1_\#\boldsymbol{\vartheta}_\varepsilon](x)\,\mathrm{H}\left(\boldsymbol{\vartheta}_\varepsilon|_{(\pi^1)^{-1}(x)}\right) \\ &= \log 8 + \mathrm{H}\left(\boldsymbol{\theta}_\varepsilon\right) \\ &\geq 2\log 8 - 2R + \varepsilon \\ &= 2\overline{R} + \varepsilon\,. \end{aligned}$$

Since $\mathcal{P}_{\mathbb{N}}\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)$ is dense in $\mathcal{P}\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)$, there exists a sequence of joint types $(\boldsymbol{\vartheta}_N)$ converging to $\boldsymbol{\vartheta}_\varepsilon$ and such that, for every N, $\boldsymbol{\vartheta}_N$ belongs to $\mathcal{P}_N\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)$. It follows from Proposition 30 that with probability one $S_N^R(\boldsymbol{\vartheta}_N)\geq 1$ definitively in N. Moreover

$$\begin{aligned} \langle\boldsymbol{\vartheta}_\varepsilon,\boldsymbol{\Delta}\rangle &= \textstyle\sum_{x,z\in\mathbb{Z}_8}\boldsymbol{\vartheta}_\varepsilon\,(x,z)\,\boldsymbol{\Delta}\,(x,z) \\ &= \textstyle\sum_{x,z\in\mathbb{Z}_8}\frac{1}{8}\boldsymbol{\theta}_\varepsilon(z-x)\boldsymbol{\delta}(z-x) \\ &= \langle\boldsymbol{\theta}_\varepsilon,\boldsymbol{\delta}\rangle \\ &= \delta^{GV}(2R-\varepsilon)\,. \end{aligned}$$

Therefore

$$\mathbb{P}\left(\limsup_{N\in\mathbb{N}}\frac{1}{N}d_{\min}(\mathcal{S}_N^R)\leq\delta^{GV}(2R-\varepsilon)\right)=\mathbb{P}\left(\limsup_{N\in\mathbb{N}}\frac{1}{N}d_{\min}(\mathcal{S}_N^R)\leq\lim_{N\in\mathbb{N}}\langle\boldsymbol{\vartheta}_N,\boldsymbol{\Delta}\rangle\right)=1\,.$$

Then (4.11) follows from the arbitrariness of $\varepsilon>0$ and the continuity of the GV-distance $\delta^{GV}(2R)$ as a function of R. ■

4.4 Minimum distance of the typical $\mathbb{Z}_8$-code

In this section we will prove that the normalized minimum Bhattacharyya-distance of the $\mathbb{Z}_8$-code ensemble converges almost surely to the GV distance. In other words, a typical $\mathbb{Z}_8$-code sequence asymptotically meets the Gilbert-Varshamov bound. In Sect. 4.4.1 we will use a first order method to prove that the GV distance is a lower bound on the asymptotics of a typical realization of the sequence $\left(\frac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{T}_N^R\right)\right)$. In Sect. 4.4.2 instead, we will apply a second order method in order to show that the GV distance also gives an almost sure upper bound on the lim sup of the same random sequence.

Recall from Section 2.5 that the minimum distance of a $\mathbb{Z}_8$-code is a function of its type-enumerating function. If we denote the type-enumerating functions of the $\mathbb{Z}_8$-code ensemble by

$$T_N^R:\mathcal{P}(\mathbb{Z}_8)\to\mathbb{Z}^+\,,\qquad T_N^R(\boldsymbol{\theta}):=\left|(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}\cap\ker\Phi_N^R\right|\,.$$

we have

$$\mathrm{d}_{\min}\left(\mathcal{T}_N^R\right)=N\min\left\{\langle\boldsymbol{\theta},\boldsymbol{\delta}\rangle\;\middle|\;\boldsymbol{\theta}\in\mathcal{P}(\mathbb{Z}_8)\setminus\{\delta_0\}\;:\;T_N^R(\boldsymbol{\theta})\geq 1\right\}\,.$$

4.4.1 A lower bound on the typical asymptotic minimum distance

A first step in our analysis consists in evaluating the average value and the variance of the random variables $S_N^R(\boldsymbol{\theta})$. It is convenient to introduce the following function:

$$l : \mathcal{P}(\mathbb{Z}_8) \to \mathbb{R}\,, \qquad l(\boldsymbol{\theta}) := \frac{8}{\gcd(\mathrm{supp}(\boldsymbol{\theta}))}\,. \tag{4.12}$$

In other words for a $\mathbb{Z}_8$-type $\boldsymbol{\theta}$, $l(\boldsymbol{\theta})$ equals the order of the smallest subgroup of $\mathbb{Z}_8$ supporting $\boldsymbol{\theta}$. Observe that the function l takes values only on the set of divisors of 8. Moreover it is lower semicontinuous since it jumps to lower values when approaching $\mathbb{Z}_8$-types supported on smaller subgroups of $\mathbb{Z}_8$.

The following result motivates definition (4.12).

Lemma 32 *For any $\boldsymbol{x}$ in $\mathbb{Z}_8^N$, the random variable $\Phi_N^R\boldsymbol{x}$ is uniformly distributed over the subgroup $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L$, with $\boldsymbol{\theta} = \boldsymbol{\upsilon}_{\mathbb{Z}_8}(\boldsymbol{x})$ denoting the type of $\boldsymbol{x}$.*

Proof Let (δ_i) for $i = 1, \ldots, N$ be the canonical basis of $\mathbb{Z}_8^N$. If we write $\boldsymbol{x} = \sum_{i=1}^N x_i\delta_i$, we have that x_i belongs to $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8$ for every i, and there exists some i^* such that $gcd(x_{i^*}, 8) = \frac{8}{l(\boldsymbol{\theta})}$. Since $\left\{\Phi_N^R\delta_i\,,\ 1 \le i \le N\right\}$ is a collection of i.i.d. random variables uniformly distributed over $\mathbb{Z}_8^L$, it follows that $x_{i^*}\Phi_N^R\delta_{i^*}$ is uniformly distributed over $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L$, and is independent from the r.v. $\sum_{i\neq i^*}\delta_i\Phi_N^R e_i$ which itself takes values in $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L$. Then, for every $\boldsymbol{z}$ in $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L$ we have

$$\begin{aligned}\mathbb{P}(\Phi_N^R\boldsymbol{x} = \boldsymbol{z}) &= \textstyle\sum_{\boldsymbol{y}\in\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L}\mathbb{P}\left(x_{i^*}\Phi_N^R\delta_{i^*} = \boldsymbol{z}-\boldsymbol{y}\,,\ \Sigma_{i\neq i^*}x_i\Phi_N^R\delta_i = \boldsymbol{y}\right)\\ &= \textstyle\sum_{\boldsymbol{y}\in\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L}\frac{1}{l(\boldsymbol{\theta})^L}\mathbb{P}\left(\Sigma_{i\neq i^*}x_i\Phi_N^R\delta_i = \boldsymbol{y}\right) = \frac{1}{l(\boldsymbol{\theta})^L}\,,\end{aligned}$$

which shows that $\Phi_N^R\boldsymbol{x}$ is uniformly distributed over $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8^L$. ■

The following is an immediate consequence of Lemma 32.

Lemma 33 *For every design rate R in $(0, \log R)$ and $\boldsymbol{\theta} \neq \delta_0$ in $\mathcal{P}_N(\mathbb{Z}_8)$, we have*

$$\mathbb{E}\left[T_N^R(\boldsymbol{\theta})\right] = \binom{N}{N\boldsymbol{\theta}}\frac{1}{l(\boldsymbol{\theta})^L}\,.$$

Proof By the linearity of the average operator

$$\mathbb{E}\left[T_N^R(\boldsymbol{\theta})\right] = \mathbb{E}\sum_{\boldsymbol{x}\in(\mathbb{Z}_8)_{\boldsymbol{\theta}}^N}\mathbb{1}_{\{\Phi_N^R\boldsymbol{x}=\boldsymbol{0}\}} = \sum_{\boldsymbol{x}\in(\mathbb{Z}_8)_{\boldsymbol{\theta}}^N}\mathbb{P}\left(\Phi_N^R\boldsymbol{x} = \boldsymbol{0}\right) = \frac{\left|(\mathbb{Z}_8)_{\boldsymbol{\theta}}^N\right|}{(l(\boldsymbol{\theta}))^L} = \binom{N}{N\boldsymbol{\theta}}\frac{1}{(l(\boldsymbol{\theta}))^L}$$

■

Observe that as a consequence of Lemma 33 we have

$$\mathbb{E}\left[T_N^R(\boldsymbol{\theta})\right] \leq \exp\left(N\left[\mathrm{H}(\boldsymbol{\theta}) - \frac{\log l(\boldsymbol{\theta})}{\log 8}\overline{R}\right]\right), \tag{4.13}$$

$$\lim_{N\in\mathcal{N}_{\boldsymbol{\theta}}} \frac{1}{N}\log \mathbb{E}\left[T_N^R(\boldsymbol{\theta})\right] = \mathrm{H}(\boldsymbol{\theta}) - \frac{\log l(\boldsymbol{\theta})}{\log 8}\overline{R},$$

for any $\mathbb{Z}_8$-type $\boldsymbol{\theta}$ in $\mathcal{P}_{\mathbb{N}}(\mathbb{Z}_8)$. For every t in $[0, \log 8]$ define the set

$$A_t := \left\{\boldsymbol{\theta} \in \mathcal{P}(\mathbb{Z}_8) \,:\, \mathrm{H}(\boldsymbol{\theta}) - \frac{\log l(\boldsymbol{\theta})}{\log 8} t \geq 0\right\}.$$

From the continuity of the entropy function $\mathrm{H}(\boldsymbol{\theta})$ and the lower semicontinuity of $l(\boldsymbol{\theta})$ it follows that A_t is closed in $\mathcal{P}(\mathbb{Z}_8)$. Moreover it is nonempty for every $t \leq \log 8$. Therefore, Lemma 64 can be applied showing that the function $t \mapsto \min\left\{\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle \,\middle|\, \boldsymbol{\theta} \in A_t\right\}$ is lower semicontinuous over the interval $[0, \log 8]$. Furthermore, observe that from (4.12), for every design rate R in $(0, \log 8)$ we have

$$\begin{aligned} A_{\overline{R}} &= \left\{l(\boldsymbol{\theta}) = 8\,, \mathrm{H}(\boldsymbol{\theta}) \geq \overline{R}\right\} \cup \left\{l(\boldsymbol{\theta}) = 4\,, \mathrm{H}(\boldsymbol{\theta}) \geq \tfrac{2}{3}\overline{R}\right\} \cup \left\{l(\boldsymbol{\theta}) = 2\,, \mathrm{H}(\boldsymbol{\theta}) \geq \tfrac{1}{3}\overline{R}\right\} \\ &\subseteq \left\{\mathrm{H}(\boldsymbol{\theta}) \geq \overline{R}\right\} \cup \left\{\mathrm{supp}(\boldsymbol{\theta}) \subseteq 2\mathbb{Z}_8\,, \mathrm{H}(\boldsymbol{\theta}) \geq \tfrac{2}{3}\overline{R}\right\} \cup \left\{\mathrm{supp}(\boldsymbol{\theta}) \subseteq 4\mathbb{Z}_8\,, \mathrm{H}(\boldsymbol{\theta}) \geq \tfrac{\overline{R}}{3}\right\}. \end{aligned}$$

It follows that

$$\min\left\{\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle \,|\, \boldsymbol{\theta} \in A_{\overline{R}}\right\} \geq \min\left\{\delta_8(R), \delta_4(\tfrac{2}{3}R), \delta_2(\tfrac{1}{3}R)\right\}, \tag{4.14}$$

where $\delta^{GV}(R)$ is the 8-PSK GV distance, while

$$\delta_4\left(\tfrac{2}{3}R\right) := \min\left\{\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle \,\middle|\, \mathrm{supp}(\boldsymbol{\theta}) \subseteq 2\mathbb{Z}_8\,,\ \mathrm{H}(\boldsymbol{\theta}) \geq \tfrac{2}{3}\overline{R}\right\}, \qquad 0 \leq t \leq \log 4\,, \tag{4.15}$$

$$\delta_2\left(\tfrac{1}{3}R\right) := \min\left\{\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle \,\middle|\, \mathrm{supp}(\boldsymbol{\theta}) \subseteq 4\mathbb{Z}_8\,,\ \mathrm{H}(\boldsymbol{\theta}) \geq \tfrac{1}{3}\overline{R}\right\}, \qquad 0 \leq t \leq \log 2\,, \tag{4.16}$$

are the GV distances associated to the subconstellations 4-PSK and 2-PSK respectively. Notice that a standard continuity argument may be used to show that (4.14) is actually an equality. Using a first order method based on (4.13) and inequality (4.14) it is possible to prove the following almost sure lower bound on the asymptotical normalized minimum distance of the GCE.

Theorem 34 *For every R in $(0, \log 8)$, with probability one*

$$\liminf_{N\in\mathbb{N}} \frac{1}{N} d_{\min}(\mathcal{T}_N^R) \geq \min\left\{\delta_8(R), \delta_4(\tfrac{2}{3}R), \delta_2(\tfrac{1}{3}R)\right\}. \tag{4.17}$$

Proof Let us fix an arbitrary $0 < \varepsilon < \overline{R}$ and define $B_\varepsilon := \mathcal{P}(\mathbb{Z}_8) \setminus A_{\overline{R}-\varepsilon}$. For every N, from Markov inequality and (4.13) we have

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\boldsymbol{\theta}\in B_\varepsilon}\{S_N(\boldsymbol{\theta}) \geq 1\}\right) &\leq \sum_{\boldsymbol{\theta}\in B_\varepsilon \cap \mathcal{P}_N(\mathbb{Z}_8)} \mathbb{E}\left[S_N(\boldsymbol{\theta})\right] \\
&\leq \sum_{\boldsymbol{\theta}\in B_\varepsilon \cap \mathcal{P}_N(\mathbb{Z}_8)} \exp\left(N(\mathrm{H}(\boldsymbol{\theta}) - \overline{R}\tfrac{\log l(\boldsymbol{\theta})}{\log 8})\right) \\
&\leq |\mathcal{P}_N(\mathbb{Z}_8)| \exp(-N\varepsilon)\,.
\end{aligned}$$

Define $f(t) := \min\left\{\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle \middle| \, \boldsymbol{\theta} \in A_t\right\}$. It follows that

$$\sum_N \mathbb{P}\Big(d_{\min}(\mathcal{T}_N^R) < N f(\overline{R} - \varepsilon)\Big) \leq \sum_N |\mathcal{P}_N(\mathbb{Z}_8)| \exp(-N\varepsilon) < +\infty\,,$$

and Borel Cantelli lemma implies that $\{\frac{1}{N} d_{\min}(\mathcal{T}_N^R) < f(\overline{R} - \varepsilon)\}$ occurs only for finitely many values of N. Then, for any $\varepsilon > 0$, $\liminf \frac{1}{N} \mathrm{d}_{\min}\left(\mathcal{T}_N^R\right) \geq f(\overline{R} - \varepsilon)$ almost surely. From the monotonicity and the semicontinuity of f it follows that

$$\begin{aligned}
\mathbb{P}\left(\liminf_N \tfrac{1}{N} \mathrm{d}_{\min}\left(\mathcal{T}_N^R\right) \geq f(\overline{R})\right) &= \mathbb{P}\left(\liminf_N \tfrac{1}{N} \mathrm{d}_{\min}\left(\mathcal{T}_N^R\right) \geq \lim_k f(\overline{R} - \tfrac{1}{k})\right) \\
&= \mathbb{P}\left(\cap_{k\in\mathbb{N}}\left\{\liminf_N \tfrac{1}{N} \mathrm{d}_{\min}\left(\mathcal{T}_N^R\right) \geq f(\overline{R} - \tfrac{1}{k})\right\}\right) \\
&= \lim_{k\in\mathbb{N}} \mathbb{P}\left(\liminf_N \tfrac{1}{N} \mathrm{d}_{\min}\left(\mathcal{T}_N^R\right) \geq f(\overline{R} - \tfrac{1}{k})\right) = 1\,.
\end{aligned}$$

Finally the claim follows from (4.14). ■

The final step of the present section consists in showing that the minimum on the righthand side of (4.17) is in fact given by the 8-PSK GV distance $\delta^{GV}(R)$. This is proved in the following.

Proposition 35 *For every design rate R in $(0, \log 8)$*

$$\delta_2(\tfrac{1}{3}R) = \delta_4(\tfrac{2}{3}R) > \delta^{GV}(R)\,, \tag{4.18}$$

Proof Thanks to the convexity property of the entropy function we can rewrite the Gilbert-Varshamov distance as

$$\delta_8(R) = \langle \boldsymbol{\theta}_\lambda, \boldsymbol{\delta}\rangle\,, \qquad \boldsymbol{\theta}_\lambda := \frac{1}{Z(\lambda)} e^{-\lambda\boldsymbol{\delta}}\,, \qquad Z(\lambda) := \sum_{x\in\mathbb{Z}_8} e^{-\lambda\boldsymbol{\delta}(x)}\,,$$

where the Lagrangian multiplier $\lambda > 0$ solves the equation $\mathrm{H}\left(\boldsymbol{\theta}_\lambda\right) = \overline{R}$. Similarly, it is possible to write

$$\delta_2\left(\tfrac{1}{3}R\right) = \frac{\boldsymbol{\delta}(4) e^{-\lambda_2 \boldsymbol{\delta}(4)}}{Z_2(\lambda_2)}\,, \qquad Z_2(\lambda_2) := 1 + e^{-\lambda_2 \boldsymbol{\delta}(4)}\,,$$

where $\lambda_2 > 0$ solves $\mathrm{H}\left(Z_2(\lambda_2)^{-1} e^{-\lambda_2 \boldsymbol{\delta}(4)}\right) = \frac{1}{3}\overline{R}$, while

$$\delta_4\left(\tfrac{2}{3}R\right) = \frac{\langle \exp(-\lambda_4 \boldsymbol{\delta}_4), \boldsymbol{\delta}_4 \rangle}{Z_4(\lambda_4)}\,, \qquad Z_4(\lambda_4) := \sum_{x \in 2\mathbb{Z}_8} e^{-\lambda \boldsymbol{\delta}(x)}\,.$$

where $\boldsymbol{\delta}_4$ denotes the restriction of the Bhattacharyya weight $\boldsymbol{\delta}$ to the subgroup $2\mathbb{Z}_8$, and $\lambda_4 > 0$ solves $\mathrm{H}\left(Z_4(\lambda_4)^{-1} e^{-\lambda_4 \boldsymbol{\delta}_4}\right) = \frac{2}{3}\overline{R}$.

Simple geometrical considerations based on Pythagoras theorems allow to show that

$$\boldsymbol{\delta}(4) = 2\boldsymbol{\delta}(2) = 2\boldsymbol{\delta}(6) \tag{4.19}$$

and

$$\boldsymbol{\delta}(1) = \boldsymbol{\delta}(7)\,, \qquad \boldsymbol{\delta}(3) = \boldsymbol{\delta}(5)\,, \qquad \boldsymbol{\delta}(1) = \boldsymbol{\delta}(4) - \boldsymbol{\delta}(3) < \frac{1}{4}\boldsymbol{\delta}(4)\,. \tag{4.20}$$

It follows from (4.19) that

$$Z_4(2s) = \left(1 + e^{-2s\boldsymbol{\delta}(2)}\right)^2 = \left(1 + e^{-s\boldsymbol{\delta}(4)}\right)^2 = \left(Z_2\left(s\right)\right)^2\,, \qquad \forall\, s \geq 0\,. \tag{4.21}$$

Define $\alpha := \dfrac{e^{-\lambda_2 \boldsymbol{\delta}(4)}}{Z_2(\lambda_2)}$. It follows from (4.19) and (4.21) that

$$\frac{e^{-2\lambda_2 \boldsymbol{\delta}(0)}}{Z_4(2\lambda_2)} = \frac{1}{\left(Z_2(\lambda_2)\right)^2} = (1-\alpha)^2\,, \qquad \frac{e^{-2\lambda_2 \boldsymbol{\delta}(2)}}{Z_4(2\lambda_2)} = \frac{e^{-2\lambda_2 \boldsymbol{\delta}(6)}}{Z_4(2\lambda_2)} = \alpha(1-\alpha)\,, \qquad \frac{e^{-2\lambda_2 \boldsymbol{\delta}(4)}}{Z_4(2\lambda_2)} = \alpha^2\,.$$

Then

$$\mathrm{H}\left(\left(Z_4(2\lambda_2)\right)^{-1} e^{-2\lambda_2 \boldsymbol{\delta}_4}\right) = 2\,\mathrm{H}\left(\alpha\right) = 2\,\mathrm{H}\left(\frac{e^{-\lambda_2 \boldsymbol{\delta}(4)}}{Z_2(\lambda_2)}\right)\,,$$

so that $2\lambda_2 = \lambda_4$. Therefore,

$$\delta_4\left(\tfrac{2}{3}R\right) = \frac{\langle \exp(-\lambda_4 \boldsymbol{\delta}_4), \boldsymbol{\delta}_4 \rangle}{Z_4(\lambda_4)} = \alpha^2 \boldsymbol{\delta}(4) + 2\alpha(1-\alpha)\boldsymbol{\delta}(2) = \alpha \boldsymbol{\delta}(4) = \frac{\boldsymbol{\delta}(4) e^{-\frac{\lambda_4}{2}\boldsymbol{\delta}(4)}}{Z_2(\lambda_4/2)} = \delta_1\left(\tfrac{1}{3}R\right)\,,$$

thus showing the equality in (4.18). It remains to show the inequality in (4.18). In order to do that we introduce the $\mathbb{Z}_8$-type $\hat{\boldsymbol{\theta}}$ defined by

$$\begin{gathered}\hat{\boldsymbol{\theta}}(0) := (1-\alpha)^3\,, \qquad \hat{\boldsymbol{\theta}}(1) := \hat{\boldsymbol{\theta}}(2) := \hat{\boldsymbol{\theta}}(7) := \alpha(1-\alpha)^2\,, \\ \hat{\boldsymbol{\theta}}(4) := \alpha^3\,, \qquad \hat{\boldsymbol{\theta}}(6) := \hat{\boldsymbol{\theta}}(5) := \hat{\boldsymbol{\theta}}(3) := \alpha^2(1-\alpha)\,.\end{gathered}$$

It is straightforward to verify that

$$\mathrm{H}(\hat{\boldsymbol{\theta}}) = 3\,\mathrm{H}\left(\alpha\right) = \overline{R}\,.$$

Moreover, it follows from (4.19) and (4.20) that

$$\begin{aligned}\langle\hat{\boldsymbol{\theta}},\boldsymbol{\delta}\rangle &= \sum_{x\in\mathbb{Z}_8}\boldsymbol{\delta}(x)\hat{\boldsymbol{\theta}}(x)\\ &= \alpha^3\boldsymbol{\delta}(4)+2\alpha^2(1-\alpha)\left(\boldsymbol{\delta}(4)-\boldsymbol{\delta}(1)\right)+\alpha(1-\alpha)\tfrac{1}{2}\boldsymbol{\delta}(4)+2\alpha(1-\alpha)^2\boldsymbol{\delta}(1)\\ &= \alpha\boldsymbol{\delta}(4)\tfrac{1}{2}\left(-2\alpha^2+3\alpha+1\right)+\alpha\boldsymbol{\delta}(1)2\left(1+2\alpha^2-3\alpha\right)\\ &= \alpha\boldsymbol{\delta}(4)+\alpha\boldsymbol{\delta}(4)\left(\left(2\boldsymbol{\delta}(1)-\tfrac{1}{2}\boldsymbol{\delta}(4)\right)\left(2\alpha^2-3\alpha+1\right)\right)\\ &< \alpha\,,\end{aligned}$$

last inequality following from (4.20) and the fact that $2\alpha^2-3\alpha+1>0$ for every $\alpha>0$. It follows that

$$\delta^{GV}(R)\le\langle\hat{\boldsymbol{\theta}},\boldsymbol{\delta}\rangle<\alpha\boldsymbol{\delta}(4)=\delta_2\left(\tfrac{1}{3}R\right)\,,$$

thus concluding the proof. ■

An immediate consequence of Theorem 34 and Proposition 35 is the following.

Corollary 36 *For every design rate R in $(0,\log 8)$ we have*

$$\mathbb{P}\left(\liminf_N\frac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{T}_N^R\right)\ge\delta^{GV}(R)\right)=1\,.$$

We observe that the geometry of the particular signal set 8-PSK plays a role only in the proof of Proposition 35. The rest of the derivation remains true for any $\mathbb{Z}_8$-symmetric channel and can in fact be generalized to Abelian group-code ensembles.

4.4.2 An upper bound on the typical asymptotic minimum distance

In this section we will prove the tightness the bound shown in Theorem 5. A second moment method will be used, and the key point consists in estimating the variance of the type spectra $\{T_N^R(\boldsymbol{\theta})\}$.

In order to do that, we need some preliminary considerations about the structure of the product set $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}\times(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$, $\boldsymbol{\theta}$ in $\mathcal{P}_N(\mathbb{Z}_8)$ being some $\mathbb{Z}_8$-type. Let $m=l(\boldsymbol{\theta})$ be the order of the smallest subgroup of $\mathbb{Z}_8$ supporting $\boldsymbol{\theta}$, and consider two non necessarily distinct N-tuples $\boldsymbol{x}$ and $\boldsymbol{z}$ both belonging to $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$, i.e. having type $\boldsymbol{\theta}$. Let $<\boldsymbol{x}>$, $<\boldsymbol{z}>$ and $<\boldsymbol{x},\boldsymbol{z}>$ the subgroups of $\mathbb{Z}_8^N$ respectively generated by $\boldsymbol{x}$, by $\boldsymbol{z}$, and by $\boldsymbol{x}$ and $\boldsymbol{z}$. It is easy to realize that both $<\boldsymbol{x}>$ and $<\boldsymbol{z}>$ are isomorphic to $\frac{8}{m}\mathbb{Z}_8$. Moreover the following diagram commutes

$$\begin{array}{ccc}<\boldsymbol{x}> & \overset{j_1}{\hookrightarrow} & <\boldsymbol{x},\boldsymbol{z}>\\ \downarrow i & & \uparrow f\\ \frac{8}{m}\mathbb{Z}_8 & \overset{j_2}{\hookrightarrow} & \frac{8}{m}\mathbb{Z}_8\oplus\frac{8}{m}\mathbb{Z}_8\end{array}$$

where:

- $i :< \boldsymbol{x} > \rightarrow \frac{8}{m}\mathbb{Z}_8$, $i(\alpha\boldsymbol{x}) = \alpha$ is an isomorphism;
- $j_1 :< \boldsymbol{x} > \rightarrow < \boldsymbol{x}, \boldsymbol{z} >$, $j_1(\alpha\boldsymbol{x}) = \alpha\boldsymbol{x}$, and $j_2 : \frac{8}{m}\mathbb{Z}_8 \rightarrow \frac{8}{m}\mathbb{Z}_8 \oplus \frac{8}{m}\mathbb{Z}_8$, $j_2(k) = (k, 0)$ are the standard injections;
- $f : \frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8 \oplus \frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8 \rightarrow < \boldsymbol{x}, \boldsymbol{z} >$, $f(a, b) = a\boldsymbol{x} + b\boldsymbol{z}$, is surjective.

It follows that $< \boldsymbol{x}, \boldsymbol{z} >$ contains a subgroup isomorphic to $\frac{8}{m}\mathbb{Z}_8$ and is itself isomorphic to a subgroup of $\frac{8}{m}\mathbb{Z}_8 \oplus \frac{8}{m}\mathbb{Z}_8$. An immediate consequence is that $< \boldsymbol{x}, \boldsymbol{z} >$ is isomorphic to a group of type $\frac{8}{m}\mathbb{Z}_8 \oplus \frac{8}{h}\mathbb{Z}_8$ for some h dividing m (possibly $h = 1$ when $\boldsymbol{x} = \boldsymbol{w}$). It is then possible to partition the set of ordered pairs of N-tuples of type $\boldsymbol{\theta}$ as follows:

$$(\mathbb{Z}_8)^N_{\boldsymbol{\theta}} \times (\mathbb{Z}_8)^N_{\boldsymbol{\theta}} = \bigcup_{h | l(\boldsymbol{\theta})} A_{N,\boldsymbol{\theta},h}\,, \tag{4.22}$$

with $A_{N,\boldsymbol{\theta},h}$ denoting the set of all pairs $(\boldsymbol{x}, \boldsymbol{z})$ in $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}} \times (\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$ such that the subgroup $< \boldsymbol{x}, \boldsymbol{z} >$ generated by $\boldsymbol{x}$ and $\boldsymbol{z}$ is isomorphic to $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8 \oplus \frac{8}{h}\mathbb{Z}_8$. The following lemma provides an estimation of the cardinality of $A_{N,\boldsymbol{\theta},h}$, with h ranging over the set of divisors of $l(\boldsymbol{\theta})$. For a subset A of a finite set B and a probability measure $\boldsymbol{\mu}$ in $\mathcal{P}(B)$ such that $\boldsymbol{\mu}(A) > 0$, we use the notation $\mu|_A$ for the conditional measure in $\mathcal{P}(A)$ defined by $\boldsymbol{\mu}_A(a) = \boldsymbol{\theta}(A)^{-1}\boldsymbol{\theta}(a)$.

Lemma 37 *For every N, $\boldsymbol{\theta}$ in $\mathcal{P}_N(\mathbb{Z}_8)$, and h dividing $l(\boldsymbol{\theta})$, we have*

$$|A_{N,\boldsymbol{\theta},h}| \leq 4\binom{N}{N\boldsymbol{\theta}} \prod_{\substack{1 \leq i \leq 8/h: \\ \boldsymbol{\theta}\left(i + \frac{8}{h}\mathbb{Z}_8\right) > 0}} \binom{N_i}{N_i\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}}\,, \tag{4.23}$$

where $N_i := N\boldsymbol{\theta}\left(i + \frac{8}{h}\mathbb{Z}_8\right)$ is the number of entries from the coset $i + \frac{8}{h}\mathbb{Z}_8$ in any N-tuple of type $\boldsymbol{\theta}$.

Proof Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be in $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$. A necessary condition for the subgroup $< \boldsymbol{x}, \boldsymbol{z} >$ to be isomorphic to $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8 \oplus \frac{8}{h}\mathbb{Z}_8$ is the existence of some α in $\mathbb{Z}_8^*$ such that

$$-h\alpha\boldsymbol{x} + h\boldsymbol{z} = \boldsymbol{0}\,. \tag{4.24}$$

For (4.24) to hold, necessarily $\boldsymbol{z}$ has to belong to $\alpha\boldsymbol{x} + \frac{8}{h}\mathbb{Z}_8^N$. Thus, whenever (4.24) holds, the set of positions of the entries of $\boldsymbol{x}$ belonging to any coset $i + \frac{8}{h}\mathbb{Z}_8$ and the set of positions of the entries of $\boldsymbol{z}$ belonging to the coset $\alpha i + \frac{8}{h}\mathbb{Z}_8$ need to coincide, i.e.

$$\boldsymbol{x}^{-1}\left(i + \tfrac{8}{h}\mathbb{Z}_8\right) = \boldsymbol{z}^{-1}\left(\alpha i + \tfrac{8}{h}\mathbb{Z}_8\right)\,, \qquad \forall\, i \in \mathbb{Z}_8\,. \tag{4.25}$$

Notice that since both $\boldsymbol{x}$ and $\boldsymbol{z}$ are assumed to be of type $\boldsymbol{\theta}$, (4.25) in particular implies

$$\boldsymbol{\theta}\left(i+\tfrac{8}{h}\mathbb{Z}_8\right)=\boldsymbol{\theta}\left(\alpha i+\tfrac{8}{h}\mathbb{Z}_8\right)\,,\qquad \forall\, i\in\mathbb{Z}_8\,. \tag{4.26}$$

For those α for which (4.26) is not satisfied there exists no pair $(\boldsymbol{x},\boldsymbol{z})$ satisfying (4.24). Thus, with no loss of generality we can restrict ourselves to considering values of α such that (4.26) is satisfied (as it is the case always for $\alpha=1$).

Notice that a necessary and sufficient condition for $\boldsymbol{x}$ and $\boldsymbol{z}$ both to belong to $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$ is the existence of an index permutation $\sigma:\{1,\dots,N\}\to\{1,\dots,N\}$ such that $\sigma\boldsymbol{x}:=\boldsymbol{x}\circ\sigma^{-1}=\boldsymbol{z}$. Equation (4.25) can be read as a constraint on the structure of σ, which has necessarily to be of the form

$$\sigma=\sigma^1\circ\sigma^2\circ\ldots\circ\sigma^{8/h}\circ\tilde{\sigma}_{\alpha,\boldsymbol{x}}\,. \tag{4.27}$$

In (4.27) $\tilde{\sigma}_{\alpha,\boldsymbol{x}}$ is the index permutation mapping, for every coset $i+\frac{8}{h}\mathbb{Z}_8$, the smallest element of $\boldsymbol{x}^{-1}\left(\alpha^{-1}i+\frac{8}{h}\mathbb{Z}_8\right)$ in the smallest element of $\boldsymbol{x}^{-1}\left(i+\frac{8}{h}\mathbb{Z}_8\right)$, the second smallest element $\boldsymbol{x}^{-1}\left(\alpha^{-1}i+\frac{8}{h}\mathbb{Z}_8\right)$ in the second smallest element of $\boldsymbol{z}^{-1}\left(i+\frac{8}{h}\mathbb{Z}_8\right)$, and so on. For every coset $i+\frac{8}{h}\mathbb{Z}_8$ instead, $\sigma^i:\{0,\dots,N\}\to\{0,\dots,N\}$ is any permutation such that

$$\sigma^i(j)=j\,,\qquad \forall\; j\in\{1,\dots,N\}\setminus\boldsymbol{x}^{-1}\left(i+\frac{8}{h}\mathbb{Z}_8\right)\,. \tag{4.28}$$

Thus, for a given $\boldsymbol{x}$ in $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$ and α in $\mathbb{Z}_8^*$ such that (4.26) is satisfied, we have that the number of $\boldsymbol{z}$ in $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$ satisfying (4.25) equals the cardinality of the orbit of $\tilde{\sigma}_{\alpha,\boldsymbol{x}}\boldsymbol{x}$ under the action of the group of index permutations

$$G^{(\boldsymbol{x})}:=\{\sigma=\sigma^1\circ\sigma^2\circ\ldots\circ\sigma^{8/h}\;:\;(4.28)\,\forall\; i=1,\dots,\tfrac{8}{h}\}\,.$$

Clearly the order of this group is $\left|G^{(\boldsymbol{x})}\right|=\prod_{i=1}^{8/h}N_i!$, while the cardinality of the stabilizer of $\tilde{\sigma}_{\alpha,\boldsymbol{x}}\boldsymbol{x}$ in $G^{(\boldsymbol{x})}$ is $\left|\mathrm{Stab}\left(\tilde{\sigma}_{\alpha,\boldsymbol{x}}\boldsymbol{x},G^{(\boldsymbol{x})}\right)\right|=\prod_{i=1}^{8}(N\boldsymbol{\theta}(i))!$, so that the orbit of $\tilde{\sigma}_{\alpha,\boldsymbol{x}}\boldsymbol{x}$ in $G^{(\boldsymbol{x})}$ has cardinality $\left|O\left(G^{(\boldsymbol{x})},\tilde{\sigma}_{\alpha,\boldsymbol{x}}\boldsymbol{x}\right)\right|=\prod_{i=1}^{8/h}N_i!/\prod_{i=1}^{8}(N\boldsymbol{\theta}(i))!=\prod_{i=1}^{8/h}\binom{N_i}{N_i\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}}$.

This allows us to estimate the cardinality of $A_{N,\boldsymbol{\theta},h}$ by

$$|A_{N,\boldsymbol{\theta},h}|\le|\mathbb{Z}_8^*|\sum_{\boldsymbol{x}\in(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}}\left|O\left(G^{(\boldsymbol{x})},\tilde{\sigma}_{\alpha,\boldsymbol{x}}\boldsymbol{x}\right)\right|=4\binom{N}{N\boldsymbol{\theta}}\prod_{i=1}^{8/h}\binom{N_i}{N_i\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}}\,.$$

■

Lemma 38 *For every N in $\mathbb{N}$ and $\boldsymbol{\theta}$ in $\mathcal{P}_N(\mathbb{Z}_8)$*

$$\mathrm{Var}\left[T_N^R(\boldsymbol{\theta})\right] \leq 4\binom{N}{N\boldsymbol{\theta}}\left(\frac{1}{l(\boldsymbol{\theta})}\right)^L \sum_{\substack{h|l(\boldsymbol{\theta}) \\ h<l(\boldsymbol{\theta})}} \left(\frac{1}{h}\right)^L \prod_{i=1}^{8/h} \binom{N_i}{N_i\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}}, \tag{4.29}$$

for N_i defined as in Lemma 37.

Proof Let $\boldsymbol{x}$ and $\boldsymbol{z}$ in $(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}$ be two N-tuples of type $\boldsymbol{\theta}$. We claim that if $<\boldsymbol{x},\boldsymbol{z}>$ is isomorphic to $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8 \oplus \frac{8}{h}\mathbb{Z}_8$, then the pair $\left(\Phi_N^R\boldsymbol{x}, \Phi_N^R\boldsymbol{z}\right)$ is uniformly distributed over a subgroup $(\mathbb{Z}_8^2)^L$ which is isomorphic to $\left(\mathbb{Z}_8 \oplus \frac{8}{h}\mathbb{Z}_8\right)^L$. To see this, notice that for every $1 \leq j \leq L$ the image of the evaluation homomorphism

$$\Psi_j : \hom(\mathbb{Z}_8^N, \mathbb{Z}_8) \to \mathbb{Z}_8^2, \qquad \Psi_j(\Phi) = (\Phi\boldsymbol{x}, \Phi\boldsymbol{z})$$

coincides with $<(x_i,z_i)>_{1\leq i\leq N}$, i.e. the subgroup of $\mathbb{Z}_8^2$ generated by $\{(x_i,z_i)\}_{1\leq i\leq N}$. It thus follows from Lemma 10 of Chapter 3 that each pair $\left((\Phi_N^R\boldsymbol{x})_j, (\Phi_N^R\boldsymbol{z})_j\right)$ is uniformly distributed over $<(x_i,z_i)>_{i=1,\ldots,N}$. Moreover the r.v.s $((\Phi_N\boldsymbol{x})_j, (\Phi_N\boldsymbol{z})_j)$ for $j=1,\ldots,L$ are mutually independent. Notice that $<(x_i,z_i)>_{i=1,\ldots,N}$ is isomorphic to the subgroup $<\boldsymbol{x},\boldsymbol{z}>$ of $\mathbb{Z}_8^L$ generated by $\boldsymbol{x}$ and $\boldsymbol{z}$, which, as previously observed, is itself isomorphic to a group of type $\frac{8}{l(\boldsymbol{\theta})}\mathbb{Z}_8 \oplus \frac{8}{h}\mathbb{Z}_8$. Recalling the partition (4.22), we have that, whenever the pair $(\boldsymbol{x},\boldsymbol{z})$ belongs to $A_{N,\boldsymbol{\theta},h}$, the joint probability that both $\boldsymbol{x}$ and $\boldsymbol{z}$ belong to $\ker\Phi_N^R$ is given by

$$\mathbb{P}\left(\Phi_N^R\boldsymbol{x} = \mathbf{0}, \Phi_N^R\boldsymbol{z} = \mathbf{0}\right) = \left(\frac{1}{hl(\boldsymbol{\theta})}\right)^L \tag{4.30}$$

It follows from (4.22), (4.23) and (4.30) that

$$\begin{aligned}
\mathrm{Var}\left[T_N^R(\boldsymbol{\theta})\right] &= \sum_{\boldsymbol{x},\boldsymbol{w}\in(\mathbb{Z}_8)^N_{\boldsymbol{\theta}}} \mathrm{Cov}\left[\mathbb{1}_{\left\{\Phi_N^R\boldsymbol{x}=\mathbf{0}\right\}}\mathbb{1}_{\left\{\Phi_N^R\boldsymbol{z}=\mathbf{0}\right\}}\right] \\
&= \sum_{h|l(\boldsymbol{\theta})} \sum_{(\boldsymbol{x},\boldsymbol{z})\in A_{N,\boldsymbol{\theta},h}} \mathbb{P}\left(\Phi_N^R\boldsymbol{x}=\mathbf{0}, \Phi_N^R\boldsymbol{z}=\mathbf{0}\right) - \mathbb{P}\left(\Phi_N^R\boldsymbol{x}=\mathbf{0}\right)\mathbb{P}\left(\Phi_N^R\boldsymbol{z}=\mathbf{0}\right) \\
&= \sum_{h|l(\boldsymbol{\theta})} |A_{N,\boldsymbol{\theta},h}| \left(\tfrac{1}{h^L l(\boldsymbol{\theta})^L} - \tfrac{1}{l(\boldsymbol{\theta})^{2L}}\right) \\
&\leq \sum_{\substack{h|l(\boldsymbol{\theta}) \\ h<l(\boldsymbol{\theta})}} 4\tbinom{N}{N\boldsymbol{\theta}} \tfrac{1}{(hl(\boldsymbol{\theta}))^L} \prod_{i=1}^{8/h} \tbinom{N_i}{N_i\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}}.
\end{aligned}$$

■

For every h dividing 8, consider the projection τ^h of $\mathbb{Z}_8$ onto the quotient group $\mathbb{Z}_8\Big/\frac{8}{h}\mathbb{Z}_8$, defined by

$$\tau^h(x) = y + \tfrac{8}{h}\mathbb{Z}_8 \qquad \Leftrightarrow \qquad x \in y + \tfrac{8}{h}\mathbb{Z}8\,,$$

and, for every $\boldsymbol{\theta}$ in $\mathcal{P}(\mathbb{Z}_8)$ denote by $\tau^h_{\#}\boldsymbol{\theta}$ in $\mathcal{P}\left(\mathbb{Z}_8\Big/\frac{8}{h}\mathbb{Z}_8\right)$ the image measure under τ^h. As an immediate consequence of Lemma 33 and Lemma 38 we have

$$\begin{aligned}
\limsup_{N\in\mathbb{N}} \frac{1}{N}\log\left(\frac{\mathrm{Var}\left[S^N_{\boldsymbol{\theta}}\right]}{\mathbb{E}\left[S^N_{\boldsymbol{\theta}}\right]^2}\right) &\leq \limsup_{N\in\mathbb{N}} \frac{1}{N}\log\binom{N}{N\boldsymbol{\theta}}^{-1}\sum_{\substack{h|l(\boldsymbol{\theta})\\ h<l(\boldsymbol{\theta})}}\left(\frac{l(\boldsymbol{\theta})}{h}\right)^L\prod_{i=1}^{8/h}\binom{N_i}{N_i\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}}\\
&= \max_{\substack{h|l(\boldsymbol{\theta})\\ h<l(\boldsymbol{\theta})}} \tfrac{\log l(\boldsymbol{\theta})/h}{\log 8}\overline{R} - \mathrm{H}(\boldsymbol{\theta}) + \sum_{i=1}^{8/h}\boldsymbol{\theta}(i+\tfrac{8}{h}\mathbb{Z}_8)\,\mathrm{H}\left(\boldsymbol{\theta}|_{i+\frac{8}{h}\mathbb{Z}_8}\right)\\
&= \max_{\substack{h|l(\boldsymbol{\theta})\\ h<l(\boldsymbol{\theta})}}\left\{\frac{\log l(\boldsymbol{\theta})/h}{\log 8}\overline{R} - \mathrm{H}\left(\boldsymbol{\tau}_h(\boldsymbol{\theta})\right)\right\}.
\end{aligned} \tag{4.31}$$

The following result shows the tightness of the almost sure lower bound of Theorem 34. Its proof relies on relies on an application of the second order method and the key point consists in showing that the $\mathbb{Z}_8$-type $\boldsymbol{\theta}$ giving the GV-distance can be approximated by $\mathbb{Z}_8$-types $\boldsymbol{\theta}_\varepsilon$ such that $\mathrm{H}\left(\boldsymbol{\tau}^h_{\#}\boldsymbol{\theta}_\varepsilon\right) > \frac{\log l(\boldsymbol{\theta}_\varepsilon)/h}{\log 8}\overline{R}$ for all $h = 1, 2, 4$.

Theorem 39 *For every R in $(0, \log 8)$,*

$$\mathbb{P}\left(\limsup_{N\in\mathbb{N}}\frac{1}{N}d_{\min}\left(\mathcal{T}^R_N\right) \leq \delta^{GV}\left(\overline{R}\right)\right) = 1\,.$$

Proof Let us fix an arbitrary ε in $(0, R)$ and denote by $\boldsymbol{\theta} = \boldsymbol{\theta}(R-\varepsilon)$ the $\mathbb{Z}_8$-type achieving the GV distance $\delta_8(R-\varepsilon)$, i.e. such that $\delta_8(R-\varepsilon) = \langle\boldsymbol{\theta}, \boldsymbol{\delta}\rangle$ and

$$\mathrm{H}(\boldsymbol{\theta}) \geq \overline{R} + \varepsilon\,. \tag{4.32}$$

We can express $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = \left(\sum_{x\in\mathbb{Z}_8} e^{-\lambda\boldsymbol{\delta}(x)}\right)^{-1} e^{-\lambda\boldsymbol{\delta}}$ where the Lagrangian multiplier λ is the unique positive solution of the equation $\mathrm{H}\left(\boldsymbol{\theta}\right) = \overline{R} + \varepsilon$. Notice that $\mathrm{supp}(\boldsymbol{\theta}) = \mathbb{Z}_8$, so that in particular $l(\boldsymbol{\theta}) = 8$. Also, observe that (4.19) and (4.20) imply that $\boldsymbol{\theta}$ has the following ordering

$$\boldsymbol{\theta}(0) > \boldsymbol{\theta}(1) = \boldsymbol{\theta}(7) > \boldsymbol{\theta}(2) = \boldsymbol{\theta}(6) > \boldsymbol{\theta}(3) = \boldsymbol{\theta}(5) > \boldsymbol{\theta}(4)\,. \tag{4.33}$$

Define

$$A_0 := \{0,1,7,2\}\,, \qquad B_0 := \{0,1,6,3\}\,, \qquad C_0 := \{0,5,6,7\}\,,$$

and let A_1, B_1 and C_1 be the complements in $\mathbb{Z}_8$ respectively of A_0, B_0 and C_0. It follows from (4.33) that

$$\begin{aligned} &\boldsymbol{\theta}(A_0) \geq \boldsymbol{\theta}(2\mathbb{Z}_8)\,, && \boldsymbol{\theta}(A_0) \geq \boldsymbol{\theta}(2\mathbb{Z}_8+1)\,, \\ &\boldsymbol{\theta}(B_0) \geq \boldsymbol{\theta}(2\mathbb{Z}_8)\,, && \boldsymbol{\theta}(B_0) \geq \boldsymbol{\theta}(2\mathbb{Z}_8+1)\,, \\ &\boldsymbol{\theta}(C_0) \geq \boldsymbol{\theta}(2\mathbb{Z}_8)\,, && \boldsymbol{\theta}(C_0) \geq \boldsymbol{\theta}(2\mathbb{Z}_8+1)\,. \end{aligned} \tag{4.34}$$

Moreover, it is easy to check that $|A_a \cap B_b \cap C_c| = 1$, for every choice of (a,b,c) in $\{0,1\}^3$. Thus, $f : \mathbb{Z}_8 \to \{0,1\}^3$, where $f(x) = (a,b,c)$ if and only if x is in $A_a \cap B_b \cap C_c$, is a bijection. Then, from (7.3), (7.5) and (4.34), it thus follows that

$$\mathrm{H}(\boldsymbol{\theta}) = \mathrm{H}\left(f_{\#}\boldsymbol{\theta}\right) \geq \mathrm{H}\left(\boldsymbol{\theta}(A_0)\right) + \mathrm{H}\left(\boldsymbol{\theta}(B_0)\right) + \mathrm{H}\left(\boldsymbol{\theta}(C_0)\right) \geq 3\,\mathrm{H}\left(\boldsymbol{\theta}(2\mathbb{Z}_8)\right) = 3\,\mathrm{H}\left(\tau_{\#}^4\boldsymbol{\theta}\right)\,. \tag{4.35}$$

Let us now introduce the sets $D := \{0,2\}$ and $E := \{1,7\}$. We have from (4.33) that

$$\begin{aligned} &\boldsymbol{\theta}(D) \geq \boldsymbol{\theta}(4\mathbb{Z}_8)\,, && \boldsymbol{\theta}(D) \geq \boldsymbol{\theta}(4\mathbb{Z}_8+2)\,, \\ &\boldsymbol{\theta}(E) \geq \boldsymbol{\theta}(4\mathbb{Z}_8+1)\,, && \boldsymbol{\theta}(E) \geq \boldsymbol{\theta}(4\mathbb{Z}_8+3)\,. \end{aligned}$$

It thus follows that

$$\begin{aligned} \mathrm{H}\left(\tau_{\#}^2\boldsymbol{\theta}\right) &= \mathrm{H}\left(\tau_{\#}^4\boldsymbol{\theta}\right) + \boldsymbol{\theta}(2\mathbb{Z}_8)\,\mathrm{H}\left(\tau_{\#}^4\boldsymbol{\theta}_4\right) + \boldsymbol{\theta}(2\mathbb{Z}_8+1)\,\mathrm{H}\left(\tau_{\#}^4\boldsymbol{\theta}|_{2\mathbb{Z}_8+1}\right) \\ &\geq \mathrm{H}(\boldsymbol{\theta}(2\mathbb{Z}_8)) + \boldsymbol{\theta}(2\mathbb{Z}_8)\,\mathrm{H}\left(\boldsymbol{\theta}_4(D)\right) + \boldsymbol{\theta}(2\mathbb{Z}_8+1)\,\mathrm{H}\left(\boldsymbol{\theta}|_{2\mathbb{Z}_8+1}(E)\right)\,. \end{aligned} \tag{4.36}$$

Observe that

$$\boldsymbol{\theta}_4(D) = \frac{\boldsymbol{\theta}(0)}{\boldsymbol{\theta}(2\mathbb{Z}_8)} + \frac{\boldsymbol{\theta}(2)}{\boldsymbol{\theta}(2\mathbb{Z}_8)} = \boldsymbol{\theta}_4(4\mathbb{Z}_8)\boldsymbol{\theta}|_{4\mathbb{Z}_8}(0) + \boldsymbol{\theta}_4(4\mathbb{Z}_8+2)\boldsymbol{\theta}|_{4\mathbb{Z}_8+2}(2)$$

so that, by the concavity of the entropy function, we get

$$\mathrm{H}\left(\boldsymbol{\theta}_4(D)\right) \geq \boldsymbol{\theta}_4(4\mathbb{Z}_8)\,\mathrm{H}\left(\boldsymbol{\theta}|_{4\mathbb{Z}_8}(0)\right) + \boldsymbol{\theta}_4(4\mathbb{Z}_8+2)\,\mathrm{H}\left(\boldsymbol{\theta}|_{4\mathbb{Z}_8+2}(2)\right)\,.$$

An analogous reasoning leads to

$$\mathrm{H}\left(\boldsymbol{\theta}|_{2\mathbb{Z}_8+1}(E)\right) \geq \boldsymbol{\theta}|_{2\mathbb{Z}_8+1}(4\mathbb{Z}_8+1)\,\mathrm{H}\left(\boldsymbol{\theta}|_{4\mathbb{Z}_8+1}(1)\right) + \boldsymbol{\theta}|_{2\mathbb{Z}_8+1}(4\mathbb{Z}_8+3)\,\mathrm{H}\left(\boldsymbol{\theta}|_{4\mathbb{Z}_8+3}(3)\right)\,.$$

Upon substituting the two inequalities above in (4.36), we get

$$\begin{aligned} \mathrm{H}\left(\tau_{\#}^2\boldsymbol{\theta}\right) &\geq \mathrm{H}\left(\tau_{\#}^4\boldsymbol{\theta}\right) + \textstyle\sum_{i=0}^3 \boldsymbol{\theta}(4\mathbb{Z}_8+i)\,\mathrm{H}\left(\boldsymbol{\theta}|_{4\mathbb{Z}_8+i}(i)\right) \\ &= \mathrm{H}\left(\tau_{\#}^4\boldsymbol{\theta}\right) + \mathrm{H}(\boldsymbol{\theta}) - \mathrm{H}\left(\tau_{\#}^2\boldsymbol{\theta}\right) \\ &\geq \tfrac{4}{3}\,\mathrm{H}(\boldsymbol{\theta}) - \mathrm{H}\left(\tau_{\#}^2\boldsymbol{\theta}\right)\,, \end{aligned}$$

last inequality following from (4.35). Then

$$\mathrm{H}\left(\tau_{\#}^{2}\boldsymbol{\theta}\right) \geq \frac{2}{3}\,\mathrm{H}(\boldsymbol{\theta})\,. \tag{4.37}$$

Now let $(\boldsymbol{\theta}_N)$ ba a sequence of $\mathbb{Z}_8$-types converging to $\boldsymbol{\theta}$, with $\boldsymbol{\theta}_N$ belonging to $\mathcal{P}_N(\mathbb{Z}_8)$ for every N. By successively applying Chabyshev inequality, (4.31), (4.35) and (4.37), we get

$$\begin{aligned}
\limsup_N \tfrac{1}{N}\log \mathbb{P}\left(T_N^R(\boldsymbol{\theta}_N)=0\right) &\leq \limsup_N \tfrac{1}{N}\log \frac{\mathrm{Var}\left[T_N^R(\boldsymbol{\theta}_N)\right]}{\mathbb{E}[S_N(\boldsymbol{\theta}_N)]^2} \\
&\leq \max\left\{\tfrac{1}{3}\overline{R}-\mathrm{H}\left(\boldsymbol{\tau}_{\#}^{4}\boldsymbol{\theta}\right), \tfrac{2}{3}\overline{R}-\mathrm{H}\left(\boldsymbol{\tau}_{\#}^{2}\boldsymbol{\theta}\right), \overline{R}-\mathrm{H}\left(\boldsymbol{\theta}\right)\right\} \\
&\leq \max\left\{\tfrac{1}{3}\left(\overline{R}-\mathrm{H}\left(\boldsymbol{\theta}\right)\right), \tfrac{2}{3}\left(\overline{R}-\mathrm{H}\left(\boldsymbol{\theta}\right)\right), \overline{R}-\mathrm{H}\left(\boldsymbol{\theta}\right)\right\} \\
&\leq -\tfrac{1}{3}\varepsilon\,,
\end{aligned}$$

last inequality following from (4.32). Thus $\sum_N \mathbb{P}(S_N(\boldsymbol{\theta}_N)=0)<\infty$ and an application of Borel-Cantelli lemma implies

$$\mathbb{P}\left(\limsup_{N\in\mathbb{N}} \frac{1}{N} d_{\min}\left(\mathcal{T}_N^R\right) > \delta^{GV}(R-\varepsilon)\right) \leq \mathbb{P}\left(\left\{T_N^R(\boldsymbol{\theta}_N)\geq 1\right\} \text{ i. o. } \ N\in\mathbb{N}\right) = 0\,.$$

Finally, the claim follows from the arbitrariness of ε in $(0,R)$ in the previous arguments and the continuity of the Gilbert-Varshamov distance $\delta^{GV}(R)$ as a function of the rate R. ■

4.5 Minimum distance of the typical binary affine code

In this section we analyze the asymptotics of the normalized minimum distance of the binary affine code ensemble $\{\mathcal{U}_N^R\}$ under an arbitrary labeling

$$\eta : \mathbb{Z}_2^3 \to \mathbb{Z}_8\,,$$

which has to be consider fixed throughout the section.

A first observation is that, since $\mathbb{Z}_2$-affine codes are not geometrically uniform, their minimum distance does not coincide with their minimum weight, as we have seen in the previous section it is the case for $\mathbb{Z}_8$-codes. Rather, similarly to what we did in Section 4.3 for the random coding ensemble, it is necessary to look at all pairs of codewords evaluate the minimum distance. We introduce the distance function

$$\boldsymbol{\Delta}_\eta : \mathbb{Z}_2^3 \times \mathbb{Z}_2^3 \to \mathbb{R}^+\,, \qquad \boldsymbol{\Delta}_\eta(x,y) = \boldsymbol{\Delta}(\eta(y),\eta(y+x))\,.$$

It is easy to check that each column $\boldsymbol{\Delta}_\eta(\cdot, z)$ of $\boldsymbol{\Delta}_\eta$ is just a permutation of $\boldsymbol{\delta}$, i.e. there exists a bijection $\sigma_z : \mathbb{Z}_2^3 \to \mathbb{Z}_8$ such that

$$\boldsymbol{\Delta}_\eta(x, z) = \boldsymbol{\delta}(\sigma_z(x)) \,, \qquad x, y \in \mathbb{Z}_2^3 \,. \tag{4.38}$$

Moreover, we recall that the isometry group of the 8-PSK is isometric to D_8. As a consequence there exist at least two non-identical columns of $\boldsymbol{\Delta}_\eta$, i.e. $\boldsymbol{\Delta}_\eta(x, z_1) \neq \boldsymbol{\Delta}_\eta(x, z_2)$ for some x, z_1, z_2 in $\mathbb{Z}_2^3$.

We introduce the following enumerating functions for the $\mathbb{Z}_2$-affine ensemble: for a joint type $\boldsymbol{\vartheta}$ in $\mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ and a blocklength N in $\mathbb{N}$

$$U_N^R(\boldsymbol{\vartheta}) := \left|\left\{(\boldsymbol{x}, \boldsymbol{y}) \in (\mathbb{Z}_2^3)_{\boldsymbol{\vartheta}}^N \,:\, \boldsymbol{x} \in \ker \Psi_N^R, \Psi_N^R \boldsymbol{y} = \boldsymbol{Z}_N\right\}\right| \,,$$

denotes the number of pairs of $\mathbb{Z}_2^3$ N-tuples $(\boldsymbol{x}, \boldsymbol{y})$ of joint type $\boldsymbol{\vartheta}$ such that both $\boldsymbol{y}$ and $\boldsymbol{x} + \boldsymbol{y}$ belong to $\mathcal{U}_N^R$. For a type $\boldsymbol{\theta}$ in $\mathcal{P}(\mathbb{Z}_2^3)$ let

$$V_N^R(\boldsymbol{\theta}) := \left|\left\{\boldsymbol{x} \in (\mathbb{Z}_2^3)_{\boldsymbol{\theta}}^N \,:\, \boldsymbol{x} \in \ker \Psi_N^R\right\}\right|$$

denote the number of type-$\boldsymbol{\theta}$ N-tuples in the kernel of Ψ_N^R. It is straightforward to check that the minimum Bhattacharyya-distance of the binary affine ensemble can be rewritten as

$$\mathrm{d}_{\min}\left(\mathcal{U}_N^R\right) = N \inf\left\{\langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle \mid \boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) : \pi_\#^1 \boldsymbol{\vartheta} \neq \delta_0,\, U_N^R(\boldsymbol{\vartheta}) \geq 1\right\} \,,$$

where $\pi_\#^1 : \mathbb{Z}_2^3 \times \mathbb{Z}_2^3 \to \mathbb{Z}_2^3$ denotes the marginal projection operator on the first component. Notice that for every $\boldsymbol{\theta}$ in $\mathcal{P}(\mathbb{Z}_2^3)$ we have

$$\begin{aligned}
|\mathcal{U}_N^R| V_N^R(\boldsymbol{\theta}) &= \Big(\textstyle\sum_{\boldsymbol{y} \in \mathbb{Z}_2^{3N}} \mathbb{1}_{\left\{\Psi_N^R \boldsymbol{y} = \boldsymbol{Z}_N\right\}}\Big)\Big(\textstyle\sum_{\boldsymbol{x} \in \left(\mathbb{Z}_2^3\right)_{\boldsymbol{\theta}}^N} \mathbb{1}_{\left\{\Psi_N^R \boldsymbol{x} = \boldsymbol{0}\right\}}\Big) \\
&= \textstyle\sum_{\boldsymbol{\vartheta} \in \mathcal{P}_N(\mathbb{Z}_2^3) : \pi_\#^1 \boldsymbol{\vartheta} = \boldsymbol{\theta}} \quad \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in \left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)_{\boldsymbol{\vartheta}}^N} \mathbb{1}_{\{\Phi_N \boldsymbol{x} = \boldsymbol{0}\}} \mathbb{1}_{\{\Phi_N \boldsymbol{y} = \boldsymbol{Z}_N\}} \\
&= \textstyle\sum_{\boldsymbol{\vartheta} \in \mathcal{P}_N(\mathbb{Z}_2^3) : \pi_\#^1 \boldsymbol{\vartheta} = \boldsymbol{\theta}} U_N^R(\boldsymbol{\vartheta}) \,,
\end{aligned}$$

so that in particular the following holds:

$$V_N^R(\boldsymbol{\theta}) = 0 \qquad \Longleftrightarrow \qquad U_N^R(\boldsymbol{\vartheta}) = 0 \,, \quad \forall\, \boldsymbol{\vartheta} \in \left(\pi_\#^1\right)^{-1}(\boldsymbol{\theta}) \,. \tag{4.39}$$

4.5.1 A lower bound on the typical asymptotic minimum distance of the binary affine code ensemble

We want to show that almost surely the sequence of the normalized minimum distance of the binary affine ensemble $\left(\frac{1}{N} \mathrm{d}_{\min}(\mathcal{U}_N^R)\right)$ has lim inf not smaller than

$$\underline{\delta}_\eta(R) := \min\left\{\langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle \,, | \,, \boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) : \mathrm{H}(\boldsymbol{\vartheta}) \geq 2\overline{R},\, \mathrm{H}\left(\pi_\#^1 \boldsymbol{\vartheta}\right) \geq \overline{R}\right\} \,. \tag{4.40}$$

We start by evaluating the expected value of the enumerating functions $U_N^R(\boldsymbol{\vartheta})$ and $V_N^R(\boldsymbol{\vartheta})$.

Lemma 40 *For every $\boldsymbol{\vartheta}$ in $\mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ and $\boldsymbol{\theta}$ in $\mathcal{P}_N\left(\mathbb{Z}_2^3\right)$ such that*

$$\pi^1_{\#}\boldsymbol{\vartheta} \neq \delta_0\,, \qquad \boldsymbol{\theta} \neq \delta_0$$

we have

$$\mathbb{E}\left[U_N^R(\boldsymbol{\vartheta})\right] = \binom{N}{N\boldsymbol{\vartheta}}\frac{1}{8^{2L}}\,, \qquad \mathbb{E}\left[V_N^R(\boldsymbol{\theta})\right] = \binom{N}{N\boldsymbol{\theta}}\frac{1}{8^{L}}\,.$$

Proof For every $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{Z}_2^{3N}$ such that $\boldsymbol{x} \neq \mathbf{0}$ we have that $\Psi_N^R\boldsymbol{x}$ and $\Psi_N^R\boldsymbol{y} - \boldsymbol{Z}_N$ are independent and both uniformly distributed over $\mathbb{Z}_2^{3L}$. It follows that

$$\begin{aligned}\mathbb{E}\left[U_N^R(\boldsymbol{\vartheta})\right] &= \mathbb{E}\Big[\textstyle\sum_{(\boldsymbol{x},\boldsymbol{y})\in\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)^N_{\boldsymbol{\vartheta}}} \mathbb{1}_{\left\{\Psi_N^R\boldsymbol{x}=\mathbf{0}\right\}}\mathbb{1}_{\left\{\Psi_N^R\boldsymbol{y}=\boldsymbol{Z}_N\right\}}\Big] \\ &= \textstyle\sum_{(\boldsymbol{x},\boldsymbol{y})\in\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)^N_{\boldsymbol{\vartheta}}} \mathbb{P}\left(\Psi_N^R\boldsymbol{x} = \mathbf{0}\,,\ \Psi_N^R\boldsymbol{y} - \boldsymbol{Z}_N = \mathbf{0}\right) = \binom{N}{N\boldsymbol{\vartheta}}\frac{1}{8^{2L}}\,.\end{aligned}$$

Analogously $\mathbb{E}\left[V_N^R(\boldsymbol{\theta})\right] = \sum_{\boldsymbol{x}\in\left(\mathbb{Z}_2^3\right)^N_{\boldsymbol{\theta}}} \mathbb{P}\left(\Psi_N^R\boldsymbol{x} = \mathbf{0}\right) = \binom{N}{N\boldsymbol{\vartheta}}\frac{1}{8^L}$. ■

For every t in $[0, \log 8]$, we define the set A_t

$$A_t := \left\{\boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) \,:\, \mathrm{H}(\boldsymbol{\vartheta}) \geq 2t\,,\ \mathrm{H}\left(\pi^1_{\#}\boldsymbol{\vartheta}\right) \geq t\right\}\,, \tag{4.41}$$

Then the function $\underline{\delta}_\eta(R)$ defined in (4.40) can be rewritten as

$$\underline{\delta}_\eta(R) = \min\left\{\langle\boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta\rangle \,\middle|\, \boldsymbol{\vartheta} \in A_{\overline{R}}\right\}\,. \tag{4.42}$$

A first-order method based on Lemma 40 allows to state that, for every joint type $\boldsymbol{\vartheta}$ not belonging to $A_{\overline{R}}$, $U_N^R(\boldsymbol{\vartheta}) = 0$ definitively in N with probability one. More precisely we have the following result.

Lemma 41 *For every ε in $(0, \overline{R})$, with probability one there exists N_0 in $\mathbb{N}$ such that*

$$U_N^R(\boldsymbol{\vartheta}) = 0\,, \qquad \forall\,\boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) \setminus A_{\overline{R}-\varepsilon}\,, \qquad \forall\, N \geq N_0\,. \tag{4.43}$$

Proof From Lemma 40, using the standard exponential bounds on the multinomials, we have that for every type $\boldsymbol{\vartheta}$ in $F_N := \mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) \setminus A_{\overline{R}-\varepsilon}$ at least one between

$$\mathbb{E}\left[U_N^R(\boldsymbol{\vartheta})\right] = \binom{N}{N\boldsymbol{\vartheta}}\frac{1}{8^{2L}} \leq \exp\left(N\left(\mathrm{H}(\boldsymbol{\vartheta}) - 2\overline{R}\right)\right) \leq \exp(-2N\varepsilon)\,, \tag{4.44}$$

and

$$\mathbb{E}\left[V_N^R\left(\pi^1_{\#}\boldsymbol{\vartheta}\right)\right] = \binom{N}{N\pi^1_{\#}\boldsymbol{\vartheta}}\frac{1}{8^L} \leq \exp\left(N\left(\mathrm{H}\left(\pi^1_{\#}\boldsymbol{\vartheta}\right) - \overline{R}\right)\right) \leq \exp(-N\varepsilon) \tag{4.45}$$

holds true. Then, successively using a union bound estimation, recalling (4.39), and applying Markov inequality, we have

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{\boldsymbol{\vartheta}\in F_N}\left\{U_N^R(\boldsymbol{\vartheta}) \geq 1\right\}\Big) &\leq \mathbb{P}\Big(\bigcup_{\boldsymbol{\vartheta}\in F_N}\left\{U_N^R(\boldsymbol{\vartheta}) \geq 1\right\} \cap \left\{V_N^R\left(\pi^1_{\#}\boldsymbol{\vartheta}\right) \geq 1\right\}\Big) \\
&\leq \sum_{\boldsymbol{\vartheta}\in F_N}\mathbb{P}\left(\left\{U_N^R(\boldsymbol{\vartheta}) \geq 1\right\} \cap \left\{V_N^R(\pi^1_{\#}\boldsymbol{\vartheta}) \geq 1\right\}\right) \\
&\leq \sum_{\boldsymbol{\vartheta}\in F_N}\min\left\{\mathbb{P}\left(S_N(\boldsymbol{\vartheta}) \geq 1\right), \mathbb{P}\left(V_N^R(\pi^1_{\#}\boldsymbol{\vartheta}) \geq 1\right)\right\} \\
&\leq \sum_{\boldsymbol{\vartheta}\in F_N}\min\left\{\mathbb{E}\left[S_N(\boldsymbol{\vartheta})\right], \mathbb{E}\left[V_N^R(\boldsymbol{\vartheta})\right]\right\} \\
&\leq \left|\mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)\right|\exp(-N\varepsilon)\,.
\end{aligned}$$

Thus $\sum_N \mathbb{P}\Big(\bigcup_{\boldsymbol{\vartheta}\in F_N}\left\{U_N^R(\boldsymbol{\vartheta}) \geq 1\right\}\Big) \leq \sum_N \left|\mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)\right|\exp(-N\varepsilon) < \infty$, and by Borel-Cantelli lemma we obtain $\mathbb{P}\left(\bigcup_{\boldsymbol{\vartheta}\in F_N}\left\{U_N^R(\boldsymbol{\vartheta}) \geq 1\right\} \text{ i.o. } N \in \mathbb{N}\right) = 0$, which is equivalent to the claim. ■

The following lower bound on the typical asymptotic minimum distance of the $\mathbb{Z}_2$-affine ensemble follows from Lemma 41 and the continuity of $\underline{\delta}_\eta(R)$ as a function of the design rate R.

Theorem 42 *For every design rate R in $(0, \log 8)$, with probability one*

$$\liminf_{N\in\mathbb{N}} \frac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) \geq \underline{\delta}_\eta(R)\,.$$

Proof Let us fix an arbitrary $\varepsilon > 0$. From the definition of $\underline{\delta}_\eta$, it follows that a sufficient condition for $\frac{1}{N}\,\mathrm{d}_{\min}\left(\mathcal{U}_N^R\right)$ to be not below $\underline{\delta}_\eta(R+\varepsilon)$ is that $U_N^R(\boldsymbol{\vartheta}) = 0$ for every joint type $\boldsymbol{\vartheta}$ in $\mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ not belonging to $A_{\overline{R}-\varepsilon}$. Then, from Lemma 41 we have

$$\mathbb{P}\left(\liminf_{N\in\mathbb{N}} \frac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) \geq \underline{\delta}_\eta(R+\varepsilon)\right) \geq \mathbb{P}\big(\exists N_0 \in \mathbb{N} : \ (4.43)\big) = 1\,.$$

Therefore, by the continuity of the function $\underline{\delta}_\eta$ we get

$$\begin{aligned}
\mathbb{P}\left(\liminf_{N\in\mathbb{N}} \tfrac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) \geq \underline{\delta}_\eta(R)\right) &= \mathbb{P}\left(\liminf_{N\in\mathbb{N}} \tfrac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) \geq \lim_{k\in\mathbb{N}} \underline{\delta}_\eta\left(R+\tfrac{1}{k}\right)\right) \\
&= \mathbb{P}\left(\bigcap_{k\in\mathbb{N}} \left\{\liminf_{N\in\mathbb{N}} \tfrac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) \geq \underline{\delta}_\eta\left(R+\tfrac{1}{k}\right)\right\}\right) \\
&= \lim_{k\in\mathbb{N}} \mathbb{P}\left(\liminf_{N\in\mathbb{N}} \tfrac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) \geq \underline{\delta}_\eta\left(R+\tfrac{1}{k}\right)\right) \\
&= 1\,,
\end{aligned}$$

showing the claim. ■

4.5.2 An upper bound on the typical asymptotic minimum distance of the binary affine code ensemble

We now want to show that the typical asymptotic normalized minimum distance of the binary affine code ensemble is upper-bounded by

$$\overline{\delta}_\eta(R) := \min\left\{\langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta\rangle\,, |\,, \boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) :\, \mathrm{H}(\boldsymbol{\vartheta}) - \mathrm{H}\left(\pi^1_{\#}\boldsymbol{\vartheta}\right) \geq \overline{R},\, \mathrm{H}(\pi^1_{\#}\boldsymbol{\vartheta}) \geq \overline{R}\right\}\,. \tag{4.46}$$

In order to do that we shall use a second moment method. As a first step we need to estimate the variance of the type spectrum $\{U_N^R(\boldsymbol{\vartheta})\}$.

Lemma 43 *Given N in $\mathbb{N}$, and a joint type $\boldsymbol{\vartheta}$ in $\mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ such that $\pi^1_{\#}\boldsymbol{\vartheta} \neq \delta_0$,*

$$\mathrm{Var}\left[U_N^R(\boldsymbol{\vartheta})\right] \leq \binom{N}{N\boldsymbol{\vartheta}}\binom{N}{N\pi^1_{\#}\boldsymbol{\vartheta}}\frac{16}{8^{3L}} + \binom{N}{N\boldsymbol{\vartheta}}^2\binom{N}{N\pi^1_{\#}\boldsymbol{\vartheta}}^{-1}\frac{1}{8^{3L}} + \binom{N}{N\boldsymbol{\vartheta}}\frac{8}{8^{2L}}\,, \tag{4.47}$$

Proof We have

$$\begin{aligned}
\mathrm{Var}\left[U_N^R(\boldsymbol{\vartheta})\right] &= \mathrm{Var}\left[\sum\nolimits_{(\boldsymbol{x},\boldsymbol{y})\in\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)^N_{\boldsymbol{\vartheta}}} \mathbb{1}_{\left\{\Psi_N^R\boldsymbol{x}=\boldsymbol{0}\right\}}\mathbb{1}_{\left\{\Psi_N^R\boldsymbol{y}=\boldsymbol{Z}_N\right\}}\right] \\
&= \sum_{(\boldsymbol{x}_1,\boldsymbol{y}_1),(\boldsymbol{x}_2,\boldsymbol{y}_2)\in\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)^N_{\boldsymbol{\vartheta}}} c\big(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}_1,\boldsymbol{y}_2\big)\,,
\end{aligned}$$

where

$$c\big(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}_1,\boldsymbol{y}_2\big) := \mathrm{Cov}\left[\mathbb{1}_{\left\{\Psi_N^R\boldsymbol{x}_1=\boldsymbol{0}\right\}}\mathbb{1}_{\left\{\Psi_N^R\boldsymbol{y}_1=\boldsymbol{Z}_N\right\}},\ \mathbb{1}_{\left\{\Psi_N^R\boldsymbol{x}_2=\boldsymbol{0}\right\}}\mathbb{1}_{\left\{\Psi_N^R\boldsymbol{y}_2=\boldsymbol{Z}_N\right\}}\right].$$

We are now going to estimate the terms $c\big(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}_1,\boldsymbol{y}_2\big)$, separately considering four possible different linear dependency structures among $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{y}_1$, and $\boldsymbol{y}_2$. Observe that,

since $\pi^1_{\#}\boldsymbol{\vartheta} \neq \delta_0$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ need to be nonzero in order for the pairs $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ and $(\boldsymbol{x}_2, \boldsymbol{y}_2)$ to have type $\boldsymbol{\vartheta}$.

Suppose first $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2)$ in $\big(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\big)^N_{\boldsymbol{\vartheta}}$ are such that $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are linear independent. Then the random variables $\Psi^R_N \boldsymbol{x}_1, \Psi^R_N \boldsymbol{x}_2, \Psi^R_N \boldsymbol{y}_1$ and $\Psi^R_N \boldsymbol{y}_2$ are independent so that

$$c\big(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2\big) = 0\,.$$

Second, consider the case when $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are linear independent but $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are not linear independent. In this case we have that the random variables $\Psi^R_N \boldsymbol{x}_1, \Psi^R_N \boldsymbol{x}_2$ and $\Psi^R_N \boldsymbol{y}_1 - \boldsymbol{Z}_N$ are independent, so that

$$c\big(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2\big) \leq \mathbb{P}\left(\Psi^R_N \boldsymbol{x}_1 = \boldsymbol{0}, \Psi^R_N \boldsymbol{x}_2 = \boldsymbol{0}, \Psi^R_N \boldsymbol{y}_2 = \boldsymbol{Z}_N\right) = \frac{1}{8^{3L}}\,.$$

Since there are at most $16\binom{N}{N\boldsymbol{\vartheta}}\binom{N}{N\pi^1_{\#}\boldsymbol{\vartheta}}$ possible choices of such pairs $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2)$ in $\big(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\big)^N_{\boldsymbol{\vartheta}}$, they contribute to the first addend in the righthand side of (4.47).

As a third case we consider pairs $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2)$, such that $\boldsymbol{x}_1 = \boldsymbol{x}_2$, and $\boldsymbol{x}_1$, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are linear independent. In this situation the random variables $\Psi^R_N \boldsymbol{x}_1$, $\Psi^R_N \boldsymbol{y}_1$ and $\Psi^R_N \boldsymbol{y}_2$ are independent so that

$$c\big(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2\big) \leq \mathbb{P}\left(\Psi^R_N \boldsymbol{x}_1 = \boldsymbol{0}, \Psi^R_N \boldsymbol{y}_1 = \boldsymbol{Z}_N, \Psi^R_N \boldsymbol{y}_2 = \boldsymbol{Z}_N\right) = \frac{1}{8^{3L}}\,.$$

Since there are at most $\binom{N}{N\boldsymbol{\vartheta}}^2\binom{N}{N\pi^1_{\#}\boldsymbol{\vartheta}}^{-1}$ possible choices of such pairs $(\boldsymbol{x}_1, \boldsymbol{y}_1), (\boldsymbol{x}_2, \boldsymbol{y}_2)$ in $\big(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\big)^N_{\boldsymbol{\vartheta}}$, they contribute to the second addend in the righthand side of (4.47).

Finally, it remains to be considered the case when $\boldsymbol{x}_1 = \boldsymbol{x}_2$, and $\boldsymbol{x}_1$, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are linear dependent. There are at most $\binom{N}{N\boldsymbol{\vartheta}}8$ possible choices of pairs $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ and $(\boldsymbol{x}_2, \boldsymbol{y}_2)$ in $\big(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\big)^N_{\boldsymbol{\vartheta}}$ satisfying these requirements and for each of them

$$c\big(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2\big) \leq \mathbb{P}\left(\Psi^R_N \boldsymbol{x}_1 = \boldsymbol{0}, \Psi^R_N \boldsymbol{y}_1 = \boldsymbol{Z}_N\right) = \frac{1}{8^{2L}}\,.$$

Therefore, they contribute to the third addend in the righthand side of (4.47). ■

Let us define, for every $t \in [0, \log 8]$ the set

$$B_t := \left\{\boldsymbol{\vartheta} \in \mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right) \,:\, \mathrm{H}(\boldsymbol{\vartheta}) - \mathrm{H}\left(\pi^1_{\#}\boldsymbol{\vartheta}\right) \geq t\,,\ \mathrm{H}\left(\pi^1_{\#}\boldsymbol{\vartheta}\right) \geq t\right\}\,. \tag{4.48}$$

Clearly, we can rewrite $\overline{\delta}_\eta(R)$ defined in (4.46) as

$$\overline{\delta}_\eta(R) := \min\left\{\langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta\rangle \,\middle|\, \boldsymbol{\vartheta} \in B_{\overline{R}}\right\} \tag{4.49}$$

A second moment method based on Lemma 40 and Lemma 43 allows to show that, given a joint type $\boldsymbol{\vartheta}$ in the interior of $B_{\overline{R}}$, almost surely $U_N^R(\boldsymbol{\vartheta}) \geq 1$ definitively in N in $\mathcal{N}_{\boldsymbol{\vartheta}}$. This idea is exploited in proving the following almost sure upper bound on the asymptotic behaviour of the normalized minimum distance sequence.

Theorem 44 *For every design rate R in $(0, \log 8)$*

$$\mathbb{P}\left(\limsup_{N\in\mathbb{N}} \mathrm{d}_{\min}(\mathcal{U}_N^R) \leq \overline{\delta}_\eta(R)\right) = 1$$

Proof Let us fix an arbitrary $\varepsilon > 0$, and let $\boldsymbol{\vartheta}_\varepsilon$ in $B_{\overline{R}+\varepsilon}$ be such that

$$\overline{\delta}_\eta(R-\varepsilon) = \langle \boldsymbol{\vartheta}_\varepsilon, \boldsymbol{\Delta}_\eta \rangle .$$

Consider a sequence $(\boldsymbol{\vartheta}_N)_{N\in\mathbb{N}}$ converging to $\boldsymbol{\vartheta}_\varepsilon$, with $\boldsymbol{\vartheta}_N$ in $\mathcal{P}_N\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ for every N (which exists since $\mathcal{P}_{\mathbb{N}}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ is dense in $\mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$). We can now apply Chebyshev inequality and use Lemma 40 and Lemma 43 obtaining

$$\mathbb{P}\left(U_N^R\left(\boldsymbol{\vartheta}_N\right) = 0\right) \leq \frac{\mathrm{Var}\left[U_N^R(\boldsymbol{\vartheta}_N)\right]}{\left(\mathbb{E}\left[U_N^R(\boldsymbol{\vartheta}_N)\right]\right)^2} \leq 16\frac{\binom{N}{N\pi_\#^1\boldsymbol{\vartheta}_N}}{\binom{N}{N\boldsymbol{\vartheta}_N}}8^L + \frac{1}{\binom{N}{N\pi_\#^1\boldsymbol{\vartheta}}}8^L + 8\frac{1}{\binom{N}{N\boldsymbol{\vartheta}_N}}8^{2L} .$$

It follows that

$$\begin{aligned}\limsup_{N\in\mathbb{N}} \frac{\log \mathbb{P}\left(U_N^R(\boldsymbol{\vartheta}_N)=0\right)}{N} &\leq \overline{R} + \max\left\{ \mathrm{H}(\pi_\#^1\boldsymbol{\vartheta}_\varepsilon) - \mathrm{H}(\boldsymbol{\vartheta}_\varepsilon), -\mathrm{H}(\pi_\#^1\boldsymbol{\vartheta}_\varepsilon), \overline{R} - 2\,\mathrm{H}(\boldsymbol{\vartheta}_\varepsilon)\right\} \\ &\leq -\varepsilon < 0\,,\end{aligned}$$

so that $\sum_N \mathbb{P}\left(U_N^R\left(\boldsymbol{\vartheta}_N\right) = 0\right) < +\infty$, and by Borel-Cantelli lemma we have

$$\mathbb{P}\left(\left\{U_N^R\left(\boldsymbol{\vartheta}_N\right) = 0\right\} \text{ i.o. } N \in \mathbb{N}\right) = 0\,.$$

Therefore,

$$\mathbb{P}\left(\limsup_{N\in\mathbb{N}} \frac{1}{N} d_{\min}\left(\mathcal{U}_N^R\right) > \overline{\delta}(R-\varepsilon)\right) \leq \mathbb{P}\left(\left\{U_N^R(\boldsymbol{\vartheta}_N) = 0\right\} \text{ i.o. } N \in \mathbb{N}\right) = 0\,,$$

so that by the monotonicity and the continuity of $\overline{\delta}_\eta(R)$ we have the claim. ■

4.5.3 Comparing

We now want to compare the distance bounds $\underline{\delta}_\eta(\overline{R})$ and $\overline{\delta}_\eta(R)$, defined in (4.40) and (4.46) respectively, with the GV distance $\delta^{GV}(R)$. First, observe that any joint type $\boldsymbol{\vartheta}$ in $\mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ such that $\mathrm{H}(\boldsymbol{\vartheta}) - \mathrm{H}\left(\pi^1_\# \boldsymbol{\vartheta}\right) \geq \overline{R}$ and $\mathrm{H}\left(\pi^1_\# \boldsymbol{\vartheta}\right) \geq \overline{R}$ clearly satisfies $\mathrm{H}(\boldsymbol{\vartheta}) \geq 2\overline{R}$. From this it immediately follows that $\overline{\delta}_\eta(R) \geq \underline{\delta}_\eta(R)$. Notice also that the inequality above holds as an equality whenever $\underline{\delta}_\eta(R) = \langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle$ for some joint type $\boldsymbol{\vartheta}$ in $\mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ such that $\mathrm{H}\left(\pi^1_\# \boldsymbol{\vartheta}\right) = \overline{R}$. It can be shown that this is the case for every binary labeling $\eta : \mathbb{Z}_2^3 \to \mathbb{Z}_8$ for large enough values of R, so that actually in most cases $\overline{\delta}_\eta(R)$ and $\underline{\delta}_\eta(R)$ do coincide.

However, we will now concentrate on comparing $\overline{\delta}_\eta(R)$ with the GV distance $\delta^{GV}(R)$, in particular showing that the former is strictly below the latter. In order to do that, we start by considering some $\boldsymbol{\theta}$ in $\mathcal{P}(Z_8)$ giving the GV distance, i.e. such that $\delta_8(R) = \langle \boldsymbol{\theta}, \boldsymbol{\delta} \rangle$ and $\mathrm{H}(\boldsymbol{\theta}) \geq \overline{R}$. Since the map $\boldsymbol{\theta} \mapsto \langle \boldsymbol{\theta}, \boldsymbol{\delta} \rangle$ is linear and the entropy function is concave, with no loss of generality we can assume that $\mathrm{H}(\boldsymbol{\theta}) = \overline{R} \log 8$. Then, we can use Lagrangian multipliers in order to express $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta}(x) = \frac{e^{-\lambda \boldsymbol{\delta}(x)}}{Z(\lambda)}\,, \qquad Z(\lambda) := \sum_{x \in \mathbb{Z}_8} e^{-\lambda \boldsymbol{\delta}(x)}\,, \tag{4.50}$$

where λ in $(0, +\infty)$ is the unique solution of the equation $\mathrm{H}\left(Z(\lambda)^{-1} e^{-\lambda \boldsymbol{\delta}}\right) = \overline{R} \log 8$. From $\boldsymbol{\theta}$ we now construct a joint type $\boldsymbol{\vartheta}^*$ in $\mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$, defined by

$$\boldsymbol{\vartheta}^*(x, z) := \frac{1}{8} \boldsymbol{\theta}\left(\sigma_z(x)\right)\,, \qquad x, z \in \mathbb{Z}_2^3\,, \tag{4.51}$$

where the bijections $\sigma_z : \mathbb{Z}_2^3 \to \mathbb{Z}_8$ have been defined in (4.38), and let $\boldsymbol{\theta}^* := \pi^1_\# \boldsymbol{\vartheta}^*$ in $\mathcal{P}(\mathbb{Z}_2^3)$ be its marginal measure. Notice that from (4.50) we have

$$\boldsymbol{\vartheta}^*(x, z) > 0\,, \qquad \boldsymbol{\theta}^*(x) > 0\,, \qquad \forall\, x,\, z \in \mathbb{Z}_2^3\,. \tag{4.52}$$

We have

$$\langle \boldsymbol{\vartheta}^*, \boldsymbol{\Delta}_\eta \rangle = \sum_{x, z \in \mathbb{Z}_2^3} \boldsymbol{\vartheta}^*(x, z) \boldsymbol{\Delta}_\eta(x, z) = \sum_{z \in \mathbb{Z}_2^3} \frac{1}{8} \boldsymbol{\theta}\left(\sigma_z(x)\right) \boldsymbol{\delta}\left(\sigma_z(x)\right) = \langle \boldsymbol{\theta}, \boldsymbol{\delta} \rangle = \delta_8(R)\,. \tag{4.53}$$

From (4.51) we have $\sum_x \boldsymbol{\vartheta}^*(x, z) = \frac{1}{8} \sum_x \boldsymbol{\theta}(\sigma_z(x)) = \frac{1}{8}$ so that the marginal $\pi^2_\# \boldsymbol{\vartheta}^*$ is the uniform measure over $\mathbb{Z}_2^3$. Again from (4.51) we have that the conditioned measures satisfy $\boldsymbol{\vartheta}|_{\mathbb{Z}_2^3 \times \{z\}} = \boldsymbol{\theta} \circ \sigma_z$ for every z in $\mathbb{Z}_2^3$. Then, by applying (7.3) we have

$$\mathrm{H}(\boldsymbol{\vartheta}^*) = \mathrm{H}\left(\pi^2_\# \boldsymbol{\vartheta}^*\right) + \sum_{x \in \mathbb{Z}_2^3} \boldsymbol{\vartheta}^*\left(\mathbb{Z}_2^3 \times \{x\}\right) \mathrm{H}\left(\boldsymbol{\vartheta}^*|_{\mathbb{Z}_2^3 \times \{x\}}\right) = \log 8 + \mathrm{H}(\boldsymbol{\theta}) = \log 8 + \overline{R}\,. \tag{4.54}$$

Moreover, $\boldsymbol{\theta}^* = \pi^1_{\#}\boldsymbol{\vartheta}^* = \frac{1}{8}\sum_x \boldsymbol{\theta}\circ\sigma_x$ is a convex combination of permutations of the vector $\boldsymbol{\theta}$. As already observed, for every labeling $\eta : \mathbb{Z}_2^3 \to \mathbb{Z}_8$ there exists at least a pair of nonequal columns of the matrix $\boldsymbol{\Delta}_\eta$, say $\boldsymbol{\Delta}_\eta(\cdot, z_1) \neq \boldsymbol{\Delta}_\eta(\cdot, z_2)$. As a consequence, we have $\boldsymbol{\delta}\circ\sigma_{z_1} \neq \boldsymbol{\delta}\circ\sigma_{z_2}$ which, together with (4.50), implies $\boldsymbol{\theta}\circ\sigma_{z_1} \neq \boldsymbol{\vartheta}\circ\sigma_{z_2}$. Hence, from the strict concavity and the permutation invariance of the entropy function H it follows that

$$\mathrm{H}(\boldsymbol{\theta}^*) = \mathrm{H}\left(\tfrac{1}{8}\textstyle\sum_x \boldsymbol{\theta}\circ\sigma_x\right) > \tfrac{1}{8}\textstyle\sum_x \mathrm{H}\left(\boldsymbol{\theta}\circ\sigma_x\right) = \mathrm{H}(\boldsymbol{\theta}) = \overline{R}\,. \tag{4.55}$$

An immediate consequence is

$$\overline{\delta}_\eta(R) = \min\left\{f(\boldsymbol{\upsilon})\,|\,\boldsymbol{\upsilon}\in\mathcal{P}(\mathbb{Z}_2^3) : \mathrm{H}(\boldsymbol{\upsilon}) \geq \overline{R}\right\} \leq f(\boldsymbol{\theta}^*)\,, \tag{4.56}$$

where, for $\boldsymbol{\upsilon}$ in $\mathbb{Z}_2^3$, we define

$$f(\boldsymbol{\upsilon}) := \min\left\{\langle\boldsymbol{\vartheta},\boldsymbol{\Delta}_\eta\rangle|\,\boldsymbol{\vartheta}\in\mathcal{P}\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right) :\ \pi^1_{\#}\boldsymbol{\vartheta} = \boldsymbol{\upsilon}\,,\ \mathrm{H}(\boldsymbol{\vartheta}) - \mathrm{H}(\boldsymbol{\upsilon}) \geq \overline{R}\right\}.$$

We now present a lemma characterizing some properties of the function $f(\boldsymbol{\upsilon})$ defined above. For every x in $\mathbb{Z}_2^3$, define the minimum of the x-th row of the distance function $\boldsymbol{\Delta}_\eta$ as $m_x := \min\{\boldsymbol{\Delta}_\eta(x,z)|\,z\in\mathbb{Z}_2^3\}$, the set of elements achieving such a minimum as $M_x := \{z\in\mathbb{Z}_2^3 : \boldsymbol{\Delta}_\eta(x,z) = m_x\}$ and its cardinality by $n_x := |M_x|$. Since $\boldsymbol{\Delta}_\eta(0,z) = 0$ for every binary labeling η and every z in $\mathbb{Z}_2^3$, we have that $m_0 = 0$ and $n_0 = 8$. However, since no binary labeling η is isometric, there must exist x and z in $\mathbb{Z}_2^3$ such that

$$m_x < \boldsymbol{\Delta}_\eta(x,z)\,. \tag{4.57}$$

Lemma 45 *Let $\boldsymbol{\upsilon}$ in $\mathcal{P}(\mathbb{Z}_2^3)$ and R in $(0,\log 8)$. Then:*

1. *if $\sum_x \boldsymbol{\upsilon}(x)\log n_x \geq \overline{R}$, then $f(\boldsymbol{\upsilon}) = \sum_x \boldsymbol{\upsilon}(x)m_x$.*
2. *if $\sum_x \boldsymbol{\upsilon}(x)\log n_x < \overline{R}$ and $\boldsymbol{\vartheta}$ in $\mathcal{P}\left(\mathbb{Z}_2^3\times\mathbb{Z}_2^3\right)$ is such that $\pi^1_{\#}\boldsymbol{\vartheta} = \boldsymbol{\upsilon}$, $\mathrm{H}(\boldsymbol{\vartheta}) \geq \mathrm{H}(\boldsymbol{\upsilon}) + \overline{R}$, and $\langle\boldsymbol{\vartheta},\boldsymbol{\Delta}_\eta\rangle = f(\boldsymbol{\upsilon})$, then $\mathrm{H}(\boldsymbol{\vartheta}) = \mathrm{H}(\boldsymbol{\upsilon}) + \overline{R}$.*

Proof In order to prove point 1, notice that for every joint type $\boldsymbol{\vartheta}$ such that $\pi^1_{\#}\boldsymbol{\vartheta} = \boldsymbol{\upsilon}$

$$\langle\boldsymbol{\vartheta},\boldsymbol{\Delta}_\eta\rangle = \sum_{x,z}\boldsymbol{\vartheta}(x,z)\boldsymbol{\Delta}_\eta(x,z) \geq \sum_{x,z}\boldsymbol{\vartheta}(x,z)m_x = \sum_x \boldsymbol{\upsilon}(x)m_x\,.$$

Moreover, the joint type $\boldsymbol{\tau}(x,z) := \frac{1}{n_x}\boldsymbol{\upsilon}(x)\mathbb{1}_{M_x}(z)$ is such that $\langle\boldsymbol{\tau},\boldsymbol{\Delta}_\eta\rangle = \sum_x \boldsymbol{\upsilon}(x)m_x$, while $\pi^1_{\#}\boldsymbol{\tau} = \boldsymbol{\upsilon}$ and $\mathrm{H}(\boldsymbol{\tau}) = \mathrm{H}(\boldsymbol{\upsilon}) + \sum_x \boldsymbol{\upsilon}(x)\log n_x$. Then, whenever $\sum_x \boldsymbol{\upsilon}(x)\log n_x \geq \overline{R}$,

$$\sum_x \boldsymbol{\upsilon}(x)m_x \geq \min\left\{\langle\boldsymbol{\vartheta},\boldsymbol{\Delta}_\eta\rangle|\,\pi^1_{\#}\boldsymbol{\vartheta} = \boldsymbol{\upsilon}\,,\ \mathrm{H}(\boldsymbol{\vartheta}) - \mathrm{H}(\boldsymbol{\upsilon}) \geq \overline{R}\right\} \geq \langle\boldsymbol{\tau},\boldsymbol{\Delta}_\eta\rangle = \sum_x \boldsymbol{\upsilon}(x)m_x\,.$$

To prove point 2, let us assume without loss of generality that $f(\boldsymbol{\upsilon}) = \langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle$ for some joint type $\boldsymbol{\vartheta}$ in $\mathcal{P}\left(\mathbb{Z}_2^3 \times \mathbb{Z}_2^3\right)$ such that $\pi^1_\# \boldsymbol{\vartheta} = \boldsymbol{\upsilon}$ and $\mathrm{H}(\boldsymbol{\vartheta}) \geq \mathrm{H}(\boldsymbol{\upsilon}) + \overline{R}$. Notice that, if $\sum_x \boldsymbol{\upsilon}(x) \log n_x < \overline{R}$, then necessarily $\boldsymbol{\vartheta}(x,z) > 0$ for some x in $\mathbb{Z}_2^3$ and z not belonging to M_x, for otherwise

$$\mathrm{H}(\boldsymbol{\vartheta}) - \mathrm{H}(\boldsymbol{\upsilon}) = \sum_x \boldsymbol{\upsilon}(x)\, \mathrm{H}(\boldsymbol{\vartheta}|_{\{x\}\times\mathbb{Z}_2^3}) < \sum_x \boldsymbol{\upsilon}(x) \log |M_x| < \overline{R}\,.$$

It follows that $\langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle > \sum_x \boldsymbol{\upsilon}(x) m_x$.

Suppose that $\mathrm{H}(\boldsymbol{\vartheta}) > \overline{R} + \mathrm{H}(\boldsymbol{\upsilon})$. For every t in $[0,1]$ define a new joint type $\boldsymbol{\vartheta}_t := (1-t)\boldsymbol{\vartheta} + t\boldsymbol{\tau}$ interpolating $\boldsymbol{\vartheta}$ and $\boldsymbol{\tau}$. By the linearity of the marginal projection we have $\pi^1_\# \boldsymbol{\vartheta}_t = (1-t)\pi^1_\# \boldsymbol{\vartheta} + t\pi^1_\# \boldsymbol{\tau} = \boldsymbol{\upsilon}$. Since the entropy function is continuous, there exists some $\varepsilon > 0$ such that $\mathrm{H}(\boldsymbol{\vartheta}_t) \geq \overline{R} + \mathrm{H}(\boldsymbol{\upsilon})$ for every t in $(0, \varepsilon)$. Once fixed any such a t, we have

$$\langle \boldsymbol{\vartheta}_t, \boldsymbol{\Delta}_\eta \rangle \geq f(\boldsymbol{\upsilon}) = \langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle > (1-t)\langle \boldsymbol{\vartheta}, \boldsymbol{\Delta}_\eta \rangle + t\langle \boldsymbol{\tau}, \boldsymbol{\Delta}_\eta \rangle = \langle \boldsymbol{\vartheta}_t, \boldsymbol{\Delta}_\eta \rangle\,,$$

which is a contradiction. Then we have shown that necessarily $\mathrm{H}(\boldsymbol{\vartheta}) = \overline{R} + \mathrm{H}(\boldsymbol{\upsilon})$. ■

We are now ready to prove the following:

Theorem 46 *For any labeling $\eta : \mathbb{Z}_2^3 \to \mathbb{Z}_8$ and for every R in $(0, \log 8)$ we have*

$$\overline{\delta}_\eta(R) < \delta_8(R)\,.$$

Proof Thanks to (4.56) it is sufficient to show that $f(\boldsymbol{\theta}^*) < \delta^{GV}(\boldsymbol{\vartheta})$. We shall separately deal with the two alternatives

$$\textstyle\sum_x \boldsymbol{\theta}^*(x) \log n_x \geq \overline{R}\,, \tag{4.58}$$

and

$$\textstyle\sum_x \boldsymbol{\theta}^*(x) \log n_x < \overline{R}\,. \tag{4.59}$$

First suppose (4.58) holds. It follows from Lemma 45, (4.52), (4.57) and (4.53) that

$$\begin{aligned} f(\boldsymbol{\theta}^*) &= \textstyle\sum_x m_x \boldsymbol{\theta}^*(x) \\ &= \textstyle\sum_x \sum_z \frac{1}{8} \boldsymbol{\vartheta}(\sigma_z(x)) m_x \\ &= \textstyle\sum_x \sum_z \frac{1}{8} \boldsymbol{\vartheta}(\sigma_z(x)) \min_w \boldsymbol{\Delta}_\eta(x,w) \\ &< \textstyle\sum_x \sum_z \frac{1}{8} \boldsymbol{\vartheta}(\sigma_z(x)) \boldsymbol{\Delta}_\eta(x,z) \\ &= \delta^{GV}(R)\,, \end{aligned}$$

thus proving the claim.

Now suppose (4.59) holds. For any $x \neq 0$ in $\mathbb{Z}_2^3$, we have

$$\boldsymbol{\theta}^*(0) = \sum_{z \in \mathbb{Z}_2^3} \boldsymbol{\vartheta}^*(0,z) = \frac{1}{Z(\lambda)} > \sum_{z \in \mathbb{Z}_2^3} \frac{e^{-\lambda \boldsymbol{\delta}(\sigma_z(x))}}{8Z(\lambda)} = \sum_{z \in \mathbb{Z}_2^3} \boldsymbol{\vartheta}^*(x,z) = \boldsymbol{\theta}^*(x)\,.$$

Hence, $\boldsymbol{\theta}^*$ is not the uniform measure over $\mathbb{Z}_2^3$ and, as a consequence $\mathrm{H}(\boldsymbol{\theta}^*) < \log 8$. Therefore, from (4.54) and (4.55),

$$\mathrm{H}(\boldsymbol{\vartheta}^*) = \log 8 + \overline{R} > \mathrm{H}(\boldsymbol{\theta}^*) + \overline{R}\,.$$

Then, we can apply Lemma 45, obtaining that $\langle \boldsymbol{\vartheta}^*, \boldsymbol{\Delta}_\eta \rangle > f(\boldsymbol{\theta}^*)$. The claim now follows by applying (4.56) and (4.53). ■

A consequence of Theorem 44 and Theorem 46 is the following.

Corollary 47 *For every labeling η and for every design rate R in $(0, \log 8)$, with probability one the $\mathbb{Z}_2$-affine ensemble does not asymptotically achieve the Gilbert-Varshamov bound.*

4.6 Conclusions

In this chapter we have analyzed the asymptotic behavior of the minimal Bhattacharyya distance of Abelian group codes over symmetric channels. We have focused the 8-PSK AWGN channel. We have proven that typical $\mathbb{Z}_8$-codes achieve the GV bound, while a typical code sampled from the random coding ensemble does not. Finally, we have shown that the binary-affine ensemble does not achieve the GV bound with probability one. A lot more needs to be understood about this problem. As a first goal, we are currently trying to extend our final result to all $p^r - PSK$ constellations (where p is a prime number). We believe that this type of analysis is a first fundamental step to understand the behavior of more structured ensembles of codes, for instance LDPC or turbo group codes over non-binary symmetric channels.

Chapter 5

Average spectra and minimum distances of LDPC codes over Abelian groups

5.1 Introduction

Low-density parity-check (LDPC) codes have received a huge amount of attention in the last years. It is indeed the family of high-performance codes for which the deepest theoretical insight has been achieved. Their definition is quite simple: they are those binary linear codes which can be described as kernels of matrices over the binary field $\mathbb{Z}_2$ with a 'small' number of non-zero elements. Since the pioneering work [30], two streams of research are easily recognizable in the literature on LDPC codes. On the one hand, structural properties of such codes have been investigated: distance-spectra, minimum distances and also capacity estimations under maximum-likelihood (ML) decoding, [48, 50, 42, 59, 42, 43, 11, 16, 54]. On the other hand, they have been studied coupled with the well-known iterative decoding schemes [55, 56, 69, 51, 72, 40, 57, 15].

LDPC codes over non-binary Abelian groups were already introduced and studied in Gallager's seminal work [30]. More recently, after the rediscovery of Gallager codes in the '90s, non-binary LDPC codes have received a considerable amount of attention by researchers, and have been studied both for binary and non-binary channels. In the former case they allow to introduce a new design parameter, the choice of the non-zero entries in the parity matrix, to be optimized jointly with the degree profile. In the latter case they alow to design high-performance bandwidth-efficient coding schemes.

An interesting difference with respect to the binary case is the way to choose the non-zero elements of parity matrix. In this chapter we will consider many different possibilities. Among them, the so-called *unlabelled ensemble* where non-zero elements are

all equal to the identity, and the *uniformly labelled ensemble* where non-zero elements are instead, each one independently, chosen to be any possible automorphism of the group G with uniform probability. We will see that the latter ensemble will outperform the former. Of course our results could be extended to irregular LDPC ensembles, where the fraction of rows and columns with different amounts of non-zero entries (degree profile) is fixed, although this extension will not be considered here.

In [30] regular ensembles of LDPC $\mathbb{Z}_m$-codes were considered with all the non-zero entries equal to 1 (unlabelled ensembles in our terminology); he studied their Hamming distance-spectra and provided bounds for their error probabilities under maximum-likelihood and suboptimal iterative decoding over some highly symmetric channels. In [14], the authors show empirical evidence that, appropriately choosing the values of the non-zero entries in the parity check matrix, LDPC codes over the Galois field $\mathbb{F}_{2^r}$ perform better than the corresponding binary LDPC codes, when used over binary-input output-symmetric channels. LDPC codes over $\mathbb{F}_{2^r}$ for binary-input output-symmetric channels have also been studied in [53] following a density-evolution approach. The works [6, 7, 19] contain quite a complete theoretical analysis of LDPC codes over finite fields for non-binary channels considering both ML and belief-propagation decoding. Average type-spectra of regular LDPC ensembles over $\mathbb{Z}_p$ in the special case when p is prime, and more in general over $\mathbb{F}_{p^r}$, have been studied in [19, 6]. In this case the structural theory of binary LDPC codes generalizes in an almost straightforward way. In particular it has been shown, using expurgation techniques and results from [62], that average type-spectra provide lower bounds to the typical error exponent of these ensembles, and that this exponent can be made arbitrarily close to the random-coding one by allowing the density of the parity matrix to grow while keeping the rate constant.

However, in the case of algebraic structures which are not fields (e.g. $\mathbb{Z}_m$ with non-prime m), the available theoretical results are very few. In [6], average type-spectra of unlabelled ensembles of LDPC $\mathbb{Z}_m$-codes have also been studied in the case when m is not prime, but there is no results on minimum Euclidean distances. In the papers [66, 2, 73] the case when m is not prime has been considered, but mainly from an iterative-decoding perspective. Computer simulation have been reported in [66, 73] showing that, when mapped over the m-PSK constellation, LDPC $\mathbb{Z}_m$-codes guarantee better performance than their binary counterparts.

In this chapter we will study in detail average type-spectra and minimum Bhattacharyya-distances of regular LDPC ensembles over any finite Abelian group G, in which the non-zero entries of the parity-check operator are randomly sampled, independently and uniformly, from an arbitrary group F of authomorphisms of G (briefly F-labelled ensembles), generalizing all the results in [30, 14, 19, 6]. This extension passes through the use of mathematical tools which do not show up in the binary case: group characters, arithmetic concepts (Möbius inversion formula, Ramanujan sums), combinatorial techniques (Cayley graphs) and convex-analytical techniques.

As a first result, we will find exact expressions in terms of combinatorial formulas for the average type-spectra of regular F-labelled ensembles of LDPC codes over G: see Theorem 53. For the unlabelled ensemble of LDPC codes over $\mathbb{Z}_m$, we will show that our results for average type-spectra coincide with those obtained in [30, 6] , while for LDPC codes over finite fields the results of [14, 19, 6] will be recovered. Theorem 53 is instead completely original, to the best of our knowledge, for the uniformly labelled ensemble of LDPC codes over $\mathbb{Z}_m$, for which the average type-spectrum has an elegant expression in terms of Ramanujan sums. Coupling this analysis with an ad hoc analysis for the low-weight average type-enumerating functions, we will finally propose upper bounds to the repartition function of the minimum Bhattacharyya distance. This will allow us to show that minimum distances grow linearly in N with probability one (see Theorem 62): in the coding terminology this means that such codes are asymptotically good with probability one. More precisely, we obtain almost sure lower bounds on the asymptotic normalized minimum distance of the LDPC ensembles. These bounds are defined as solution of $(|G|-1)$-dimensional optimization problems. Proving the tightness of these bounds would require second-moment estimations for the type-enumerating functions, and is a problem left for future research. However, concentration results available in the literature for the Hamming distance-spectra of regular ensembles of binary LDPC codes (see [54]) make us optimistic about the tightness of our bounds for regular ensembles of LDPC G-codes as well. Finally, we will present some numerical results for the average distance-spectra showing how strongly the choice of the label group F affects the value of the typical minimum distance. In particular, we will show that, for the 8-PSK AWGN channel, the distance properties of the uniformly labelled ensemble of LDPC $\mathbb{Z}_8$-codes are significantly better than those of the unlabelled ensemble. This is confirmed by Monte-Carlo simulations of these codes which we have run, and it is coherent with some of the simulation results reported in [6].

The remainder of this chapter is organized as follows. In Section 2 we introduce LDPC code ensembles over Abelian groups. In Section 3 we study the average type-enumerating functions of these ensembles and we determine their exact growth-rate, namely the so-called average type-spectrum: the main result is Theorem 53. Section 4 is a technical one devoted to a detailed probabilistic analysis of low-weight codewords: the main result is Theorem 59. Using the results of Sections 3 and 4 we are finally able to prove, in Section 5, a probabilistic lower bound on the growth of minimum Euclidean distances for the LDPC ensembles when the block-length N goes to infinity: see Theorems 62 and 63. Finally, in Section 6 we report some numerical simulations showing that the uniformly labelled ensemble of LDPC $\mathbb{Z}_8$-codes definitely outperforms the unlabelled one on the 8-PSK AWGN channel, and we draw some final conclusions. An appendix completes the paper, containing some of the more technical proofs and a technical lemma on semicontinuous functions.

5.1.1 Low-density parity-check codes over Abelian groups

For any finite Abelian group G, we now describe the ensembles of LDPC G-codes which will be considered in this paper. For every given degree pair (c, d) in $\mathbb{N}^2$, we consider the set of admissible block-lengths $\mathcal{N}_{(c,d)} := \{N \in \mathbb{N} \text{s.t. } d \mid Nc\}$, and for every N in $\mathcal{N}_{(c,d)}$ define $L = Nc/d$. Consider the c-repetition operator

$$\mathrm{Rep}_c^N : G^N \to G^{Nc}\,, \qquad (\mathrm{Rep}_c^N\, \boldsymbol{x})_i = x_{\lceil i/c \rceil}\,, \tag{5.1}$$

where $\lceil x \rceil$ denotes the lowest integer not below x, and the d-check summation operator

$$\mathrm{Sum}_d^N : G^{Nc} \to G^L\,, \qquad (\mathrm{Sum}_d^N\, \boldsymbol{x})_i = \textstyle\sum_{k=i(d-1)+1}^{id} x_k\,. \tag{5.2}$$

Consider the group of permutations on Nc elements, S_{Nc}, and let Π'_N be a random variable uniformly distributed over S_{Nc}. Moreover, consider a subgroup F of $\mathrm{Aut}(G)$, the automorphism group of G, and let $(\Lambda_j)_{1 \le j \le Nc}$ be a family of independent random variables identically distributed uniformly on F, independent of Π'_N. Define the random diagonal automorphism $\Pi''_N \in \mathrm{Aut}(G^{Nc})$ by $(\Pi''_N \boldsymbol{x})_j := \Lambda_j x_j$ for $1 \le j \le Nc$. Finally, for every $N \in \mathcal{N}_{(c,d)}$ define the random syndrome homomorphism

$$\Phi_N : G^N \to G^L\,, \qquad \Phi_N := \mathrm{Sum}_d^N\, \Pi'_N \Pi''_N\, \mathrm{Rep}_c^N\,, \tag{5.3}$$

and the associated random G-code $\mathcal{C}_N := \ker \Phi_N$. This is called the (c, d)-*regular* F-*labelled ensemble*. F will be called the *label group*. The two extreme cases $F = \{1\}$ and $F = \mathrm{Aut}(G)$ will be referred to respectively as the *unlabelled* and the *uniformly labelled* (c, d)-regular ensembles.

The reason for considering only automorphisms as possible labels, avoiding the use of non-invertible labels, is clarified by the following proposition. For any group H, we denote the set of endomorphisms of H by $\mathrm{End}(H)$.

Proposition 48 *Assume that, for all $N \in \mathcal{N}_{(c,d)}$, $\Phi_N : G^N \to G^L$ is defined as in (5.3) with Π'_N uniformly distributed over S_{Nc} and $\Pi''_N \in \mathrm{End}(G^{Nc})$ is defined by $(\Pi''_N \boldsymbol{x})_j := \Lambda_j x_j$ for $1 \le j \le Nc$ where (Λ_j) are i.i.d. according to some probability distribution $\boldsymbol{\mu} \in \mathcal{P}(\mathrm{End}(G))$ such that $\mathrm{supp}(\boldsymbol{\mu}) \nsubseteq \mathrm{Aut}(G)$. Then, for al $k \in G \setminus \{0\}$ such that $\Lambda k = 0$ for some $\Lambda \in \mathrm{supp}(\boldsymbol{\mu})$*

$$\mathbb{P}(\mathrm{d}_{\min}(\ker \Phi_N) \le \boldsymbol{\delta}(k)) \ge 1 - (1 - \boldsymbol{\mu}(\Lambda)^c)^N \overset{N\to\infty}{\longrightarrow} 1.$$

Proof Consider $\Lambda \in \mathrm{supp}(\boldsymbol{\mu}) \setminus \mathrm{Aut}(G)$, and $k \in \ker \Lambda \setminus \{0\}$. For $1 \le s \le N$, let $e_s^k \in G^N$ be the N-tuple with all-zero entries but the s-th one which is equal to k. If

$\Lambda_j = \Lambda$ for all $(s-1)c+1 \leq j \leq sc$, then $\Pi''_N \operatorname{Rep}_c^N e_s^k = \mathbf{0}$, so that $\Phi_N e_s^k = \mathbf{0}$, and $\mathrm{d}_{\min}(\ker \Phi_N) \leq \boldsymbol{\delta}(k)$. Since the events

$$E_s^N := \bigcap_{(s-1)c+1 \leq j \leq sc} \{\Lambda_j = \Lambda\}$$

are independent for $1 \leq s \leq N$ and all have probability $1 - \boldsymbol{\mu}(\Lambda)^c$, it follows that

$$\mathbb{P}(\mathrm{d}_{\min}(\ker \Phi_N) \leq \boldsymbol{\delta}(k)) \geq \mathbb{P}\left(\bigcup_{1 \leq s \leq N} E_s^N\right) = 1 - \left(1 - \mathbb{P}\left(E_s^N\right)\right)^N = \left(1 - \boldsymbol{\mu}(\Lambda)^c\right)^N .$$

We wish to underline the fact that the proof of Proposition 48 strongly relied on the independence assumption we made for the labels Λ_j. Indeed, by introducing proper dependance structures for the random labels which allow to avoid certain configurations, it is possible to consider ensembles of LDPC G-codes with non-invertible labels as well. This possibility will not considered in the present paper, but will be explored in a future work.

As LDPC G-codes are special G-codes admitting sparse kernel representation, they suffer from all the limitations in performance of G-codes. In particular, the capacity they can achieve on a G-symmetric channel is upper bounded by the G-capacity of that channel. This explains why the authors of [6] had to restrict themselves to prime values of m while studying LDPC $\mathbb{Z}_m$-codes, albeit the average type-spectra they obtained for the unlabelled ensemble did not need such an assumption. In fact, they noticed that for non-prime m 'expurgation is impossible' and LDPC $\mathbb{Z}_m$-codes result 'bounded away from the random-coding spectrum'. The same restriction to prime values of m (or more in general to groups G admitting Galois field structure) was required both in [6] and [19] in order to study the uniformly labelled ensemble.

In this chapter regular ensembles of F-labelled *LDPC* G-codes will be studied for any finite Abelian group G. In particular we will find estimations for their average type-enumerating functions $\overline{W_{\mathcal{C}_N}(\boldsymbol{\theta})}$ and explicit combinatorial formulas for their average type-spectra defined as the limit of $N^{-1} \log \overline{W_{\mathcal{C}_N}(\boldsymbol{\theta})}$. Coupling this analysis with an ad hoc analysis of the type-enumerator functions for small weight codewords, we will finally propose upper bounds to the repartition function of the minimum normalized distance $\frac{1}{N} \mathrm{d}_{\min}(\mathcal{C}_N)$. This will allow to show that, if $c > 2$, minimum distances grow linearly in N with high probability. We will also show that the typical minimum distance (more precisely the lower bound on it -conjecture to be tight- provided by the average type-spectra) of the uniformly-labelled LDPC ensemble is significantly larger than the typical minimum distance of the corresponding unlabelled ensemble.

5.2 Average type-spectra of LDPC G-codes

In this section we first present some considerations about semidirect-product group actions. Then, in Sect.3.2 we introduce LDPC codes in a slightly more general setting and we show how regular F-labelled ensembles of LDPC G-codes introduced in Sect. 2.4 can be cast in this framework. In Sect.3.3 we prove the main result, Theorem 53, characterizing the average type-spectra of regular F-labelled ensembles. Finally, in Sect.3.4 we show how previous results in the literature can be recovered as particular cases of Theorem 53 and we provide an explicit formula for the average type-spectrum of the uniformly labelled ensemble over the cyclic group, which is instead an original result.

5.2.1 Group actions

We recall here some basic facts about semidirect group actions; the reader is referred to the standard textbook [37] for further details. Assume that a group F acts on a set A. A subset $B \subseteq A$ is said to be F-invariant if $fb \in B$ for every $b \in B$ and $f \in F$. Clearly, if B is F-invariant, F acts on B as well. For every a in A, the relative orbit $Fa := \{b \in A \text{s.t. } b = fa \text{ for some } f \in F\}$ is F-invariant and its action on it is transitive. The set of the orbits is denoted by A/F and called the quotient of A by the action of F. There is a canonical surjection $\pi_F : A \to A/F$ which associates an element a with the orbit it belongs to. Given $a \in A$, we define its stabilizer as $\mathrm{Stab}_F(a) := \{f \in F \text{s.t. } fa = a\}$. The well-known class formula gives: $|F| = |Fa| \cdot |\mathrm{Stab}_F(a)|$.

If A and B are sets and the group F acts on A, a map $\phi : A \to B$ is said to be F-invariant if $\phi(fa) = \phi(a)$ for every $a \in A$ and $f \in F$. As an example, the canonical surjection $\pi_F : A \to A/F$ is a F-invariant map. Suppose we have a F-invariant map $\phi : A \to B$, then it is immediate to see that we can define a map $\tilde{\phi} : A/F \to B$ such that $\phi = \tilde{\phi} \circ \pi_F$. Notice that if it happens that ϕ is onto and moreover $\phi(a) = \phi(a')$ if and only if $Fa = Fa'$, then the map $\tilde{\phi}$ is a bijection and thus A/F and B are in one-to-one correspondence. We will often use this fact in order to characterize quotient spaces.

We now introduce an example which will play a fundamental role in our future derivations. Given any set A, the permutation group S_N acts naturally on A^N: given $\boldsymbol{a} \in A^N$ and σ in S_N, we define $\sigma\boldsymbol{a} \in A^N$ by $(\sigma\boldsymbol{a})_j = \boldsymbol{a}_{\sigma^{-1}(j)}$. Orbits can easily described using types. Given $\boldsymbol{a}, \boldsymbol{b} \in A^N$, it is immediate to see that

$$\exists\, \sigma \in S_N \;:\; \sigma\boldsymbol{a} = \boldsymbol{b} \;\Leftrightarrow\; \boldsymbol{\theta}_A(\boldsymbol{a}) = \boldsymbol{\theta}_A(\boldsymbol{b})\,.$$

This says that the subsets $A^N_{\boldsymbol{\theta}}$ of type-$\boldsymbol{\theta}$ N-tuples are exactly the orbits for the action of the permutation group S_N on A^N, and we have a natural bijection $A^N/S_N \simeq \mathcal{P}_N(A)$ (obtained through the mapping $\boldsymbol{a} \mapsto \boldsymbol{\theta}(\boldsymbol{a})$.

Suppose now we are given an action of a group F on the set A. This extends to an action of F^N on A^N with the orbit set $A^N/F^N \simeq (A/F)^N$. We would like to combine this action, with the action of the permutation group on A^N and the way to do this is as follows: we consider the semidirect product

$$S_N \ltimes F^N\,, \quad (\sigma_1, \boldsymbol{g}_1)(\sigma_2, \boldsymbol{g}_2) = (\sigma_1\sigma_2, (\sigma_2^{-1}\boldsymbol{g}_1)\boldsymbol{g}_2)\,,$$

and the action on A^N given by $(\sigma, \boldsymbol{g})\boldsymbol{a} = \sigma(\boldsymbol{g}\boldsymbol{a})$.

We now want to characterize the set of orbits of this semidirect action. Notice that the map $\pi_F : A \to A/F$ induces a natural map $\pi_F^\sharp : \mathcal{P}(A) \to \mathcal{P}(A/F)$ where $[\pi_F^\sharp \boldsymbol{\theta}](Fa) = \sum_{b \in Fa} \boldsymbol{\theta}(b)$. It is easy to see that the following diagram commutes

$$\begin{array}{ccc} A^N & \stackrel{\pi_{F^N}}{\to} & (A/F)^N \\ \downarrow \boldsymbol{\theta}_A & & \downarrow \boldsymbol{\theta}_{A/F} \\ \mathcal{P}_N(A) & \stackrel{\pi_F^\sharp}{\to} & \mathcal{P}_N(A/F) \end{array} \tag{5.4}$$

(i.e. $\boldsymbol{\theta}_{A/F} \circ \pi_{F^N} = \pi_F^\sharp \circ \boldsymbol{\theta}_A$).

In the sequel we will use the notation $\boldsymbol{v}_{A,F} = \boldsymbol{\theta}_{A/F} \circ \pi_{F^N}$ and call $\boldsymbol{v}_{A,F}(\boldsymbol{a})$ the (A,F)-*type* of $\boldsymbol{a}$. Clearly, The (A,F)-type is exactly what is needed to describe orbits with respect to the action of the semidirect group $S_N \ltimes F^N$. Indeed, it is immediate to check that $\mathcal{P}_N(A/F)$ is in bijection with the quotient $A^N/(S_N \ltimes F^N)$: given $\boldsymbol{a}, \boldsymbol{b} \in A^N$ we have that

$$\exists (\sigma, \boldsymbol{g}) \in S_N \ltimes F^N \text{ s.t. } (\sigma, \boldsymbol{g})\boldsymbol{a} = \boldsymbol{b} \;\Leftrightarrow\; \boldsymbol{v}_{A,F}(\boldsymbol{a}) = \boldsymbol{v}_{A,F}(\boldsymbol{b})\,.$$

If $\boldsymbol{v} \in P_N(A/F)$ we will use the notation $A^N_{\boldsymbol{v}} := \{\boldsymbol{a} \in A^N \mid \boldsymbol{v}_{A,F}(\boldsymbol{a}) = \boldsymbol{v}\}$. Using the fact that $\boldsymbol{v}_{A,F} = \boldsymbol{\theta}_{A/F} \circ \pi_{F^N}$ we obtain that

$$|A^N_{\boldsymbol{v}}| = \tbinom{N}{N\boldsymbol{v}} \prod_{\alpha \in A/F} |\pi_F^{-1}(\alpha)|^{N\boldsymbol{v}(\alpha)}\,. \tag{5.5}$$

Define now $\mathcal{O}^N_{\boldsymbol{v}} := \left\{\boldsymbol{\theta} \in \mathcal{P}_N(A) \text{s.t. } \pi_F^\sharp(\boldsymbol{\theta}) = \boldsymbol{v}\right\}$. For every given $\boldsymbol{v} \in \mathcal{P}(A/F)$, and N in $\mathbb{N}$, we have

$$A^N_{\boldsymbol{v}} = \bigcup_{\boldsymbol{\theta} \in \mathcal{O}^N_{\boldsymbol{v}}} A^N_{\boldsymbol{\theta}}\,, \tag{5.6}$$

the union being disjoint. Notice that we also have $\left|\mathcal{O}^N_{\boldsymbol{v}}\right| = \prod_{\alpha \in A/F} |\pi_F^{-1}(\alpha)|^{N\boldsymbol{v}(\alpha)}$.

5.2.2 A general framework for LDPC ensembles over Abelian groups

Fix an infinite subset $\mathcal{N} \subseteq \mathbb{N}$, a group U, two sequences of finite Abelian groups $Z^{(N)}$ and $Y^{(N)}$ (with $N \in \mathcal{N}$), and two sequences of homomorphisms

$$\Xi_o^N : U^N \to Z^{(N)}\,, \quad \Xi_i^N : Z^{(N)} \to Y^{(N)}\,.$$

Consider moreover a sequence I_N of subgroups of $\mathrm{Aut}(Z^{(N)})$, and assume that the actions of I_N on $Z^{(N)}$ satisfy the following property: there exists a fixed finite set A and a sequence of invariant maps $\Theta_N : Z^{(N)} \to \mathcal{P}(A)$ such that $\boldsymbol{x}$, $\boldsymbol{y} \in Z^{(N)}$ are in the same orbit if and only if $\Theta_N(\boldsymbol{x}) = \Theta_N(\boldsymbol{y})$. In this way the quotient space $Z^{(N)}/I_N$ can be naturally identified with the image of Θ_N inside $\mathcal{P}(A)$.

Let now Π_N be a sequence of random variables uniformly distributed over I_N. For every $N \in \mathcal{N}$ define

$$\Phi_N := \Xi_i^N \Pi_N \Xi_o^N \,, \tag{5.7}$$

The triple (Ξ_o^N, Ξ_i^N, I_N) is called an *interconnected ensemble* while $(\ker \Phi_N)$ will be the *random code sequence* associated to the ensemble. The set A will be called the *interconnection type alphabet* of the ensemble.

Consider now the type-enumerating function $W_N(\boldsymbol{\theta})$ for the ensemble. By taking the expectation with respect to our probability space, we get

$$\overline{W_N(\boldsymbol{\theta})} = \mathbb{E}\Big[\sum_{\boldsymbol{x} \in U_{\boldsymbol{\theta}}^N} \mathbb{1}_{\{\mathbf{0}\}}(\Phi_N \boldsymbol{x}) \Big] = \sum_{\boldsymbol{x} \in U_{\boldsymbol{\theta}}^N} \mathbb{P}(\Phi_N \boldsymbol{x} = \mathbf{0}) \,. \tag{5.8}$$

Put $Z_{\boldsymbol{\upsilon}}^{(N)} := \Theta_N^{-1}(\boldsymbol{\upsilon})$ and define the following sets: for every $\boldsymbol{\upsilon} \in \mathcal{P}(A)$, $\boldsymbol{\theta} \in \mathcal{P}(U)$

$$Z_{\boldsymbol{\upsilon}}^{i,N} := \Big\{ \boldsymbol{w} \in Z_{\boldsymbol{\upsilon}}^{(N)} \mid \Xi_i^N \boldsymbol{w} = \mathbf{0} \Big\} \,, \quad U_{\boldsymbol{\theta},\boldsymbol{\upsilon}}^{o,N} := \{ \boldsymbol{x} \in U^N \mid \boldsymbol{\theta}_U(\boldsymbol{x}) = \boldsymbol{\theta} \,,\, \Theta_N(\Xi_o^N \boldsymbol{x}) = \boldsymbol{\upsilon} \} \,. \tag{5.9}$$

We have the following simple result.

Proposition 49 *For every $\boldsymbol{\theta}$ in $\mathcal{P}_N(U)$*

$$\overline{W_N(\boldsymbol{\theta})} = \sum_{\boldsymbol{\upsilon} \in \mathcal{P}(A)} \frac{|U_{\boldsymbol{\theta},\boldsymbol{\upsilon}}^{o,N}||Z_{\boldsymbol{\upsilon}}^{i,N}|}{|Z_{\boldsymbol{\upsilon}}^{(N)}|} \,. \tag{5.10}$$

Proof If $\boldsymbol{x} \in U_{\boldsymbol{\theta},\boldsymbol{\upsilon}}^{o,N}$, using the fact that I_N acts transitively on $Z_{\boldsymbol{\upsilon}}^{(N)}$ and the class formula, we obtain

$$\mathbb{P}(\Phi_N \boldsymbol{x} = \mathbf{0}) = \mathbb{P}(\Pi_N \Xi_o^N \boldsymbol{x} \in Z_{\boldsymbol{\upsilon}}^{i,N}) = \frac{|Z_{\boldsymbol{\upsilon}}^{i,N}||\mathrm{Stab}_{I_N}(\Xi_o^N(\boldsymbol{x}))|}{|I_N|} = \frac{|Z_{\boldsymbol{\upsilon}}^{i,N}|}{|Z_{\boldsymbol{\upsilon}}^{(N)}|} \,.$$

Using now (5.8), (5.10) follows immediately.

We now frame the LDPC ensembles introduced in Section 2 into this more general setting. We use the notation introduced in Section 5.1.1. Given $(c,d) \in \mathbb{N}^2$ and $N \in \mathcal{N}_{(c,d)}$, consider $L = Nc/d$. Take $U = G$, $Z^{(N)} = G^{Nc}$, $Y^{(N)} = G^L$. Also, take $\Xi_o^N =$

Rep_c^N, $\Xi_i^N = \mathrm{Sum}_d^N$, $I_N = S_{Nc} \ltimes F^{Nc}$. The ensemble $(\mathrm{Rep}_c^N, \mathrm{Sum}_d^N, S_{Nc} \ltimes F^{Nc})$ is the (c,d)-regular F-labelled ensemble. The type alphabet in this case is simply $A = G/F$.

Irregular ensembles can be framed into this setting by simply replacing the repetition operator. Also other interesting cases can be obtained by considering the interconnections among the inner and outer encoder done through some vector structured channels and allowing only independent permutations on the various channels. However, will now focus on the evaluation of the type-spectra of the regular F-labelled LDPC G-code ensembles. This will be done in the following subsection by explicitly calculating the three terms entering in the formula (5.10).

5.2.3 The average type-spectrum of the (c,d)-regular F-labelled ensemble

In order to prove the main result of this section we will use some generating functions techniques. For a finite set A, consider the ring of complex-coefficients multivariable polynomials (briefly multinomials) $\mathbb{C}[A]$. Given $p \in \mathbb{C}[A]$ and $\boldsymbol{k} \in \mathbb{Z}_+^A$ we denote by $\lfloor p(\boldsymbol{z}) \rfloor_{\boldsymbol{k}}$ the coefficient of the term $\boldsymbol{z}^{\boldsymbol{k}}$ in $p(\boldsymbol{z})$, i.e. $p(\boldsymbol{z}) = \sum_{\boldsymbol{k} \in \mathbb{Z}_+^A} \lfloor p(\boldsymbol{z}) \rfloor_{\boldsymbol{k}} \boldsymbol{z}^{\boldsymbol{k}}$. In particular, we will consider type-enumerating multinomials, i.e. homogeneous-degree multinomials of the form $p(\boldsymbol{z}) = \sum_{\boldsymbol{\theta} \in \mathcal{P}_N(A)} \lfloor p(\boldsymbol{z}) \rfloor_{N\boldsymbol{\theta}} \boldsymbol{z}^{N\boldsymbol{\theta}}$, where each coefficient $\lfloor p(\boldsymbol{z}) \rfloor_{N\boldsymbol{\theta}}$ equals the number of N-tupes $\boldsymbol{a} \in A^N$ of A-type $\boldsymbol{\theta}$, satisfying certain properties. The easiest case is provided by the multinomial $(\sum_{a \in A} z_a)^N = \sum_{\boldsymbol{\theta} \in \mathcal{P}_N(A)} \binom{N}{N\boldsymbol{\theta}} \boldsymbol{z}^{N\boldsymbol{\theta}}$, simply enumerating the N-tuples of different A-types. The following result, proved in [11], characterizes the asymptotic growth-rate of the coefficients of powers of enumerating multinomials.

Theorem 50 *Let A be a finite set, and $p(\boldsymbol{z}) \in \mathbb{R}_+[A]$ be a homogeneous-degree, non-negative real-coefficients multinomial. For all $\boldsymbol{\theta} \in \mathcal{P}_{\mathbb{N}}(A)$ and $\boldsymbol{z} \in \mathcal{P}(A)$ such that $\mathrm{supp}(\boldsymbol{z}) = \mathrm{supp}(\boldsymbol{\theta})$, we have*

$$\left\lfloor p(\boldsymbol{z})^N \right\rfloor_{N\boldsymbol{\theta}} \leq \frac{p(\boldsymbol{z})^N}{\boldsymbol{z}^{N\boldsymbol{\theta}}}\,, \qquad \lim_{N \in \mathcal{N}_{\boldsymbol{\theta}}} \frac{1}{N} \log \left\lfloor p(\boldsymbol{z})^N \right\rfloor_{N\boldsymbol{\theta}} = \inf_{\substack{\boldsymbol{z} \in \mathcal{P}(A): \\ \mathrm{supp}(\boldsymbol{z}) = \mathrm{supp}(\boldsymbol{\theta})}} \log \frac{p(\boldsymbol{z})}{\boldsymbol{z}^{\boldsymbol{\theta}}}\,. \tag{5.11}$$

Moreover the left-hand side of (5.11) is a concave (and thus upper semicontinuous) $[-\infty, +\infty)$-valued function on $\mathcal{P}(A)$.

The first type-enumerating multinomial which we will need in our derivations is the one enumerating the 0-sum d-tuples over a finite Abelian group G:

$$\beta_d(\boldsymbol{z}) \in \mathbb{C}[z_g,\, g \in G]\,, \qquad \beta_d(\boldsymbol{z}) := \sum_{g_1, \ldots, g_d} \mathbb{1}_{\{0\}} \left(\textstyle\sum_{k=1}^d g_k \right) \prod_{1 \leq k \leq d} z_{g_k}\,.$$

By introducing the group $\hat{G}$ of characters of G, i.e. homomorphisms of G in the multiplicative group $\mathbb{C}^*$ of non-zero complex numbers, it is possible to find an explicit expression for $\beta_d(\boldsymbol{z})$ as stated in the following lemma.

Lemma 51 *For every finite Abelian group G and $d \in \mathbb{N}$*

$$\beta_d(\boldsymbol{z}) = \frac{1}{|G|} \sum_{\chi \in \hat{G}} \left(\sum_{g \in G} z_g \chi(g) \right)^d .$$

Proof The inversion formula for the discrete Fourier transform (see [67, pag. 168]) $f(g) = \frac{1}{|G|} \sum_\chi \langle f, \chi \rangle \chi(g)$, applied to $f = \delta_0 \in L^2(G)$, gives $\frac{1}{G} \sum_\chi \chi(g) = \mathbb{1}_{\{0\}}(g)$. Then,

$$\begin{aligned}
\beta_d(\boldsymbol{z}) &= \sum_{g_1,\dots,g_d} \mathbb{1}_{\{0\}} \left(\textstyle\sum_{1 \le k \le d} g_k \right) \prod_{1 \le k \le d} z_{g_k} \\
&= \sum_{g_1,\dots,g_d} \tfrac{1}{|G|} \sum_\chi \chi \left(\textstyle\sum_{1 \le k \le d} g_k \right) \prod_{1 \le k \le d} z_{g_k} \\
&= \tfrac{1}{|G|} \sum_\chi \sum_{g_1,\dots,g_d} \prod_{1 \le k \le d} \chi(g_k)\, z_{g_k} \\
&= \tfrac{1}{|G|} \sum_\chi \Big(\sum_g z_g \chi(g) \Big)^d .
\end{aligned}$$

Recall that, given any subgroup F of $\mathrm{Aut}(G)$ and a degree pair (c,d) in $\mathbb{N}^2$, the (c,d)-regular F-labelled ensemble of LDPC G-codes is described by the triple $\big(\mathrm{Rep}_c^N, \mathrm{Sum}_d^N, S_{Nc} \ltimes F^{Nc}\big)$. Let $\pi_F : G \to G/F$ be the canonical projection on the quotient, and $\pi_F^\sharp : \mathcal{P}(G) \to \mathcal{P}(G/F)$ be the associated action on probabilities. Also, define

$$\boldsymbol{\varphi} : G/F \to \mathbb{N}\,, \qquad \boldsymbol{\varphi}(q) = \left| \pi_F^{-1}(q) \right| , \tag{5.12}$$

to be the map giving the cardinalities of the orbits of G under the action of F.

Consider some admissible block-length N in $\mathcal{N}_{(c,d)}$. Formula (5.5) shows that $|Z_{\boldsymbol{v}}^{(N)}| = \binom{Nc}{Nc\boldsymbol{v}} \boldsymbol{\varphi}^{Nc\boldsymbol{v}}$ for every $\boldsymbol{v} \in \mathcal{P}_{Nc}(G/F)$. Moreover, in this case $|U_{\boldsymbol{\theta},\boldsymbol{v}}^{o,N}| = \binom{N}{N\boldsymbol{\theta}} \mathbb{1}_{\{\pi_F^\sharp \boldsymbol{\theta}\}}(\boldsymbol{v})$. Substituting in (5.10), and defining $\boldsymbol{v} := \pi_F^\sharp \boldsymbol{\theta}$, we obtain

$$\overline{W_N(\boldsymbol{\theta})} = \binom{N}{N\boldsymbol{\theta}} \binom{Nc}{Nc\boldsymbol{v}}^{-1} \boldsymbol{\varphi}^{-Nc\boldsymbol{v}} |Z_{\boldsymbol{v}}^{i,N}| \,. \tag{5.13}$$

It remains to evaluate the enumerating weights $|Z_{\boldsymbol{v}}^{i,N}|$ relative to the check summation operator. In order to do that, we inroduce the multinomial

$$\alpha_{F,d}(\boldsymbol{t}) \in \mathbb{C}[t_q,\, q \in G/F]\,, \qquad \alpha_{F,d}(\boldsymbol{t}) := \frac{1}{|G|} \sum_{\chi \in \hat{G}} \left(\sum_{q \in G/F} \frac{1}{\boldsymbol{\varphi}(q)} \sum_{g \in q} \chi(g) t_q \right)^d , \tag{5.14}$$

and present the following result, stating that the L-th power of $\alpha_{F,d}(\boldsymbol{t})$ is the type-enumerating multinomial of the normalized weights $|Z_{\boldsymbol{v}}^{i,N}|/\boldsymbol{\varphi}^{Nc\boldsymbol{v}}$.

Lemma 52 *For every* $N \in \mathcal{N}_{(c,d)}$

$$\sum_{\boldsymbol{v}\in\mathcal{P}_{Nc}(G/F)} \frac{|Z_{\boldsymbol{v}}^{i,N}|}{\boldsymbol{\varphi}^{Nc\boldsymbol{v}}}\boldsymbol{t}^{Nc\boldsymbol{v}} = \left(\alpha_{F,d}(\boldsymbol{t})\right)^{L} . \tag{5.15}$$

Proof First consider the type-enumerating multinomial $B(\boldsymbol{z}) \in \mathbb{C}[z_g,\, g \in G]$ for the kernel of the inner homomorphism $\Xi_i^N = \mathrm{Sum}_d^N$. Since any $\boldsymbol{x}$ in G^{Nc} belongs to $\ker \mathrm{Sum}_d^N$ iff it is the concatenation of L 0-sum d-tuples, from Lemma 51 we have $B(\boldsymbol{z}) = \left(\beta_d(\boldsymbol{z})\right)^L$. Consider now the map

$$\Psi : \mathbb{C}[z_g,\, g \in G] \to \mathbb{C}[t_q\,,\ q \in G/F] \qquad \Psi : p(\boldsymbol{z}) \to p(t_{\pi_F(g)},\, g \in G)\,.$$

It follows from (5.6) that, for all $\boldsymbol{v}$ in $\mathcal{P}(G/F)$, we have

$$\frac{|Z_{\boldsymbol{v}}^{i,N}|}{\boldsymbol{\varphi}^{Nc\boldsymbol{v}}} = \sum_{\boldsymbol{\theta}\in\mathcal{O}_{\boldsymbol{v}}^{Nc}} \frac{\lfloor B(\boldsymbol{z})\rfloor_{Nc\boldsymbol{\theta}}}{\boldsymbol{\varphi}^{Nc\boldsymbol{v}}} = \sum_{\boldsymbol{v}\in\mathcal{P}_{Nc}(G/F)} \frac{\lfloor \Psi B\,(\boldsymbol{t})\rfloor_{Nc\boldsymbol{v}}}{\boldsymbol{\varphi}^{Nc\boldsymbol{v}}} = \sum_{\boldsymbol{v}\in\mathcal{P}_{Nc}(G/F)} \left\lfloor \Psi B\left(\frac{\boldsymbol{t}}{\boldsymbol{\varphi}}\right)\right\rfloor_{Nc\boldsymbol{v}} . \tag{5.16}$$

Then, the claim follows by observing that $\Psi B\left(\boldsymbol{t}/\boldsymbol{\varphi}\right) = \left(\Psi\beta_d\left(\boldsymbol{t}/\boldsymbol{\varphi}\right)\right)^L = \alpha_{F,d}(\boldsymbol{t})^L$.

We are now ready to prove the main result of this section, stating that the average type-spectrum of the (c,d)-regular F-labelled ensemble of LDPC G-codes is given by

$$\Gamma_{(F,c,d)}(\boldsymbol{\theta}) := \mathrm{H}(\boldsymbol{\theta}) + \frac{c}{d} \inf_{\substack{\boldsymbol{t}\in\mathcal{P}(G/F):\\ \mathrm{supp}(\boldsymbol{t})=\mathrm{supp}(\boldsymbol{\pi}_F^{\sharp}\boldsymbol{\theta})}} \left\{\log \alpha_{F,d}(\boldsymbol{t}) + dD\big(\boldsymbol{\pi}_F^{\sharp}\boldsymbol{\theta}||\boldsymbol{t}\big)\right\} . \tag{5.17}$$

From Theorem 50 it follows that the spectrum $\Gamma_{(F,c,d)}(\boldsymbol{\theta})$ is an upper semicontinuous function on the probability simplex $\mathcal{P}(G)$. Notice that, by choosing $\boldsymbol{t} = \boldsymbol{\pi}_F^{\sharp}\boldsymbol{\theta}$, we immediately obtain the estimation

$$\Gamma_{(F,c,d)}(\boldsymbol{\theta}) \le \frac{c}{d}\log \alpha_{F,d}\left(\pi_F^{\sharp}\boldsymbol{\theta}\right) + \mathrm{H}(\boldsymbol{\theta})\,.$$

Theorem 53 *For the* (c,d)*-regular* F*-labelled ensemble of LDPC* G*-codes*

$$\overline{W_N(\boldsymbol{\theta})} \le \exp(N\Gamma_{(F,c,d)}(\boldsymbol{\theta}))\ , \qquad \lim_{N\in\mathcal{N}_{\boldsymbol{\theta}}\cap\mathcal{N}_{(c,d)}} \frac{1}{N}\log \overline{W_N(\boldsymbol{\theta})} = \Gamma_{(F,c,d)}(\boldsymbol{\theta})\,.$$

Proof From (5.13), by recalling that $Nc = Ld$, we get

$$\frac{1}{N}\log\overline{W_N(\boldsymbol{\theta})} = \frac{1}{N}\log\binom{N}{N\boldsymbol{\theta}} + \frac{c}{d}\frac{1}{L}\log\frac{|Z_{\boldsymbol{\upsilon}}^{i,N}|}{\binom{Ld}{Ld\boldsymbol{\upsilon}}\boldsymbol{\varphi}^{Ld\boldsymbol{\upsilon}}}\,.$$

We have $\lim\frac{1}{N}\log\binom{N}{N\boldsymbol{\theta}} = \mathrm{H}(\boldsymbol{\theta})$. Then we can apply first Lemma 52 and then Theorem 50 (notice that (5.15) with $L = 1$ implies that $\alpha_{F,d}(\boldsymbol{t})$ has non-negative real coefficients and homogeneous degree), obtaining

$$\lim_N \frac{1}{L}\log\frac{|Z_{\boldsymbol{\upsilon}}^{i,N}|}{\binom{Ld}{Ld\boldsymbol{\upsilon}}\boldsymbol{\varphi}^{Ld\boldsymbol{\upsilon}}} = \lim_N \frac{1}{L}\log\frac{\left\lfloor\alpha_{F,d}(\boldsymbol{t})^L\right\rfloor_{Ld\boldsymbol{\upsilon}}}{\binom{Ld}{Ld\boldsymbol{\upsilon}}\boldsymbol{\varphi}^{Ld\boldsymbol{\upsilon}}} = \inf_{\substack{\boldsymbol{t}\in\mathcal{P}(G/F):\\ \mathrm{supp}(\boldsymbol{t})=\mathrm{supp}(\boldsymbol{\upsilon})}}\left\{\log\frac{\alpha_{F,d}(\boldsymbol{t})}{\boldsymbol{t}^{d\boldsymbol{\upsilon}}} - d\,\mathrm{H}(\boldsymbol{\upsilon})\right\}$$

Similarly, the inequality is proven.

5.2.4 Special cases of Theorem 53

Now, we particularize Theorem 53 to some important special cases, showing as all the previous results in the literature of non-binary LDPC codes can be reobtained, and other interesting cases can be studied as well.

LDPC codes over Galois fields

Suppose $G \simeq \mathbb{Z}_p^r$ for some prime number p and positive integer r. First let F coincide with the whole automorphism group $\mathrm{Aut}(\mathbb{Z}_p^r)$, which is isomorphic to the general linear group of $r \times r$ invertible matrices on $\mathbb{Z}_p$. In this case the probability that an N-tuple $\boldsymbol{x}$ in G^N belongs to the random LDPC code $\mathcal{C}_N = \ker\left(\mathrm{Sum}_d^N\,\Pi_N\,\mathrm{Rep}_c^N\right)$ only depends on the Hamming weight (i.e. number of non-zero entries) of $\boldsymbol{x}$. Indeed, it is easily seen that the action of $\mathrm{Aut}(\mathbb{Z}_p^r)$ on $\mathbb{Z}_p^r$ has only two orbits: one containing the zero element only, and one containing all the non-zero elements of $\mathbb{Z}_p^r$. Thus, the quotient space is $G/F = \{q_0, q_1\}$, with $\boldsymbol{\varphi}(q_0) = 1$, $\boldsymbol{\varphi}(q_1) = p^r - 1$. Moreover, since all nontrivial characters are orthogonal to the trivial one $\chi_0 \equiv 1$, it follows that $\sum_{g\in q_1}\chi(g) = -\chi(0) = -1$ for all $\chi \in \hat{G}\setminus\{\chi_0\}$. Then, the average type-spectra of the (c,d)-regular $\mathrm{Aut}(\mathbb{Z}_p^r)$-labelled ensemble of LDPC $\mathbb{Z}_p^r$-codes is given by

$$\Gamma_{(c,d,\mathrm{Aut}(\mathbb{Z}_p^r))}(\boldsymbol{\theta}) = \mathrm{H}(\boldsymbol{\theta}) + \tfrac{c}{d}\inf_{t\in(0,1)}\left\{\log\left(\tfrac{1}{p^r} + \tfrac{p^r-1}{p^r}\left(1 - \tfrac{p^r}{p^r-1}t\right)^d\right) + dD(\lambda||t)\right\}\,, \quad (5.18)$$

where $\lambda := 1 - \boldsymbol{\theta}(0)$ and $D(\lambda||t) := \lambda\log\frac{\lambda}{t} + (1-\lambda)\log\frac{1-\lambda}{1-t}$.

Consider now the case $G \simeq \mathbb{Z}_p^r$ again, but now with label group $F \simeq \mathbb{F}_{p^r}^*$, the multiplicative group of non-zero elements of the Galois field $\mathbb{F}_{p^r}$. Observe that $\mathbb{F}_{p^r}^*$

can always be identified with a subgroup (proper if $r > 1$) of $\mathrm{Aut}(\mathbb{Z}_p^r)$. Nevertheless, the action of $\mathbb{F}_{p^r}^*$ on $\mathbb{Z}_p^r$ has the same two orbits as the action of the whole $\mathrm{Aut}(\mathbb{Z}_{p^r})$ on $\mathbb{Z}_p^r$. This shows that the (c,d)-regular $F_{p^r}^*$-labelled ensemble has the same average type-spectrum of the $\mathrm{Aut}(\mathbb{Z}_p^r)$-labelled ensemble, i.e.

$$\Gamma_{(\mathbb{F}_{p^r}^*,c,d)}(\boldsymbol{\theta}) = \Gamma_{(c,d,\mathrm{Aut}(\mathbb{Z}_p^r))}(\boldsymbol{\theta})\,, \qquad \forall\, \boldsymbol{\theta} \in \mathcal{P}(\mathbb{Z}_p^r)\,. \tag{5.19}$$

The expression (5.18) coincides with the spectrum of the $\mathbb{F}_{p^r}^*$-labelled ensemble obtained in [6, 19]. We observe that in [53] it was numerically observed that the density-evolution dynamical system [55] exhibits the same threshold value for the $\mathbb{F}_{p^r}^*$-labelled and the $\mathrm{Aut}(\mathbb{Z}_p^r)$-labelled ensembles over the binary erasure channel. Formula (5.19) shows that these ensembles have identical average type-spectra.

Unlabelled LDPC ensembles over cyclic groups

We now consider the case when $G \simeq \mathbb{Z}_m$ and $F = \{1\}$. In this case, the characters of $\mathbb{Z}_m$ are given by $\chi_k(h) := e^{\frac{2\pi}{m}hki}$ for $h, k \in \mathbb{Z}_m$, while, trivially, the quotient space $\mathbb{Z}_m/F$ coincides with $\mathbb{Z}_m$ itself and $\boldsymbol{\varphi} \equiv 1$ (see (5.12)). It follows that

$$\alpha_{\{1\},d}(\boldsymbol{t}) = \beta_d(\boldsymbol{t}) = \tfrac{1}{m} \sum_{1\le k\le m} \left(\sum_{1\le h\le m} e^{\frac{2\pi}{m}hki} z_h \right)^d .$$

Then, the average type-spectrum takes the following form

$$\Gamma_{(\{1\},c,d)}(\boldsymbol{\theta}) = \mathrm{H}(\boldsymbol{\theta}) + \frac{c}{d} \inf_{\substack{\boldsymbol{z}\in\mathcal{P}(\mathbb{Z}_m)\\ \mathrm{supp}(\boldsymbol{z})=\mathrm{supp}(\boldsymbol{\theta})}} \left\{ \log\left(\tfrac{1}{m} \sum_k \left(\sum_h e^{\frac{2\pi}{m}hki} z_h \right)^d \right) + dD\big(\boldsymbol{\theta}||\boldsymbol{z}\big) \right\} . \tag{5.20}$$

The above spectrum coincides with the one obtained in [6].

Uniformly labelled ensembles over cyclic groups

Finally, consider the case when $G \simeq \mathbb{Z}_m$ again, but this time with F isomorphic to $\mathbb{Z}_m^*$, the multiplicative group of units of $\mathbb{Z}_m$. Notice that $\mathbb{Z}_m^*$ acts by multiplication on the ring $\mathbb{Z}_m$. It is immediate to see that two $a, b \in \mathbb{Z}_m$ are in the same orbit with respect to this group action, if and only if $(m, a) = (m, b)$, where (k, h) denotes the greatest common divisor of two naturals k and h. The quotient space $\mathbb{Z}_m/\mathbb{Z}_m^*$ can be identified with the set of divisors of m $\mathbb{D}_m := \{l \in \mathbb{N}\text{s.t. } l \mid m\}$. We have $|\mathbb{Z}_m^*| = \varphi(m)$, where $\varphi : \mathbb{N} \to \mathbb{N}$, $\varphi(n) = \big|\big\{m \in \mathbb{N}\text{s.t. } m \le n\,,\, (n,m) = 1\big\}\big|$, is the Euler φ-function. The projection map is

$$\pi_{\mathbb{Z}_m^*} : \mathbb{Z}_m \to \mathbb{D}_m\,, \qquad \pi_{\mathbb{Z}_m^*}(a) = \frac{m}{(m,a)}\,.$$

Notice that, for every $l \in \mathbb{D}_m$, the orbit $\pi^{-1}_{\mathbb{Z}^*_m}(l)$ coincides with $\frac{m}{l}\mathbb{Z}^*_m$ and it is in bijection with $\mathbb{Z}^*_l$ through the map $h \mapsto \frac{m}{l}h$. Then, $\boldsymbol{\varphi}(l) = |\pi^{-1}_{\mathbb{Z}^*_m}(l)| = |\mathbb{Z}^*_l| = \varphi(l)$.

In order to evaluate the average-type spectra of the (c,d)-regular $\mathbb{Z}^*_m$-labelled ensemble of LDPC $\mathbb{Z}_m$-codes, it is convenient to introduce the so-called Ramanujan sums

$$r_l(k) := \textstyle\sum_{j\in\mathbb{Z}^*_l} e^{\frac{2\pi}{l}jki}\,, \qquad l,k \in \mathbb{N}\,.$$

The Ramanujan sums are well-known in number theory and can be explicitly evaluated in terms of both the Euler φ-function and Möbius function

$$\mu : \mathbb{N} \to \mathbb{Z}\,, \qquad \mu(n) = \begin{cases} 1 & \text{if} \quad n = 1 \\ 0 & \text{if} \quad p^2 \mid n \text{ for some prime } p \\ (-1)^k & \text{if} \quad m = p_1 p_2 \dots p_k \text{ for distinct primes } p_i\,. \end{cases}$$

For every $l,k \in \mathbb{N}$ it holds [35, pag. 237]

$$r_l(k) = \mu\left(\frac{l}{(l,k)}\right)\frac{\varphi(l)}{\varphi\left(\frac{l}{(l,k)}\right)}\,. \tag{5.21}$$

We now can now explicitly evaluate the multinomial $\alpha_{\mathbb{Z}^*_m,d}(\boldsymbol{t})$, obtaining

$$\begin{aligned} \alpha_{\mathbb{Z}^*_m,d}(\boldsymbol{t}) &= \tfrac{1}{m}\sum_{1\le k\le m}\left(\sum_{l\mid m}\tfrac{1}{\varphi(l)}\sum_{j\in\mathbb{Z}^*_l} e^{\frac{2\pi}{l}jki}t_l\right)^d \\ &= \tfrac{1}{m}\sum_{1\le k\le m}\left(\sum_{l\mid m}\tfrac{1}{\varphi(l)}r_l(k)t_l\right)^d \\ &= \tfrac{1}{m}\sum_{k\mid m}\varphi\left(\tfrac{m}{k}\right)\left(\sum_{l\mid m}\frac{\mu\left(\frac{l}{(l,k)}\right)}{\varphi\left(\frac{l}{(l,k)}\right)}t_l\right)^d\,. \end{aligned}$$

It follows that the average type-spectrum of the (c,d)-regular $\mathbb{Z}^*_m$-labelled LDPC ensemble of $\mathbb{Z}_m$-codes is given by

$$\Gamma_{(\mathbb{Z}^*_m,c,d)}(\boldsymbol{\theta}) = \mathrm{H}(\boldsymbol{\theta}) + \frac{c}{d}\inf_{\boldsymbol{t}}\left\{\log\left(\tfrac{1}{m}\sum_{k\mid m}\varphi\left(\tfrac{m}{k}\right)\left(\sum_{l\mid m}\frac{\mu\left(\frac{l}{(l,k)}\right)}{\varphi\left(\frac{l}{(l,k)}\right)}t_l\right)^d\right) + dD\big(\pi_{\mathbb{Z}^*_m}\boldsymbol{\theta}||\boldsymbol{z}\big)\right\}\,, \tag{5.22}$$

where the above infimum has to be considered with respect to all $\boldsymbol{t}$ in $\mathcal{P}(\mathbb{D}_m)$ such that $\mathrm{supp}(\boldsymbol{t}) = \mathrm{supp}(\pi_{\mathbb{Z}^*_m}\boldsymbol{\theta})$. Of course, when m is prime, formula (5.22) above reduces to (5.18). In particular, when $m = 2$, (5.18), (5.20), and (5.22) coincide. For non-prime m instead, (5.22) is novel, to the best of our knowledge.

5.3 On low-weight type-spectra

In this section we will deal with estimations of the average type-spectra of the regular F-labelled LDPC G-code ensembles for G-types very close to the all-zero type δ_0. We will consider the variational distance on $\mathcal{P}(G)$, $||\boldsymbol{\theta} - \boldsymbol{\theta}'|| := \sup_{B \subseteq G}\{\boldsymbol{\theta}(B) - \boldsymbol{\theta}'(B)\}$.

Recall that, since we are dealing with LDPC G-codes, the all-zero N-tuple is always a codeword. Then, $W_N(\delta_0) = 1$ deterministically, i.e. for any realization of of Π_N in the interconnection group $S_{Nc} \ltimes F^{Nc}$. Hence clearly $\Gamma_{(F,c,d)}(\delta_0) = 0$. The main result of this section is that there exists a punctured neighborhood of δ_0 in $\mathcal{P}(G)$, over which the spectra $\Gamma_{(F,c,d)}(\boldsymbol{\theta})$ are strictly negative. Actually, much more precise results will be derived, characterizing the exact rate of decay (asymptotically in N) of the sum of the average enumerating coefficients over all G-types $\boldsymbol{\theta}$ such that $0 < ||\boldsymbol{\theta} - \delta_0|| < \frac{2}{d}$.

Throughout this section we will often use the following notation: for a, t in $\mathbb{N}$ we define the discrete intervals $I_t^a := [(t-1)a + 1, ta] \cap \mathbb{N}$. Notice that, given a degree pair (c, d), for every admissible blocklength N in $\mathcal{N}_{(c,d)}$ we have $\{1, 2, \ldots, Nc\} = \bigcup_{1 \leq t \leq L} I_t^d = \bigcup_{1 \leq s \leq N} I_s^c$.

5.3.1 An upper bound to low-weight spectra

We start by deriving an upper bound to low-weight type-enumerating coefficients for the inner encoder $\left|Z_{\boldsymbol{\theta}}^{i,N}\right| := \left|G_{\boldsymbol{\theta}}^{Nc} \cap \ker \mathrm{Sum}_d^N\right|$.

Lemma 54 *Let (c, d) be a degree pair, and let $N \in \mathcal{N}_{(c,d)}$. For every $\boldsymbol{\theta}$ in $\mathcal{P}_{Nc}(G)$ such that*

$$||\boldsymbol{\theta} - \delta_0|| \leq 1 - \frac{2}{d}\,, \tag{5.23}$$

we have

$$\left|Z_{\boldsymbol{\theta}}^{i,N}\right| \leq \binom{L}{\lfloor w/2 \rfloor} \binom{\lfloor w/2 \rfloor\, d}{w} \binom{w}{\boldsymbol{\omega}}\,, \tag{5.24}$$

where $\boldsymbol{\omega} \in \mathbb{N}^{G \setminus \{0\}}$ is defined by $\boldsymbol{\omega}(k) := Nc\,\boldsymbol{\theta}(k)$, and $w := \sum_{k=1}^{m-1} \boldsymbol{\omega}(k)$ is the number of non-zero entries in an Nc-tuple of type $\boldsymbol{\theta}$.

Proof Let $\boldsymbol{y}$ in $G_{\boldsymbol{\theta}}^{Nc}$ be any Nc-tuple of type $\boldsymbol{\theta}$. A necessary condition for $\boldsymbol{y}$ to be in $\ker \mathrm{Sum}_d^N$ is that each of the first L intervals I_t^d contains either none or at least two non-zero entries of $\boldsymbol{y}$. Clearly, $\left|\left\{t \leq L : \left|\mathrm{supp}(\boldsymbol{y}) \cap I_t^d\right| \geq 2\right\}\right| \leq \lfloor w/2 \rfloor$ while, for any choice of a dissection $1 \leq t_1 < \ldots < t_{\lfloor w/2 \rfloor} \leq L$, (notice that (5.23) implies $w/2 \leq L$)

we have $\left|\left\{\boldsymbol{y} \in G_{\boldsymbol{\theta}}^{Nc} : \operatorname{supp}(\boldsymbol{y}) \subseteq \bigcup_{j=1}^{\lfloor w/2 \rfloor} I_{t_j}^d\right\}\right| \leq \binom{d\lfloor w/2 \rfloor}{w}\binom{w}{\boldsymbol{\omega}}$. It follows that

$$\begin{aligned}
|Z_{\boldsymbol{\theta}}^{i,N}| &\leq \left|\bigcup_{1\leq t\leq L}\left\{\boldsymbol{y} \in G_{\boldsymbol{\theta}}^{Nc} : \left|\operatorname{supp}(\boldsymbol{y}) \cap I_t^d\right| \neq 1\right\}\right| \\
&\leq \left|\bigcup_{1\leq t_1<\ldots<t_{\lfloor w/2 \rfloor}\leq L}\left\{\boldsymbol{y} \in G_{\boldsymbol{\theta}}^{Nc} : \operatorname{supp}(\boldsymbol{y}) \subseteq \bigcup_{j=1}^{\lfloor w/2 \rfloor} I_{t_j}\right\}\right| \\
&\leq \binom{L}{\lfloor w/2 \rfloor}\binom{d\lfloor w/2 \rfloor}{w}\binom{w}{\boldsymbol{\omega}}.
\end{aligned}$$

We now obtain an estimation for the average low-weight type-enumerators.

Lemma 55 *Let (c,d) be a degree pair, $F \leq \operatorname{Aut}(G)$ and $N \in \mathcal{N}_{(c,d)}$. For every $\boldsymbol{\theta} \in \mathcal{P}_N(G)$ satisfying (5.23) the average type-enumerator function of the (c,d)-regular F-labelled ensemble satisfies*

$$\overline{W_N(\boldsymbol{\theta})} \leq \binom{N}{N\boldsymbol{\theta}}\binom{L}{\lfloor w/2 \rfloor}\left(\frac{w}{2L}\right)^w, \tag{5.25}$$

where $w := Nc(1-\boldsymbol{\theta}(0))$.

Proof Consider the projection map $\pi_F : G \to G/F$ and the associated map for types $\pi_F^\sharp : G \to G/F$. Define $\boldsymbol{\upsilon} := \pi_F^\sharp \boldsymbol{\theta}$, and $\boldsymbol{u} \in \mathbb{Z}_+^{G/F\setminus\{0\}}$ by $\boldsymbol{u}(k) = Nc\boldsymbol{\upsilon}(k)$. Also, for every $\boldsymbol{\theta}'$ in $\mathcal{P}(G)$, define $\boldsymbol{\omega}'$ in $\mathbb{Z}_+^{G\setminus\{0\}}$ by $\boldsymbol{\omega}'(k) := Nc\boldsymbol{\theta}'(k)$. Notice that $\sum_{\boldsymbol{\theta}' \in \mathcal{O}_{\boldsymbol{\upsilon}}^{Nc}}\binom{w}{\boldsymbol{\omega}'} = \binom{w}{Nc\boldsymbol{u}}\boldsymbol{\varphi}^{Nc\boldsymbol{\upsilon}}$. From (5.13), (5.16) and (5.24) we get

$$\begin{aligned}
\overline{W_N(\boldsymbol{\theta})} &= \binom{N}{N\boldsymbol{\theta}}\binom{Nc}{Nc\boldsymbol{\upsilon}}^{-1}\boldsymbol{\varphi}^{-Nc\boldsymbol{\upsilon}}\sum_{\boldsymbol{\theta}' \in \mathcal{O}_{\boldsymbol{\upsilon}}^{Nc}}|Z_{\boldsymbol{\theta}'}^{i,N}| \\
&\leq \binom{N}{N\boldsymbol{\theta}}\binom{Nc}{w}^{-1}\binom{L}{\lfloor w/2 \rfloor}\binom{\lfloor w/2 \rfloor d}{w}\binom{w}{Nc\boldsymbol{u}}^{-1}\boldsymbol{\varphi}^{-Nc\boldsymbol{\upsilon}}\sum_{\boldsymbol{\theta}' \in \mathcal{O}_{\boldsymbol{\upsilon}}^{Nc}}\binom{w}{\boldsymbol{\omega}'} \\
&= \binom{N}{N\boldsymbol{\theta}}\binom{L}{\lfloor w/2 \rfloor}\binom{Nc}{w}^{-1}\binom{\lfloor w/2 \rfloor d}{w} \\
&= \binom{N}{N\boldsymbol{\theta}}\binom{L}{\lfloor w/2 \rfloor}\frac{\lfloor w/2 \rfloor d(\lfloor w/2 \rfloor d-1)\ldots(\lfloor w/2 \rfloor d-w+1)}{Ld(Ld-1)\ldots(Ld-w+1)} \\
&\leq \binom{N}{N\boldsymbol{\theta}}\binom{L}{\lfloor w/2 \rfloor}\left(\frac{w}{2L}\right)^w.
\end{aligned}$$

A first consequence of Lemma 55 is the following upper bound on the average type-spectra of the F-labelled LDPC ensembles.

Proposition 56 *For every degree pair (c,d) such that $c \geq 3$ we have*

$$\Gamma_{(F,c,d)}(\boldsymbol{\theta}) \leq f_{c,d}(x)\ , \qquad \forall\, \boldsymbol{\theta}:\ ||\boldsymbol{\theta} - \delta_0|| \leq \frac{2}{d}\,,$$

where $x := 1 - \boldsymbol{\theta}(0)$, and

$$f_{c,d}(x) := \mathrm{H}(x) + x\log(|G|-1) + \tfrac{c}{d}\,\mathrm{H}\left(\tfrac{d}{2}x\right) + cx\log\left(\tfrac{d}{2}x\right)\ ,$$

with $\mathrm{H}(x) := -x\log x - (1-x)\log(1-x)$ denoting the binary entropy.

Proof From (5.25) it follows that for every $||\boldsymbol{\theta} - \delta_0|| < \frac{2}{d}$, for the F-labelled (c,d)-regular ensemble we have

$$\begin{aligned}
\frac{1}{N}\log\overline{W_N(\boldsymbol{\theta})} \quad &\leq \quad \frac{1}{N}\log\binom{N}{N\boldsymbol{\theta}} + \frac{1}{N}\log\binom{L}{\lfloor xN\frac{c}{2}\rfloor} + \frac{1}{N}\log\left(\frac{cNx}{2L}\right)^{cNx} \\
&\overset{N\in\mathcal{N}_{(c,d)}}{\longrightarrow} \ \mathrm{H}(\boldsymbol{\theta}) + \frac{c}{d}\,\mathrm{H}\left(\frac{d}{2}x\right) + +cx\log\left(\frac{d}{2}x\right) \\
&\leq \quad \mathrm{H}(x) + x\log(|G|-1) + cx\log\left(\frac{d}{2}x\right)\ .
\end{aligned}$$

It is easy to see that, whenever $c > 2$, $\frac{\mathrm{d}}{\mathrm{d}x} f_{c,d}\big|_{x=0} = -\infty$. Therefore, Proposition 56 implies that the spectra $\Gamma_{(F,c,d)}(\boldsymbol{\theta})$ are strictly negative in a sufficiently small punctured neighborhood of δ_0 in $\mathcal{P}(G)$. In Section 5.4 this fact will be used in order to show that the minimum $\boldsymbol{\Delta}$-distance grows linearly with N with high probability. Here we derive more precise estimations for the average type-enumerating functions.

Proposition 57 *Let F be any subgroup of $\mathrm{Aut}(G)$, (c,d) a degree pair and $N \in \mathcal{N}_{(c,d)}$. There exists a positive constant K such that the type-enumerator function of the (c,d)-regular F-labelled ensemble satisfies*

$$\sum_{\frac{2}{N}\leq||\delta_0-\boldsymbol{\theta}||\leq\frac{2}{d}} \overline{W_N(\boldsymbol{\theta})} \leq KN^{2-c}\,.$$

Proof For every N in $\mathcal{N}_{(c,d)}$ we define the quantities

$$g_w(N) := \sum_{||\delta_0-\boldsymbol{\theta}||=\frac{w}{N}} \overline{W_N(\boldsymbol{\theta})}\,, \qquad w \in \mathbb{N}\,.$$

For $\boldsymbol{\theta}$ in $\mathcal{P}_N(G)$ define $\boldsymbol{\omega}$ as in Lemma 54. For all $w = 2, \ldots, \lfloor \frac{2}{d}N \rfloor$, (5.25) implies

$$g_w(N) \leq \sum_{\boldsymbol{\theta}(0)=\frac{N-w}{N}} \binom{N}{N\boldsymbol{\theta}}\binom{L}{\lfloor c\frac{w}{2}\rfloor}\left(\frac{wc}{2L}\right)^{wc} = \binom{L}{\lfloor c\frac{w}{2}\rfloor}\left(\frac{wc}{2L}\right)^{wc}\binom{N}{w}(|G|-1)^w =: g'_w(N)\,.$$

We have, for every $2 \leq w \leq \lfloor 2dN \rfloor$,

$$\frac{g'_{w+2}(N)}{g'_w(N)} \leq (|G|-1)^2 \left(\frac{N-w}{w}\right)^2 \left(\frac{L - \lfloor c\frac{w}{2} \rfloor}{\lfloor c\frac{w}{2} \rfloor 2L}\right)^c \left(1+\frac{2}{w}\right)^{(w+2)c} \leq (|G|-1)^2 (3e)^{2c} N^{2-c}.$$

It follows that if $c \geq 3$, then there exists N_0 in $\mathbb{N}$ such that, for all N in $\mathcal{N}_{(c,d)}$ such that $N \geq N_0$, $\frac{g'_{w+2}(N)}{g'_w(N)} \leq \frac{1}{2}$ for all $1 \leq w \leq \lfloor \frac{2}{d} N \rfloor$. Then, we have

$$\sum_{\frac{2}{N} \leq ||\delta_0 - \boldsymbol{\theta}|| \leq \frac{2}{d}} \overline{W_N(\boldsymbol{\theta})} \leq g'_2(N) \sum_{w=2}^{\lfloor \frac{2}{d} N \rfloor} 2^{-w} + g'_3(N) \sum_{w=2}^{\lfloor \frac{2}{d} N \rfloor} 2^{-w} \leq 2g'_2(N) + 2g'_3(N) \leq K N^{2-c}$$

for some positive constants K', K'', K, all independent of N.

5.3.2 On weight-one codewords

We now derive a more precise estimation of the average enumerating functions for G-types of N-tuples with all but one entries equal to zero. Fixed any N in $\mathbb{N}$, k in G we define the G-type

$$\boldsymbol{\tau}_k := \left(1 - \frac{1}{N}\right) \delta_0 + \frac{1}{N} \delta_k \in \mathcal{P}_N(G),$$

and we look for upper bounds on the average spectra $\overline{W_N(\boldsymbol{\tau}_k)}$ for the (c,d)-regular F-labelled LDPC ensembles. We will show how these estimations depend on the choice of F among the subgroups of the automorphism group $\mathrm{Aut}(G)$.

We start with a few elementary considerations about closed paths and cycles in directed graphs. A closed path of length n in a directed graph $\mathcal{G} = (V, E)$ (where V is a finite set and $E \subseteq V^2$) is a $\mathbb{Z}_n$-labelled string of vertices $\boldsymbol{v} \in V^{\mathbb{Z}_n}$ such that any two consecutive vertices are adjacent, i.e. $(v_k, v_{k+1}) \in E$ for all $k \in \mathbb{Z}_n$. A cycle of length n is a closed path $\boldsymbol{v} \in V^{\mathbb{Z}_n}$ such that $v_k \neq v_j$ for all $k \neq j \in \mathbb{Z}_n$. A self-loop is a cycle of length 1. Every closed path $\boldsymbol{v}$ of length n is the concatenation of k cycles $\boldsymbol{v}^1, \ldots, \boldsymbol{v}^k$ such that the sum of the lengths of $\boldsymbol{v}^1, \ldots, \boldsymbol{v}^k$ equals n. Observe that in general $k \leq n$, while $k \leq \lfloor n/2 \rfloor$ provided that the directed graph $\mathcal{G}$ contains no self-loops.

Given a finite Abelian group G, and a subset S of G we denote by $\mathcal{G}(G, S)$ the directed Cayley graph with vertex set G and edge set $\{(g, g+s) | \, g \in G, \, s \in S\}$. It is straightforward that closed paths $\boldsymbol{v}$ of length n in an Abelian Cayley graph $\mathcal{G}(G, S)$ starting in any fixed vertex $g \in G$ (i.e. such that $v_0 = g$) are in one-to-one correspondence with 0-sum n-tuples $\boldsymbol{x}$ in S^n.

For a subset $S \subseteq G$ and a positive integer n, consider a closed path $\boldsymbol{v}$ of leght n in $\mathcal{G}$. By the the previous considerations, $\boldsymbol{v}$ is the concatenation of $k(\boldsymbol{v})$ cycles. We put

$b(S,n)$ equal to the maximum of $k(\boldsymbol{v})$ over all possible closed paths $\boldsymbol{v}$ of length n in $\mathcal{G}(G,S)$, with the agreement that $b(S,n)=0$ whenever no closed path in $\mathcal{G}(G,S)$ has length n. The reason for this notation becomes evident with the following result.

Lemma 58 *Let F be any subgroup of $\mathrm{Aut}(G)$, (c,d) a degree pair and $N \in \mathcal{N}_{(c,d)}$. Then, for all k in G, the type-enumerator function of the (c,d)-regular F-labelled ensemble satisfies*

$$\overline{W_N(\boldsymbol{\tau}_k)} \leq N \binom{L}{b(Fk,c)} \left[\frac{b(Fk,c)}{L}\right]^c . \tag{5.26}$$

Proof Define $\boldsymbol{v} := \pi_F^\sharp \boldsymbol{\tau}_k \in \mathcal{P}(G/F)$. Let $\boldsymbol{y}$ be any element of $G_{\boldsymbol{v}}^{Nc}$. Then, for $\mathrm{Sum}_d^N \boldsymbol{y} = \boldsymbol{0}$ in G^L it is necessary that $\sum_{1\leq j\leq Nc} y_j = 0$ in G. Since $\boldsymbol{y} \in G_{\boldsymbol{v}}^{Nc}$ has exactly c non-zero entries all belonging to Fk, it follows that $\left|Z_{\boldsymbol{v}}^{i,N}\right| = 0$ iff there are no closed paths of length c in the Cayley graph $\mathcal{G}(G,Fk)$. Then, (5.26) immediately follows in the case $b(Fk,c)=0$.

Now, assume that there exist closed paths of length c in $\mathcal{G}(G,Fk)$. By the previous considerations, each such a path decomposes in at most $b(Fk,c)$ cycles. If we consider the intervals I_t^d, for $1\leq t\leq L$, and put $\mathrm{supp}(\boldsymbol{y}) \cap I_t^d := \{j_1^t, j_2^t, \ldots, j_{n_t}^t\}$, we have

$$\left(\mathrm{Sum}_d^N \boldsymbol{y}\right)_t = \sum_{j\in I_t^d} y_j = \sum_{1\leq i\leq n_t} y_{j_i^t}\,, \qquad \forall\, 1\leq t\leq L\,.$$

Therefore, if $\mathrm{Sum}_d^N \boldsymbol{y} = \boldsymbol{0}$, then it is necessary that $\boldsymbol{v} \in G^{\mathbb{Z}_{n_t}}$, $v_l := \sum_{1\leq i\leq l} y_{j_i^t}$ is a closed path in $\mathcal{G}(G,Fk)$ for all t such that $\mathrm{supp}(\boldsymbol{y}) \cap I_t^d$ is non-empty. It follows that $\mathrm{supp}(\boldsymbol{y}) \cap I_t^d$ is non-empty for at most $b(Fk,c)$ values of t. Therefore, by taking into account the $\binom{L}{b(Fk,c)}$ possible choices of $b(Fk,c)$ intervals out of L possible ones, the $\binom{b(Fk,c)}{c}$ choices of c positions out of $b(Fk,c)d$ available ones, and the $\boldsymbol{\varphi}(Fk)^c$ choices of an ordered c-tuple with entries from the orbit Fk, we get

$$\left|Z_{\boldsymbol{v}}^{i,N}\right| = \left|\ker \mathrm{Sum}_d^N \cap G_{\boldsymbol{v}}^{Nc}\right| \leq \binom{L}{b(Fk,c)} \binom{b(Fk,c)d}{c} \boldsymbol{\varphi}(Fk)^c\,.$$

Then, from (5.13) it follows that

$$\overline{W_N(\boldsymbol{\tau}_k)} = \frac{N\left|Z_{\boldsymbol{v}}^{i,N}\right|}{\binom{Nc}{c}\boldsymbol{\varphi}(Fk)^c} \leq \frac{N}{\binom{Nc}{c}} \binom{L}{b(Fk,c)} \binom{b(Fk,c)d}{c} \leq N \binom{L}{b(Fk,c)} \left[\frac{b(Fk,c)}{L}\right]^c$$

5.3.3 Main result

Building on the results of Sect.5.3.1 and 5.3.2, we are now ready to present the main result of this section. For a subgroup F of $\mathrm{Aut}(G)$, and a positive integer c we define

$$a(F,c) := 1 - c + \max\left(\{1\} \cup \{b(Fk,c) |\, k \in G \setminus \{0\}\}\right) , \tag{5.27}$$

where we recall that $b(S,c)$ was defined in Sect.5.3.2 as the minimum number of cycles in $\mathcal{G}(G,S)$ of total length c, with the agreement that $b(S,c) = 0$ when no closed path in $\mathcal{G}(G,S)$ has length c.

Before stating the main result, we need a simple property of $a(F,c)$. For every $k \neq 0$, Fk does not contain 0, so that there are no self-loops in $\mathcal{G}(G,Fk)$ and then $b(Fk,c) \leq \lfloor c/2 \rfloor$. It immediately follows that

$$2 - c \leq a(F,c) \leq 1 - \lceil c/2 \rceil . \tag{5.28}$$

Theorem 59 *For every degree pair (c,d) such that $c \geq 3$, and every subgroup F of $\mathrm{Aut}(G)$, there exists a positive constant K such that for the (c,d)-regular F-labelled ensemble it holds*

$$\sum_{0<||\delta_0 - \boldsymbol{\theta}|| \leq \frac{2}{d}} \overline{W_N(\boldsymbol{\theta})} \leq K N^{a(F,c)} , \qquad N \in \mathcal{N}_{(c,d)} .$$

Proof First, we consider weight-one types. From (5.26) we have

$$\sum_{\boldsymbol{\theta}(0)=\frac{N-1}{N}} \overline{W_N(\boldsymbol{\theta})} \leq \sum_{k \in G\setminus\{0\}} N \binom{L}{b(Fk,c)} \frac{b(Fk,c)^c}{L^c} \leq K' \sum_{k \in G\setminus\{0\}} N^{1+b(Fk,c)-c} \leq K'|G|N^{a(F,c)}$$

for some positive constant K'. The claim then follows by combining Proposition 57 with the previous estimation, and observing that $a(F,c) \leq 2 - c \leq -1$.

Now, we explicitly evaluate $a(F,c)$ for the three examples studied in the previous section.

Example 14 *Consider the case when $G \simeq \mathbb{Z}_p^r$ and either $F \simeq \mathrm{Aut}(\mathbb{Z}_p^r)$ or $F \simeq \mathbb{F}_{p^r}^*$. In both cases $Fk = \mathbb{Z}_p^r \setminus \{0\}$ for all $k \in \mathbb{Z}_p^r \setminus \{0\}$. Then $\mathcal{G}(\mathbb{Z}_p^r, Fk) = \mathcal{G}(\mathbb{Z}_p^r, \mathbb{Z}_p^r \setminus \{0\})$ is the complete graph with p^r vertices. It follows that $\mathcal{G}(\mathbb{Z}_p^r, \mathbb{Z}_p^r \setminus \{0\})$ contains closed paths of any length $n \geq 2$ whenever $p^r \neq 2$, while $\mathcal{G}(\mathbb{Z}_2, \{1\})$ contains closed paths of even length only. Therefore, for $G \simeq \mathbb{Z}_p^r$ with $p^r \neq 2$, $a(F,c) = 1 - \lceil c/2 \rceil$ for all c, while for $G \simeq \mathbb{Z}_2$ $a(F,c) = 1 - c/2$ for even c and $2 - c$ for odd c.*

Example 15 *Consider the unlabelled ensemble over the cyclic group, i.e.* $G \simeq \mathbb{Z}_m$ *with* $F = \{1\}$. *If* $(m, c) = 1$, *then* $m|ck$ *if and only if* $m|k$. *Then, for all* $k \in \mathbb{Z}_m \setminus \{0\}$, *the Cayley graph* $\mathcal{G}(\mathbb{Z}_m, Fk) = \mathcal{G}(\mathbb{Z}_m, \{k\})$ *has no closed paths of length* c. *In this case clearly* $a(\{1\}, c) = 2 - c$.

Then, consider the case when $(m, c) > 1$ *and let* $\mathrm{lpcf}(c, m)$ *be the smallest prime common factor between* c *and* m. *Consider any* k *in* $\mathbb{Z}_m \setminus \{0\}$ *such that* $\mathcal{G}(\mathbb{Z}_m, \{k\})$ *has a closed path of length* c, *i.e. such that* $m \mid ck$. *The length of the shortest such a path is given by* $\frac{m}{(m,k)} = \frac{(m,ck)}{(m,k)} = \left(\frac{m}{(m,k)}, c\right)$. *Thus,* $\frac{m}{(m,k)} \mid c$, *while clearly* $\frac{m}{(m,k)} \mid m$. *But* $(m, k) < m$, *so that necessarily the shortest cycle in* $\mathcal{G}(\mathbb{Z}_m, \{k\})$ $\frac{m}{(m,k)}$ *is not less than* $\mathrm{lpcf}(m, c)$, *with equality if and only if* $k \in \frac{m}{\mathrm{lpcf}(m,c)}\mathbb{Z}_m \setminus \{0\}$. *Thus,* $b(\{k\}, c) = \frac{c}{\mathrm{lpcf}(m,c)}$ *for* $k \in \frac{m}{\mathrm{lpcf}(m,c)}\mathbb{Z}_m \setminus \{0\}$, *and* $b(\{k\}, c) < \frac{c}{\mathrm{lpcf}(m,c)}$ *for* $k \in \mathbb{Z}_m \setminus \frac{m}{\mathrm{lpcf}(m,c)}\mathbb{Z}_m$. *It immediately follows that, whenever* $(m, c) > 1$, $a(\{1\}, c) = 1 - c + \frac{c}{\mathrm{lpcf}(m,c)}$.

Example 16 *Consider the uniformly-labelled ensemble over the cyclic group, i.e.* $G \simeq \mathbb{Z}_m$ *with* $F \simeq \mathbb{Z}_m^*$. *First we claim that, for* $n \geq 2$*:*

- *if* n *is even all closed paths in* $\mathcal{G}(\mathbb{Z}_n, \mathbb{Z}_n^*)$ *have even length and there exists a 2-cycle;*
- *if* n *is odd, then there exist both a 2-cycle and a 3-cycle.*

To see this, first, since $1, -1 \in \mathbb{Z}_n^*$, $(0, 1)$ *is a 2-cycle in* $\mathcal{G}(\mathbb{Z}_n, \mathbb{Z}_n^*)$, *both for even and odd* n. *Then, consider the case when* n *is even: clearly all* $k \in \mathbb{Z}_n^*$ *are odd, so that the modulo-*n *sum of an odd number of elements of* $\mathbb{Z}_n^*$ *cannot be equal to 0 modulo* n. *Thus every closed path in* $\mathcal{G}(\mathbb{Z}_n, \mathbb{Z}_n^*)$ *must be of even length. On the other hand, if* n *is odd, then* $2 \in \mathbb{Z}_n^*$, *so that* $(0, 2, 1)$ *is a* 3*-cycle in* $\mathcal{G}(\mathbb{Z}_n, \mathbb{Z}_n^*)$.

Let us now consider some $k \in \mathbb{Z}_m \setminus \{0\}$. *Then, by applying the previous observation with* $n = \frac{m}{(m,k)}$, *one gets that, if* c *is odd and* $\frac{m}{(m,k)}$ *is even, there are no closed paths of length* c *in* $\mathcal{G}(\mathbb{Z}_m, \mathbb{Z}_m^* k)$ *so that* $b(\mathbb{Z}_m^* k, c) = 0$, *while otherwise, if* c *is even or* $\frac{m}{(m,k)}$ *is odd,* $b(\mathbb{Z}_m^* k, c) = \lfloor c/2 \rfloor$. *It thus follows that* $a(\mathbb{Z}_m^*, c) = 1 - \lceil c/2 \rceil$ *unless* c *is odd and* m *is an integer power of* 2*; in the latter case* $a(\mathbb{Z}_m^*, c) = 2 - c$.

5.3.4 Lower bounds on low-weight type-enumertors

In this section we present some results, of independent interest, which show that the estimations given by Theorem 59 are tight. All the proofs are deferred to the appendix.

First we deal with weight-one type-enumerators.

Proposition 60 *Let* (c, d) *be a degree pair such that* $c \geq 3$, *and let* F *be any subgroup of* $\mathrm{Aut}(G)$. *Then, there exists a constant* $K > 0$ *such that for all* k *in* $G \setminus \{0\}$ *such that*

$a(F,c) = 1 - c + b(Fk,c)$ *the type-enumerator function of the* (c,d)*-regular* F*-labelled LDPC ensemble satisfies*

$$\mathbb{P}\left(W_N(\boldsymbol{\tau}_k) \geq 1\right) \geq K N^{a(F,c)}\,. \qquad N \in \mathcal{N}_{(c,d)}\,. \tag{5.29}$$

Finally, we propose a lower bound on weight-two type-enumerators. For every k in G, define

$$\hat{\boldsymbol{\tau}}_k := \frac{1}{N}\delta_k + \frac{1}{N}\delta_{-k} + \frac{N-2}{N}\delta_0 \in \mathcal{P}(G)\,.$$

Proposition 61 *For every degree pair* (c,d) *there exists a constant* $K > 0$ *such that for every* k *in* $G \setminus \{0\}$ *the type-enumerator function of the* (c,d)*-regular* F*-labelled LDPC ensemble satisfies*

$$\mathbb{P}\left(W_N(\hat{\boldsymbol{\tau}}_k) \geq 1\right) \geq K N^{2-c}\,, \qquad \forall\, N \in \mathcal{N}_{(c,d)}\,. \tag{5.30}$$

5.4 Asymptotic lower bounds on the typical minimum distance

Throughout this section we will assume we have fixed a G-symmetric MC $(\mathcal{X}, \mathcal{Y}, P)$ with associated Bhattacharyya distance $\boldsymbol{\Delta}$ and weight $\boldsymbol{\delta}$, and we study the asymptotics of the minimum $\boldsymbol{\Delta}$-distance of regular LDPC G-code ensembles.

Given a degree pair (c,d), a natural candidate for the typical normalized minimum $\boldsymbol{\Delta}$-distance of the (c,d)-regular F-labelled ensemble is the quantity

$$\gamma_{(F,c,d)} := \inf\left\{ \langle \boldsymbol{\delta}, \boldsymbol{\theta} \rangle \,\middle|\, \boldsymbol{\theta} \in \mathcal{P}(G) \setminus \{\delta_0\} \text{s.t. } \Gamma_{(F,c,d)}(\boldsymbol{\theta}) \geq 0 \right\}\,. \tag{5.31}$$

It turns out that $\gamma_{(F,c,d)}$ actually is a lower bound on the asymptotic normalized minimum distance for the (c,d)-regular F-labelled ensemble. This does not follow directly from Theorem 53 since $\lim_{\boldsymbol{\theta} \to \delta_0} \Gamma_{(F,c,d)}(\boldsymbol{\theta}) = 0$. However, using both Theorem 53 and 59 the following result.

Theorem 62 *Let* (c,d) *be a degree pair such that* $a(F,c) < -1$*. Then, for the* (c,d)*-regular* F*-labelled LDPC ensemble the following holds*

$$\mathbb{P}\left(\liminf_{N \in \mathcal{N}_{(c,d)}} \tfrac{1}{N} \mathrm{d}_{\min}\left(\ker \Phi_N\right) \geq \gamma_{(c,d)} \right) = 1\,.$$

Proof By (2.10) we have that

$$\frac{1}{N}\,\mathrm{d}_{\min}\left(\ker\Phi_N\right)=\inf\left\{\langle\boldsymbol{\delta},\boldsymbol{\theta}\rangle\,\middle|\,\boldsymbol{\theta}\in\mathcal{P}(G)\setminus\{\delta_0\}\text{s.t. }W_N(\boldsymbol{\theta})\geq 1\right\}=\min\left\{\kappa'_N,\kappa''_N\right\},$$

where for every N in $\mathcal{N}_{(c,d)}$ we define

$$\begin{aligned}\kappa'_N&:=\inf\left\{\langle\boldsymbol{\delta},\boldsymbol{\theta}\rangle\big|\,0<||\boldsymbol{\theta}-\delta_0||<\tfrac{2}{d}\,:\,W_N(\boldsymbol{\theta})\geq 1\right\}\,,\\ \kappa''_N&:=\inf\left\{\langle\boldsymbol{\delta},\boldsymbol{\theta}\rangle\big|\,||\boldsymbol{\theta}-\delta_0||\geq\tfrac{2}{d}\,:\,W_N(\boldsymbol{\theta})\geq 1\right\}\,.\end{aligned}$$

Clearly $\liminf_N\frac{1}{N}\,\mathrm{d}_{\min}\left(\ker\Phi_N\right)=\min\{\rho',\rho''\}$, where we put $\rho':=\liminf_N\kappa'_N$ and $\rho'':=\liminf_N\kappa''_N$.

We start by establishing a lower bound on ρ''. Define $\Omega:=\left\{\boldsymbol{\theta}:\,||\boldsymbol{\theta}-\delta_0||\geq\frac{2}{d}\right\}$, and, for each x in $\mathbb{R}$, the set

$$\Omega_x:=\left\{\boldsymbol{\theta}\in\Omega\cap\mathcal{P}_{\mathbb{N}}(G)\text{s.t. }\Gamma_{(F,c,d)}(\boldsymbol{\theta})<x\right\}\,. \tag{5.32}$$

Consider now the quantity $\eta(x):=\inf\left\{\langle\boldsymbol{\delta},\boldsymbol{\theta}\rangle\big|\,\boldsymbol{\theta}\in\Omega\setminus\Omega_x\right\}$. Since $\Gamma_{(F,c,d)}(\boldsymbol{\theta})$ is an upper semicontinuous function of $\boldsymbol{\theta}$ and Ω is a closed subset of $\mathcal{P}(G)$, standard analytical arguments (see Lemma 64 in the appendix) allow us to conclude that η is a nondecreasing and lower semicontinuous function.

Let us now fix some arbitrary $\varepsilon>0$. By successively applying a union bound estimation, Markov inequality, Theorem 53 and (5.32), we get

$$\mathbb{P}\left(\bigcup_{\boldsymbol{\theta}\in\Omega_{-\varepsilon}}\{W_N(\boldsymbol{\theta})\geq 1\}\right)\leq\sum_{\boldsymbol{\theta}\in\Omega_{-\varepsilon}}\mathbb{P}\left(W_N(\boldsymbol{\theta})\geq 1\right)\leq\sum_{\boldsymbol{\theta}\in\Omega_{-\varepsilon}}\overline{W_N(\boldsymbol{\theta})}\leq\tbinom{N+|G|-1}{|G|-1}\exp(-N\varepsilon).$$

It follows that $\sum_N\mathbb{P}\big(\bigcup_{\boldsymbol{\theta}\in\Omega_{-\varepsilon}}\{W_N(\boldsymbol{\theta})\geq 1\}\big)<+\infty$, and thus Borel-Cantelli lemma implies that with probability one the event $\bigcup_{\boldsymbol{\theta}\in\Omega_{-\varepsilon}}\{W_N(\boldsymbol{\theta})\geq 1\}$ occurs only for finitely many N in $\mathcal{N}_{(c,d)}$. Hence,

$$\mathbb{P}\left(\rho''<\eta(-\varepsilon)\right)\leq\mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}\in\Omega_{-\varepsilon}}\{W_N(\boldsymbol{\theta})>0\}\right\}\text{ i.o. }\;N\in\mathcal{N}_{(c,d)}\right)=0\,,\qquad\forall\,\varepsilon>0\,.$$

Notice that $\gamma_{(F,c,d)}=\eta(0)$. Hence, monotonicity and lower semicontinuity of the function η allow us to conclude that

$$\mathbb{P}\left(\rho''<\gamma_{(F,c,d)}\right)=\mathbb{P}\left(\rho''<\eta(0)\right)\leq\mathbb{P}\left(\rho''<\lim_k\eta\left(-\tfrac{1}{k}\right)\right)=\lim_k\mathbb{P}\left(\rho''<\eta\left(-\tfrac{1}{k}\right)\right)=0\,. \tag{5.33}$$

Now let us consider the term ρ'. By sequentially applying a union bound estimation, Markov inequality and Theorem 59, we get for every N in $\mathcal{N}_{(c,d)}$

$$\mathbb{P}\left(\bigcup_{0<||\boldsymbol{\theta}-\delta_0||<\frac{2}{d}}\{W_N(\boldsymbol{\theta})\geq 1\}\right)\leq\sum_{0<||\boldsymbol{\theta}-\delta_0||<\frac{2}{d}}\overline{W_N(\boldsymbol{\theta})}\leq KN^{a(F,c)}\,, \tag{5.34}$$

where K is a positive constant independent of N. Since $a(F,c) < -1$, we get

$$\sum_N \mathbb{P}\left(\bigcup_{0<||\boldsymbol{\theta}-\delta_0||<\frac{2}{d}} \{W_N(\boldsymbol{\theta}) \geq 1\}\right) \leq K \sum_N N^{a(F,c)} < +\infty\,.$$

By Borel-Cantelli lemma we get that the event $\bigcup_{0<||\boldsymbol{\theta}-\delta_0||<\frac{2}{d}} \{W_N(\boldsymbol{\theta}) \geq 1\}$ occurs only for finitely many N in $\mathcal{N}_{(c,d)}$ with probability one. This yields $\mathbb{P}(\rho' = +\infty) = 1$, which, together with (5.33), implies the claim.

We have proved the previous theorem under the assumption $a(F,c) < -1$. In fact, for $c = 2$ it is known, since Gallager's work [30], that deterministically the minimum distance cannot grow faster than logarithmically with the blocklength N. From (5.28) it follows that if $c \geq 5$ then $a(F,c) < -1$ for any F, if $c = 3$ then $a(F,c) = -1$ for any F, while, when $c = 4$, $a(F,c) < -1$ for some choices of F. However, one can weaken the assumption $a(F,c) < -1$ requiring only that $a(F,c) < 0$ (thus including the cases $c = 3$, and $c = 4$ for some F). In these cases, $\gamma_{(F,c,d)}$ still gives an asymptotic lower bound for the normalized minimum distances $\frac{1}{N}\,\mathrm{d}_{\min}(\ker \Phi_N)$, in a weaker probabilistic sense. In fact, a more detailed analysis enlightens a non-concentration phenomenon. In order to describe it, first, for every degree pair (c,d) and every subgroup F of $\mathrm{Aut}(G)$, we define the following quantity

$$\zeta_{(F,c)} := \begin{cases} \min\{\boldsymbol{\delta}(k) |\, k \in G \setminus \{0\} : a(F,c) = 1 - c + b(Fk,c)\} & \text{if} \quad a(F,c) \neq 2 - c \\ \min\{(2 - b(Fk,c))\boldsymbol{\delta}(k) |\, k \in G \setminus \{0\}\} & \text{if} \quad a(F,c) = 2 - c\,. \end{cases} \tag{5.35}$$

We have the following result:

Theorem 63 *Let (c,d) be a degree pair such that $a(F,c) = -1$. Then,*

$$\lim_{N \in \mathcal{N}_{(c,d)}} \mathbb{P}\left(\frac{1}{N}\,\mathrm{d}_{\min}(\ker \Phi_N) \geq \gamma_{(F,c,d)}\right) = 1\,.$$

Moreover, if the random variables Π_N defining the (c,d)-regular unlabelled LDPC ensemble are mutually independent, we have

$$\mathbb{P}\left(\liminf_{N \in \mathcal{N}_{(c,d)}} \mathrm{d}_{\min}(\ker \Phi_N) = \zeta_{(F,c)}\right) = 1\,.$$

Theorem 63 is proven in the appendix. The probabilistic interpretation is as follows. In the case $a(F,c) = -1$, with probability one, the sequence of the unnormalized minimum distances $(\mathrm{d}_{\min}(\ker \Phi_N))$ contains a subsequence converging to $\zeta_{(F,c)}$. Thus, while with increasing probability the minimum $\boldsymbol{\Delta}$-distance is growing linearly with the blocklength N, almost surely a subsequence with constant minimum distance shows up. We observe that, for irregular binary LDPC ensembles, even more evident non-concentration phenomena are known to arise: see [16, 51].

5.5 Numerical results

In this section we present some numerical results for the minimum distances of the LDPC ensembles which have been studied in this paper. We focus on a particular channel, the $\mathbb{Z}_8$-symmetric 8-PSK AWGN channel, and we compare the average distance-spectra of the regular unlabelled and uniformly labelled LDPC $\mathbb{Z}_8$-code ensembles. Our results indicate a strong superiority of the uniformly labelled (i.e. the one with label group $F \simeq \mathbb{Z}_8^*$) ensemble with respect to the unlabelled one (i.e. $F = \{1\}$). Then, we compare these results with some contradicting analysis of the average error probability of these ensembles, and discuss how this seeming paradox can be explained by invoking so-called expurgation arguments.

5.5.1 Numerical results for the average distance-spectra

Let us start with some general considerations. Suppose we are given any ensemble of G-codes with average type-spectrum $\Gamma(\boldsymbol{\theta})$. Let $\gamma := \inf\{\langle \boldsymbol{\theta}, \boldsymbol{\delta}\rangle | \, \boldsymbol{\theta} \in \mathcal{P}(G) \setminus \{\delta_0\} \text{s.t. } \Gamma(\boldsymbol{\theta}) \geq 0\}$ be its designated typical normalized minimum distance which we are interested in computing. Notice that Γ is a map defined over the $(|G|-1)$-dimensional simplex $P(G)$ and therefore in general of difficult visualization whenever $|G| > 2$. It is then convenient and natural to introduce the average distance-spectra as a one dimensional projection of Γ:

$$\Upsilon : [0, \max\{\boldsymbol{\delta}(x) | \, x \in G\}] \to [-\infty, +\infty)\,, \quad \Upsilon(t) := \sup\left\{\Gamma(\boldsymbol{\theta}) \middle| \, \boldsymbol{\theta} \in \mathcal{P}(G) : \, \langle \boldsymbol{\delta}, \boldsymbol{\theta}\rangle = t\right\}. \tag{5.36}$$

It is immediate to verify that $\gamma = \inf\{t \in [0, \max\{\boldsymbol{\delta}(x) | \, x \in G\}] : \Upsilon(t) \geq 0\}$. Notice also that, for $|G| = 2$ and $|G| = 3$, all Bhattacharyya distances are proportional to the Hamming distance, so that the average distance spectrum Υ is independent (up to a rescaling factor) of the chosen G-symmetric channel. For $|G| \geq 4$ instead, Υ really depends on the choice of the Bhattacharyya distance $\boldsymbol{\Delta}$.

In Fig.1 the average distance-spectra of two regular LDPC $\mathbb{Z}_8$-code ensembles are reported. We considered the Bhattacharyya distance $\boldsymbol{\Delta}$ of the 8-PSK AWGN channel, normalized it in such a way that $\max\{\boldsymbol{\delta}(x) | \, x \in \mathbb{Z}_8\} = \boldsymbol{\Delta}(0, 4) = 1$. In each picture a degree pair (c, d) is fixed. The dash-dotted curve is the graph of the distance-spectrum $\Upsilon_{(\{1\},c,d)}(t)$ of the (c, d)-regular unlabelled LDPC ensemble, while the solid curve is the graph of the distance-spectrum $\Upsilon_{(\mathbb{Z}_8^*,c,d)}(t)$ of the (c, d)-regular uniformly labelled LDPC ensemble.

As a reference two dotted curves are also plotted in each picture. The one taking the value 0 for $t = 0$ is the distance spectrum of the binary (c, d)-regular LDPC ensemble $\Upsilon^2_{(c,d)}(t)$. It is straightforward to check that it is a lower bound for the distance spectrum of any $\mathbb{Z}_8$-LDPC ensemble: it suffices to restrict the optimization in (5.36) to $\mathbb{Z}_8$-types $\boldsymbol{\theta}$ supported on the binary subgroup $4\mathbb{Z}_8$.

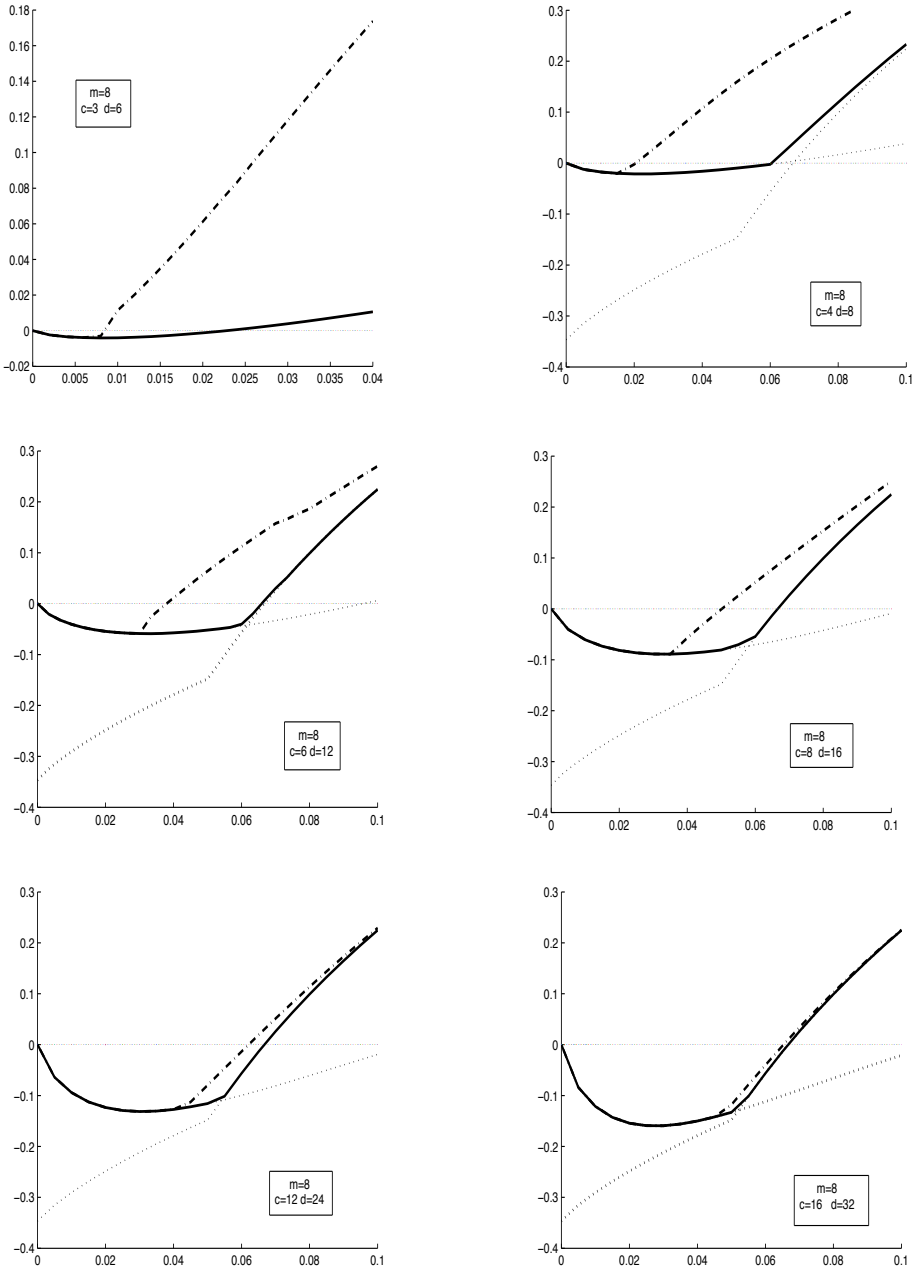

Figure 5.1: Distance spectra of (c, d)-regular LDPC ensembles over $\mathbb{Z}_8$ for 8-PSK: the solid curve corresponds to the uniformly-labelled ensemble, the dash-dotted one to the unlabelled ensemble, the two dotted curves correspond respectively to the $\mathbb{Z}_8$-linear ensemble and to the binary LDPC ensemble.

The second dotted curve instead, taking value $\frac{1}{2}\log\frac{1}{2}$ for $t = 0$, corresponds to the distance-spectra of the $\mathbb{Z}_8$-code ensemble (with no sparsity constraints) of the same rate $R = \frac{1}{2}\log 8$. This ensemble is defined as a sequence of kernels of random homomorphisms $(\ker \Phi_N)$, each Φ_N being uniformly distributed over $\mathrm{Hom}(\mathbb{Z}_8^N, \mathbb{Z}_8^{N/2})$, the group of all homomorphisms from $\mathbb{Z}_8^N$ to $\mathbb{Z}_8^{N/2}$, with no sparsity constraint. In Chapter 4 their average type-spectra have been characterized; for the $\mathbb{Z}_8$-code ensemble of rate $\frac{1}{2}\log 8$ this is given by

$$\Gamma_{\mathbb{Z}_8}(\boldsymbol{\theta}) := \mathrm{H}(\boldsymbol{\theta}) - \frac{1}{2}\log l_8(\boldsymbol{\theta})\,, \qquad l_8(\boldsymbol{\theta}) := \frac{8}{\gcd\left(\mathrm{supp}\left(\boldsymbol{\theta}\right)\right)}\,.$$

Notice that $\Gamma_{\mathbb{Z}_8}(\boldsymbol{\theta})$ is an upper semicontinuous function over the simplex $\mathcal{P}(\mathbb{Z}_8)$, and its discontinuities correspond to types supported on the subgroups $2\mathbb{Z}_8$ and $4\mathbb{Z}_8$. In fact a salient point is easily recognizable in the graphs reported around the abscissa $t = 0.05$, corresponding to the intersection between the average spectrum of the binary subchannel and that of the $\mathbb{Z}_8$-subchannel. This salient point occurs before the curve crosses the t-axis, which is coherent with the fact, proved in Chapter 4, that the typical normalized minimum distance of the $\mathbb{Z}_8$-code ensemble equals the corresponding Gilbert-Varshamov bound. In other words, while for low values of t the distance spectrum of the $\mathbb{Z}_8$-code ensemble is dominated by the term corresponding to the smallest non-trivial subgroup (a phenomenon generally observable for Abelian group code ensembles), the value of the typical minimum distance is determined by types which are not supported in any proper subgroup of $\mathbb{Z}_8$ (this is instead related to the particular constellation chosen, although conjectured to be true for many constellation of interest).

Analogous considerations can be made about the LDPC distance-spectra based on the simulations reported. In particular, for distances close to 0, the average distance-spectra of both the unlabelled and the uniformly labelled $\mathbb{Z}_8$-LDPC ensembles are dominated by the binary subgroup supported types. However, these components do affect the value of the typical normalized minimum distances ($\gamma_{(\{1\},c,d)}$ and $\gamma_{(\mathbb{Z}_8^*,c,d)}$ respectively) only for low values of the degrees ($c = 3, 4$). For all the other values of the parameters, the typical minimum distance is instead determined by types which are not supported in any proper subgroup of $\mathbb{Z}_8$. Another observation which can be made is that, not surprisingly, as the values of the degrees (c, d) are increased while keeping their ratio constant, the distance-spectra of both the unlabelled and the uniformly labelled ensembles approach the one of the $\mathbb{Z}_8$-linear ensemble.

However, the most important conclusion which can be drawn from the graphics reported concerns the different behaviors of the unlabelled and the uniformly labelled ensembles. Indeed, it appears evident that the latter drastically outperforms the former at the distance level. In particular, already for relatively low values of the degrees ($c = 8$, $d = 16$) the uniformly labelled ensemble typical minimum distance $\gamma_{(\mathbb{Z}_8^*,c,d)}$ is very close

(practically equal) to the Gilbert-Varshamov bound. For the same values of the degrees instead, the unlabelled ensemble suffers from a remarkable gap; this gap seems to be slowly filled up as the values of the degrees are increased, but it still remains significant for relatively high values of c and d. This indicates that structural properties of these two ensembles are remarkably different. Some prudence is nevertheless justified by the fact that ours are only lower bounds on the typical asymptotic normalized minimum distance, while, as already mentioned in the introduction, a concentration result for the type-spectra is needed in order to prove their tightness. However, while this phenomenon appears here only at the distance level, computer simulations of the performance of these codes reveal that a drastic superiority of the labelled ensemble with respect to the unlabelled one is evident also under belief-propagation decoding. We observe that this is coherent with Monte-Carlo simulations reported in [6], where the labelled ensemble was shown to be closer to capacity than the unlabelled ensemble.

5.5.2 The average word error probability of the LDPC codes ensembles

In our analysis of the minimum distance properties of LDPC G-code ensembles, the quantities $\zeta_{(F,c)}$ show up as an almost sure $\liminf$ for the unnormalized minimum distance only when $a(F,c) = -1$. However, these quantities characterize the asymptotic maximum-likelihood average performance of these ensembles for all values of $a(F,c)$.

For instance, let us consider in some detail the case $G \simeq \mathbb{Z}_{p^r}$ for some prime p and some positive integer r. Let us fix an admissible degree pair (c,d) and denote by $\overline{p_e(\mathcal{C}_N)}^{(F,c,d)}$ the average maximum-likelihood error probability of the (c,d)-regular F-labelled ensemble of LDPC $\mathbb{Z}_{p^r}$-codes over an arbitrary $\mathbb{Z}_{p^r}$-symmetric memoryless channel. Then, it is possible to show that there exist a threshold $(1-\frac{c}{d})\log p^r < C_{(F,c,d)} < \log p^r)$ such that, for every $\mathbb{Z}_{p^r}$-symmetric channel whose $\mathbb{Z}_{p^r}$-capacity exceeds to $C_{(F,c,d)}$, the average error probability $\overline{p_e(\mathcal{C}_N)}^{(F,c,d)}$ goes to zero in the limits of large N. Moreover, if one considers an increasing sequence of degree pairs (c_n,d_n) with a given designed rate $(1-\frac{c_n}{d_n})\log p^r$ converging to R, then the corresponding LDPC thresholds $C_{(c_n,d_n,F)}$ converge to R.

More precisely, it is possible to show that over any $\mathbb{Z}_{p^r}$-symmetric channel whose $\mathbb{Z}_{p^r}$-capacity exceeds $C_{(F,c,d)}$ we have

$$K_1 N^{a(F,c)} \leq \overline{p_e(\mathcal{C}_N)}^{(F,c,d)} \leq K_2 N^{a(F,c)}\,, \tag{5.37}$$

for some positive constants K_1, K_2, both independent of N. Moreover, it can be proven that

$$\limsup_{N\in\mathcal{N}_{(c,d)}} \frac{\overline{p_e(\mathcal{C}_N)}^{(F,c,d)}}{N^{a(F,c)}} \leq K_3 \exp(\zeta_{(F,c)})\,, \tag{5.38}$$

for some positive constants K_3, independent of the channel (and thus from $\mathbf{\Delta}$). The results (5.37) are known in the binary case (see [50]). Proofs of (5.37), (5.38) in their full generality can be gathered coupling the estimations of Section 5.3 with the standard bounding techniques used in [48, 62, 50, 6], and will be given elsewhere.

Observe that if $F \leq F' \leq \mathrm{Aut}(G)$

$$a(F,c) \leq a(F',c)\,, \qquad \zeta_{(F,c)} \geq \zeta_{(F',c)}\,. \tag{5.39}$$

Thus, from the point of view of the average performance, the smaller the label group is, the better parameters are. This stands in contrast with the numerical results presented in the previous paragraph, indicating that at the distance level the uniformly labelled ensembles perform much better than their unlabelled counterparts. An explanation for this seeming paradox can be obtained using so-called expurgation techniques. Indeed, it can be proved that, while the average error probability of the LDPC ensembles is affected by a vanishingly small fraction of codes with low minimum distance and decays to zero only as a negative power of N, almost surely a sequence of codes sampled from the same ensemble has error probability decreasing to zero exponentially fast with N. It is this typical exponential behavior that has to be considered representative of the ensemble, rather than the one of the average error probability. It is also worth to mention that the typical error exponent can be estimated in terms of the average type-spectra, using techniques presented in [62]. This phenomenon is well-known in the LDPC codes literature [30, 50]; proofs for LDPC codes over Galois fields can be found in [19, 6].

5.6 Conclusions

The following issues are left for future research:

- proving concentration results for the spectra of the LDPC ensembles for instance using a second order method (see [54]);
- giving an analytical explanation of the different behavior of the labelled and unlabelled ensembles;
- generalizing the analysis to irregular ensembles following the approach of [16, 51];
- considering generalizations of the so-called stopping sets and pseudoweight distributions which in the binary case characterize the iterative decoding performance of LDPC codes (see [51, 72, 40]); while the distribution of stopping sets has been studied for binary LDPC ensembles, the distribution of pseudocodewords is unknown even in the binary case.

Chapter 6

Conclusions

In this thesis we have developed a theoretical analysis of the performance of Abelian group codes over symmetric channels.

We have characterized the capacity achievable by Abelian group codes over symmetric channels. We have shown that in many important cases, like the AWGN channel with m-PSK modulation as input, this capacity coincides with the Shannon capacity. This generalizes a well-known result of classical information theory, namely that binary-linear codes allow to achieve capacity of binary-input output-symmetric channels.

For the AWGN channel with 8-PSK as input we have shown that the typical cyclic group code asymptotically meets the Gilbert-Varshamov bound, while a random binary-affine code is bounded away from it with probability one. The results obtained can be extended to Abelian group codes over symmetric channels, and similar results can be inferred for the typical error exponent.

We have analyzed two ensembles of regular LDPC codes over the cyclic group $\mathbb{Z}_m$, establishing precise combinatorial results for the exponential growth rate of their type-enumerating functions with respect to the code-length. We have shown that in both cases minimum distances grow linearly with probability one, and we have obtained lower bounds for their typical normalized minimum distance.

Some of the main problems left for future research are:

- extending the theory to non-Abelian group codes;
- proving concentration results for the spectra of the LDPC ensembles using a second order method;
- analyzing LDPC codes over Abelian groups under iterative decoding, generalizing some of the tools used for binary LDPC codes: density evolution, stopping sets and pseudo-weight distribution.

Chapter 7

Appendix

7.1 A few properties of the discrete entropy function

Consider the set $\mathcal{P}(A)$ of probability measures over A, which can be identified with the set of nonnegative real valued functions $\boldsymbol{\mu}$ on A satisfying the linear constraint $\sum_{a\in A}\boldsymbol{\mu}(a)=1$. If $\boldsymbol{\theta}$ is in $\mathcal{P}(A)$ and $B\subseteq A$ is such that $\boldsymbol{\theta}(B):=\sum_{b\in B}\boldsymbol{\theta}(b)>0$, the conditioned measure of $\boldsymbol{\theta}$ on B is defined as

$$\boldsymbol{\theta}|_B\in\mathcal{P}(B)\qquad \boldsymbol{\theta}|_B(b):=\boldsymbol{\theta}(B)^{-1}\boldsymbol{\theta}(b)\,,\qquad \forall\, b\in B\,. \tag{7.1}$$

Recall the definition of the entropy $\mathrm{H}(\boldsymbol{\theta}):=-\sum_{a\in\mathrm{supp}(\boldsymbol{\theta})}\boldsymbol{\theta}(a)\log\boldsymbol{\theta}(a)$. A straightforward property of the entropy function is its strict concavity.

To any function $\pi:A\to B$ between two nonempty finite sets A and B we can associate a map $\pi_\#:\mathcal{P}(A)\to\mathcal{P}(B)$ sending the probability measure $\boldsymbol{\theta}$ in $\mathcal{P}(A)$ to its image measure through π defined by

$$\pi_\#\boldsymbol{\theta}\in\mathcal{P}(B)\qquad [\pi_\#\boldsymbol{\theta}]\,(b):=\boldsymbol{\theta}\left(f^{-1}(b)\right)=\sum_{a:\,f(a)=b}\boldsymbol{\theta}(a)\,,\qquad \forall\, b\in B\,. \tag{7.2}$$

The entropy of a measure $\boldsymbol{\theta}$ and that of its image measure $\pi_\#\boldsymbol{\theta}$ are related by the following equality

$$\mathrm{H}\,(\boldsymbol{\theta})=\mathrm{H}\,(\pi_\#\boldsymbol{\theta})+\sum_{b\in\mathrm{supp}\left(\pi_\#\boldsymbol{\theta}\right)}[\pi_\#\boldsymbol{\theta}]\,(b)\,\mathrm{H}\left(\boldsymbol{\theta}|_{f^{-1}(b)}\right)\,. \tag{7.3}$$

It follows from (7.3) that $\mathrm{H}\,(\boldsymbol{\theta})=\mathrm{H}\,(\pi_\#\boldsymbol{\theta})$ if and only if the restriction of π to $\mathrm{supp}(\boldsymbol{\theta})$ is injective. In particular, this is the case for any $\boldsymbol{\theta}$ in $\mathcal{P}(A)$ when $\pi:A\to B$ is a bijection. From this and the concavity of the entropy function, a standard argument

shows that the maximum of H over $\mathcal{P}(A)$ is achieved by the uniform measure over A, $\boldsymbol{u}_A(a) = |A|^{-1}$.

When $A = B_1 \times B_2$ for some finite sets B_1 and B_2, we introduce the marginal projection operators

$$\pi^j : A \to B_j\,, \qquad \pi^j(b_1, b_2) = b_j\,, \qquad \forall\, b_j \in B_j\,,\ j = 1, 2\,. \tag{7.4}$$

It follows from (7.3) and the concavity of the entropy function that

$$\begin{aligned} \mathrm{H}(\boldsymbol{\theta}) &= \mathrm{H}\left(\pi^1_\#\boldsymbol{\theta}\right) + \textstyle\sum_{b_1 \in \mathrm{supp}\left(\pi^1_\#\boldsymbol{\theta}\right)} \left[\pi^1_\#\boldsymbol{\theta}\right](b_1)\, \mathrm{H}\left(\boldsymbol{\theta}|_{(\pi^1)^{-1}(b_1)}\right) \\ &\geq \mathrm{H}\left(\pi^1_\#\boldsymbol{\theta}\right) + \mathrm{H}\left(\textstyle\sum_{b_1 \in \mathrm{supp}\left(\pi^1_\#\boldsymbol{\theta}\right)} \left[\pi^1_\#\boldsymbol{\theta}\right](b_1)\boldsymbol{\theta}|_{(\pi^1)^{-1}(b_1)}\right) \\ &= \mathrm{H}\left(\pi^1_\#\boldsymbol{\theta}\right) + \mathrm{H}\left(\pi^2_\#\boldsymbol{\theta}\right)\,, \end{aligned}$$

with equality if and only if $\boldsymbol{\theta} = \pi^1_\#\boldsymbol{\theta} \otimes \pi^2_\#\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}$ equals the tensor product of its marginals. A simple inductive argument can then be used to generalize this property as follows. For every positive integer n, if $A = B_1 \times \ldots B_n$ and $\boldsymbol{\theta}$ is in $\mathcal{P}(A)$, then

$$\mathrm{H}\left(\boldsymbol{\theta}\right) \geq \sum_{i=1}^{n} \mathrm{H}(\pi^i_\#\boldsymbol{\theta})\,, \tag{7.5}$$

where $\pi^i : A \to B_i$ denotes the marginal projection on the i-th component.

7.2 Continuity lemmas

Let Ω be a compact metric space. It is a standard fact that any lower semicontinuous function achieves its minimum on every closed nonempty subset $C \subseteq \Omega$,

$$f : \Omega \to \overline{\mathbb{R}}\,,\ f \text{ l.s.c.}\,;\ \Longrightarrow \exists\, \overline{x} \in C \text{s.t. } f(\overline{x}) = \inf\left\{f(x)\,|\,x \in C\right\}\,. \tag{7.6}$$

Consider two functions $g : \Omega \to \overline{\mathbb{R}}$ and $h : \Omega \to \overline{\mathbb{R}}$, and define

$$f : \mathbb{R} \to \overline{\mathbb{R}}\,, \qquad f(y) := \inf\left\{g(x)\middle|x \in \Omega \text{s.t. } h(x) \leq y\right\}\,. \tag{7.7}$$

It is immediate to verify that f is nonincreasing. We are interested in the continuity properties of f.

Lemma 64 *If g and h are both lower semicontinuous, then f defined in (7.7) is lower semicontinuous.*

Proof Suppose we are given a sequence $(y_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ converging to some $y \in \overline{\underline{\mathbb{R}}}$. We want to show that

$$\liminf_{n\in\mathbb{N}} f(y_n) \geq f(y)\,. \tag{7.8}$$

Observe that with no loss of generality we can restrict to the case when $y_n \geq \min\{h(x) \,|\, x \in \Omega\}$ since otherwise the set $\{x \in \Omega \text{s.t. } h(x) \leq y_n\}$ is empty and $f(y_n) = +\infty$. From the lower semicontinuity of h it follows that the sets $\{x \in \Omega \text{s.t. } h(x) \leq y_n\}$ are closed in Ω. Since they are nonempty by the previous observation, and the function g is lower semicontinuous, from (7.6) we have that

$$\forall n \in \mathbb{N}\,,\ \exists x_n \in \Omega \quad \text{s.t.} \quad f(y_n) = g(x_n)\,,\quad h(x_n) \leq y_n\,.$$

Since the space Ω is compact, from the sequence $(x_n)_{n\in\mathbb{N}}$ we can extract a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ converging to some $\overline{x}$ in Ω. From the lower semicontinuity of h we get

$$h(\overline{x}) \leq \liminf_{k\in\mathbb{N}} h\left(x_{n_k}\right) \leq \liminf_{k\in\mathbb{N}} y_{n_k} = y\,.$$

It immediately follows that

$$g(\overline{x}) \geq f(y)\,.$$

Finally, from the lower semicontinuity of g we get

$$\liminf_{n\in\mathbb{N}} f(y_n) = \liminf_{k\in\mathbb{N}} g\left(x_{n_k}\right) \geq g(\overline{x})\,,$$

which, together with the previous inequality, implies (7.8). ■

Lemma 65 *If $g : \Omega \to \mathbb{R}$ is continuous and $h : \Omega \to \mathbb{R}$ is lower semicontinuous and such that every local minimum of h is also a global minimum, then f defined in (7.7) is continuous on $[h^*, +\infty)$ where $h^* := \min\{h(x) \,|\, x \in \Omega\}$.*

Proof Since f is nonincreasing and l.s.c. by Lemma 64, it remains to show that

$$\lim_{n\in\mathbb{N}} f(y_n) \leq f(y) \tag{7.9}$$

for every increasing sequence $(y_n) \subset [h^*, +\infty)$ converging to some $y > g^*$. Notice that the existence of the limit in the righthand side of (7.9) is guaranteed by the monotonicity of f. From the semicontinuity of g and h, and (7.6), there exists some x in Ω such that $f(y) = g(x)$ and $h(x) \leq y$. If $h(x) < y$, then $h(x) \leq y_n$ for sufficiently large n, so that $f(y_n) \leq g(x) = f(y)$ definitively in n and (7.9) follows. Thus we can assume that $h(x) = y$. Since $y > y^*$ the point x is not a global minimum for h. Hence it is not even a local minimum for h. It follows that every neighborhood of x in Ω contains some $\overline{x}$ such that $h(\overline{x}) < h(x)$. It is then possible to construct a sequence (x_n) in Ω converging to

x and such that $h(x_n) < y$ for every n. From (x_n) we can extract a subsequence (x_{n_k}) such that $h(x_{n_k}) \leq y_k$ for every k. Therefore we have $f(y_k) \leq g(x_{n_k})$ and so

$$\lim_{n\in\mathbb{N}} f(y_n) \leq \limsup_{k\in\mathbb{N}} g(x_{n_k}) \leq g(x) = f(y)\,.$$

■

We use Lemma 65 with $\Omega = \mathcal{P}(A)$ for some nonempty finite set A and h given by minus the entropy function H. Since $-\mathrm{H}$ is strictly convex, its only local maximum is the uniform measure over A, $\boldsymbol{u}_A$, which is also a global maximum.

7.2.1 Proofs for Section 5.3.4

Recall that the interconnection group For the F-labelled ensemble is $S_{Nc} \ltimes F^{Nc}$. We will write the r.v. $\Pi_N = (\Pi'_N, \Lambda)$ where Π'_N is uniformly distributed over S_{Nc} and Λ is uniformly distributed over F^{Nc}. For all $s = 1, \ldots, N$, and $k \in G$, let e_s^k in G^N be the vector whose components are all zero but for the s-th which is equal to k.

Proof of Proposition 60

Let k in $G \setminus \{0\}$ be such that $a(F,c) = 1 - c + b(Fk, c)$, and define the event $E_s^N := \{e_s^k \in \ker \Phi_N\}$. We have $W_N(\boldsymbol{\tau}_k) = \sum_{s=1}^N \mathbb{1}_{\ker \Phi_N}(e_s^k) = \sum_{s=1}^N \mathbb{1}_{E_s^N}$.

For $1 \leq t \leq L$, define the r.v. $N_t := |\Pi'_N(I_s^c) \cap I_t^d|$. Define the event

$$\tilde{E}_s^N := \bigcap_{1\leq t\leq L} \{N_t = 0\} \cup \{N_t > 0 \text{ and } \exists \text{ closed path of length } N_t \text{ in } \mathcal{G}(G, Fk)\}\,.$$

It is not hard to check that $\tilde{E}_s^N \supseteq E_s^N$. Moreover, $\mathbb{P}(E_s^N | \tilde{E}_s^N) \geq |F|^{-c}$, since, given $\tilde{E}_s^N$, there exists at least one realization of the c entries $\Lambda_{(s-1)c+1}, \ldots, \Lambda_{sc}$ in F such that $\Phi_N e_s^k = \mathbf{0}$.

Observe that $\Pi_N(I_s^c)$ is uniformly distributed over the class of all subsets of $\{1, \ldots, Nc\}$ of cardinality c, and that there exist at least $\binom{L}{b(Fk,c)}$ possible realizations of $\Pi_N(I_s^c)$ such that, for all $1 \leq t \leq L$, N_t is either 0 or equals the length of a closed path in $\mathcal{G}(G, Fk)$. It follows that

$$\mathbb{P}(E_s^N) \geq \frac{1}{|F|^c}\mathbb{P}\left(\tilde{E}_s^N\right) \geq \frac{1}{|F|^c}\binom{Nc}{c}^{-1}\binom{L}{b(Fk,c)} \geq K' N^{b(Fk,c)-c}\,, \tag{7.10}$$

for some $K' > 0$ independent of N.

We now estimate the probability of the intersections $E_s^N \cap E_r^N$ for $1 \leq r \neq s \leq N$. We have that, given that E_r^N occurred, $\Pi'_N(I_s^c)$ is uniformly distributed over the class

of subsets of of cardinality c of $\{1,\dots,Nc\}\setminus\Pi'_N(I_r^c)$. It follows that

$$\mathbb{P}(E_s^N|E_r^N)\leq\mathbb{P}(\tilde{E}_s^N|E_r^N)\leq\binom{(N-1)c}{c}^{-1}\binom{L}{b(Fk,c)}\binom{b(Fk,c)d}{c}\leq K''N^{b(Fk,c)-c}\,, \tag{7.11}$$

for some $K''>0$ independent of N. By applying a union-intersection bound, and using (7.10) and (7.11), we get

$$\mathbb{P}\left(W_N(\boldsymbol{\tau}_k)\geq 1\right)\geq\sum_s\mathbb{P}\left(E_s^N\right)-\sum_{r\neq s}\mathbb{P}\left(E_s^N\cap E_r^N\right)\geq K'N^{a(F,c)}-K''N^{2a(F,c)}\geq KN^{a(Fk,c)}$$

last equality holding true for some constant $K>0$ and N large enough, since $a(F,c)<0$. □

Proof of Proposition 61

For $1\leq s\neq r\leq N$ and $1\leq t\leq L$, define the event

$$E_{r,s}^N:=\bigcap_{t=1}^L\left\{\left|\Pi_N(I_r^c)\cap I_t^d\right|=\left|\Pi_N(I_s^c)\cap I_t^d\right|\right\}\,.$$

In the unlabelled (c,d)-regular ensemble $E_{r,s}^N$ is sufficient for the N-tuple $e_r^k-e_s^k$, (whose G-type is $\hat{\boldsymbol{\tau}}_k$) to be in $\ker\Phi_N$. Indeed in this case each check ends up summing an equal amount of entries equal to k and $-k$. For the F-labelled ensemble it easy to see that $\mathbb{P}\left(e_r^k-e_s^k\in\ker\Phi_N\middle|E_{r,s}^N\right)\geq|F|^{-2c}$, since, given that $E_{r,s}^N$ occurred, for $\Phi_N(e_r^k-e_s^k)$ to be $\mathbf{0}$ it is sufficient that the $2c$ corresponding labels equal the identity automorphism. Thus,

$$\mathbb{P}\left(W_N\left(\hat{\boldsymbol{\tau}}_k\right)\geq 1\right)\geq\mathbb{P}\left(\textstyle\sum_{s>r}\mathbb{1}_{\ker\Phi_N}(e_r^k-e_s^K)\geq 1\right)\geq|F|^{-2c}\mathbb{P}\left(\bigcup_{s>r}^N E_{r,s}^N\right)\,.$$

Now we introduce the events $F_r^N:=\bigcup_{t=1}^L\left\{\left|\Pi_N(I_r^c)\cap I_t^d\right|>\frac{d}{2}\right\}$. We have

$$\mathbb{P}\left(F_r^N\right)\leq L\sum_{a=\lfloor d/2\rfloor+1}^{c}\binom{c}{a}\binom{d}{a}\binom{dL}{a}^{-1}\leq AN^{-\lfloor d/2\rfloor}\,,$$

for some positive A independent of N and r. Clearly we have that F_r^N implies $\overline{E_{r,s}^N}$, so that $\mathbb{P}\left(E_{r,s}^N\middle|F_r^N\right)=0$. Instead, we have $\mathbb{P}\left(E_{r,s}^N\middle|\overline{F_r^N}\right)\geq\binom{(N-1)c}{c}^{-1}\geq(cN)^{-c}$. Thus, there exists some positive N_0 and K' such that, for every $N\geq N_0$,

$$\mathbb{P}\left(E_{r,s}^N\right)\geq\mathbb{P}\left(E_{r,s}^N\middle|\overline{F_r^N}\right)\mathbb{P}\left(\overline{F_r^N}\right)\geq(cN)^{-c}\left(1-AN^{-\lfloor d/2\rfloor}\right)\geq K'N^{-c}\,.$$

For every unordered triple $\{q,r,s\}\subseteq\{1,\dots,N\}$ we consider the event

$$E_{q,r,s}^N:=\bigcap_{t=1}^L\left\{\left|\Pi_N(I_q^c)\cap I_t^d\right|=\left|\Pi_N(I_r^c)\cap I_t^d\right|=\left|\Pi_N(I_s^c)\cap I_t^d\right|\right\}.$$

We have that

$$\mathbb{P}\left(E^N_{q,r,s}\right) \leq (d-1)^c c! \binom{(N-1)c}{c}^{-1} (d-2)^c c! \binom{(N-2)c}{c}^{-1} \leq K'' N^{-2c}\,,$$

for some positive K'' independent of N. For every unordered 4-tuple $\{p,q,r,s\}$ define

$$E^N_{p,q,r,s} := \bigcap_{t=1}^{L} \left\{ \left|\Pi_N(I^c_p) \cap I^d_t\right| = \left|\Pi_N(I^c_q) \cap I^d_t\right| = \left|\Pi_N(I^c_r) \cap I^d_t\right| = \left|\Pi_N(I^c_s) \cap I^d_t\right| \right\}.$$

We have that

$$\mathbb{P}\left(E^N_{p,q,r,s}\right) \leq (d-1)^c c! \binom{(N-1)c}{c}^{-1} (d-2)^c c! \binom{(N-2)c}{c}^{-1} (d-3)^c c! \binom{(N-3)c}{c}^{-1} \leq K''' N^{-3c}\,,$$

for some positive K'' independent of N. It follows that

$$\begin{aligned}
\mathbb{P}\left(W_N\left(\hat{\boldsymbol{\tau}}_k\right) \geq 1\right) &\geq |F|^{-2c} \mathbb{P}\left(\bigcup_{s>r} E^N_{r,s}\right) \\
&\geq \sum_{r<s} \mathbb{P}\left(E^N_{r,s}\right) - \sum_{q<r<s} \mathbb{P}\left(E^N_{q,r,s}\right) - \sum_{p<q<r<s} \mathbb{P}\left(E^N_{p,q,r,s}\right) \\
&\geq \binom{N}{2} K' N^{-c} - \binom{N}{3} K'' N^{-2c} - \binom{N}{4} K''' N^{-3c} \\
&\geq K N^{2-c}\,,
\end{aligned}$$

for some positive K independent of N and $N \in \mathcal{N}_{(c,d)}$ large enough. □

7.2.2 Proof of Theorem 63

In order to show the first part of the claim one follows the steps of the proof of Theorem 62 until obtaining (5.33) and (5.34). Then (5.33) implies that $\lim_N \mathbb{P}\left(\kappa''_N < \gamma_{(c,d)}\right) = 0$, while from (5.34), since $a(F,c) \leq -1$, one gets $\lim_N \mathbb{P}\left(\kappa'_N < \gamma_{(F,c,d)}\right) \leq K N^{a(F,c)} = 0$.

For the second part of the claim, we first show that

$$\mathbb{P}\left(\liminf_N \mathrm{d}_{\min}\left(\ker \Phi_N\right) \leq \zeta_{(F,c)}\right) = 1\,. \tag{7.12}$$

Indeed, let us first consider the case $a(F,c) = -1 > 2-c$. From Proposition 60 it follows that, for every $k \in G \setminus \{0\}$ such that $b(Fk,c) = a(F,c) - 1 + c = c - 2$,

$$\sum_{N \in \mathcal{N}_{(c,d)}} \mathbb{P}(W_N(\boldsymbol{\tau}_k) \geq 1) \geq \sum_{N \in \mathcal{N}_{(c,d)}} K N^{a(F,c)} = K \sum_{N \in \mathcal{N}_{(c,d)}} N^{-1} = +\infty\,.$$

We now recall that by assumption (Π_N) is a sequence of independent random variables, so that the events $\{W_N\left(\hat{\boldsymbol{\tau}}_k\right) \geq 1\}$, for N in $\mathcal{N}_{(c,d)}$, are independent. We can thus apply the converse part of Borel-Cantelli lemma [10] to conclude that with probability one

the event $\{W_N(\hat{\boldsymbol{\tau}}_k) \geq 1\}$ occurs for infinitely many $N \in \mathcal{N}_{(c,d)}$. It follows that, for all $K \in G \setminus \{0\}$ such that $b(Fk, c) = c - 2$

$$\mathbb{P}\big(\liminf_N \mathrm{d}_{\min}(\ker \Phi_N) \leq \boldsymbol{\delta}(k)\big) \geq \mathbb{P}\big(\{W_N(\hat{\boldsymbol{\tau}}_k) \geq 1\} \text{ i. o. } N \in \mathcal{N}_{(c,d)}\big) = 1\,, \quad (7.13)$$

so that (7.12) follows. The case when $c = 3$ can be treated similarly using Proposition 60 and 61 and the converse part of Borel-Cantelli lemma.

It remains to prove that $\liminf_N \mathrm{d}_{\min}(\ker \Phi_N) \geq \zeta_{(F,c)}$ with probability one. First consider the case $c = 3$. For every k such that $b(Fk, c) = 0$ we have $W_N(\boldsymbol{\tau}_k) = 0$ for every realization of Π_N in the interconnection group $S_{Nc} \ltimes F^{Nc}$. It follows that deterministically

$$\mathrm{d}_{\min}(\ker \Phi_N) \geq \min\big\{(2 - \mathbb{1}_{\{1\}}(b(Fk, c)))\boldsymbol{\delta}(k) \big| k \in G \setminus \{0\}\big\} = \zeta_{(F,c)}\,.$$

When $c \geq 4$, for every k in $G \setminus \{0\}$ such that $b(Fk, c) < 2 - c$, Lemma 58 and Borel-Cantelli lemma imply that with probability one $\{W_N(\boldsymbol{\tau}_k) = 0\}$ occurs only finitely often. Then, using an argument similar to that in the proof of Proposition 57 it is possible to show that $\sum_{\frac{1}{N} < ||\boldsymbol{\theta} - \delta_0|| < \frac{2}{d}} \overline{W_N(\boldsymbol{\theta})} \leq KN^{-2}$, and then $\sum_{\frac{1}{N} < ||\boldsymbol{\theta} - \delta_0|| < \frac{2}{d}} W_N(\boldsymbol{\theta}) = 0$ for all but a finitely many N. This implies (7.12). □

Bibliography

[1] N. Alon, J. H. Spencer, *The probabilistic method*, Wiley, New York, 2000.

[2] M. A. Armand, "Decoding LDPC Codes Over Integer Residue Rings", *IEEE Trans. Inf. Theory*, vol. 52 (10), pp.4680-4686, Oct. 2006.

[3] R. B. Ash, *Information Theory*, Dover, New York, 1990.

[4] A. Barg, G. D. Forney, Jr., "Random Codes: Minimum Distances and Error Exponents", *IEEE Trans. Inform. Theory*, vol. 48, pp. 2568-2573, 2001.

[5] S. Benedetto, R. Garello, M. Mondin, G. Montorsi, "Geometrically uniform TCM codes based on $L\times$MPSK constellations", *IEEE Trans. Inf. Theory*, vol. 40, pp.137-152, 1994.

[6] A. Bennatan, D. Burshetein, "On The Application of LDPC Codes to Arbitrary Discrete Memoryless Channels", *IEEE Trans. Inf. Theory*, vol. 50, pp.417-438, Mar. 2004.

[7] A. Bennatan, D. Burshetein , "Design and Analysis of Nonbinary LDPC Codes for Arbitrary Discrete Memoryless Channels", submitted to *IEEE Trans. Inf. Theory*, avalaible at http://arxiv.org/PS_cache/cs/pdf/0511/0511040.pdf ,2005.

[8] C. Berrou, A. Glavieux, P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes", *Proc. 1993 IEEE Int. Conf. on Comm.*, 1064-1070, 1993.

[9] R. Blahut, "Composition Bounds for Channel Block Codes", *IEEE Trans. Inform. Theory*, vol. 23 , no. 6, 656-674, 1977.

[10] V. S. Borkar, *Probability Essentials*, Springer, New York, 1995.

[11] D. Burshetein , U. Miller, "Asymptotic Enumeration Methods for Analyzing LDPC codes", *IEEE Trans. Inf. Theory*, vol. 50 (6), pp. 1115-1131, 2004.

[12] G. Caire, E. Biglieri, "Linear block codes over cyclic groups", *IEEE Trans. Inf. Theory*, vol. 41, pp.1246-1256, 1995.

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

[14] M. C. Davey and D. J. C. MacKay, "Low density parity check codes over GF(q)",*IEEE Communications Letters*, 2(6):159166, June 1998.

[15] A. Dembo and A. Montanari, " Finite size scaling for the core of large random hypergraphs", Annals of Applied Probability, 2008.

[16] C. Di, T.J. Richardson, R. Urbanke, "Weight Distribution of Low-Density Parity-Check Codes", *IEEE Trans. Inform. Theory*, vol. 52(11), pp. 4839-4855, 2006.

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

[18] P. Erdős, "Some remarks on the theory of graphs", *Bull. Amer. Math. Soc.* , vol. 53, pp. 292-294, 1947.

[19] U. Erez, G. Miller, "The ML Decoding Performance of LDPC Ensembles Over $\mathbb{Z}_q$", *IEEE Trans. Inform. Theory*, vol. 51, pp. 1871-1879, 2005.

[20] F. Fagnani, R. Garello, B. Scanavino, S. Zampieri, "Geometrically Uniform Parallel Concatenated Coded Modulation Schemes – Part 1: Analysis", submitted to *IEEE Trans. Inform. Theory*, 2004.

[21] F. Fagnani, R. Garello, B. Scanavino, S. Zampieri, "Geometrically Uniform Parallel Concatenated Coded Modulation Schemes – Part 2: Design", submitted to *IEEE Trans. Inform. Theory*, 2004.

[22] F. Fagnani, F. Garin, "Analysis of Serial Concatenation Schemes for Non-binary Modulations", in Proceedings of ISIT 2005 (Adelaide, SA, Australia), pp. 745–749, 5-9 Sept. 2005.

[23] F. Fagnani, S. Zampieri, "Minimal Syndrome Formers for Group Codes", *IEEE Trans. Inform. Theory*, vol. 45, pp. 1-31, 1998.

[24] F. Fagnani, S. Zampieri, "System Theoretic Properties of Convolutional Codes Over Rings", *IEEE Trans. Inform. Theory*, vol. 47, pp. 2256-2274, 2001.

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

[26] G. D. Forney, Jr., "Geometrically Uniform Codes", *IEEE Trans. Inform. Theory*, vol. 37, pp. 1241-1260, 1991.

[27] G. D. Forney, Jr., "On the Hamming distance of group codes", *IEEE Trans. Inform. Theory*, vol. 38, pp. 1797-1801, 1992.

[28] G. D. Forney, Jr., M.D. Trott, "The dynamics of group codes: state spaces, trellis diagrams and canonical encoders", *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491-1513, 1993.

[29] G. D. Forney, Jr., M.D. Trott, "The dynamics of group codes: Dual Abelian Group Codes and Systems", *IEEE Trans. Inform. Theory*, vol. 50, pp. 2935-2965, 2004.

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

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

[32] R. G. Gallager, *"The random coding bound is tight for the average code"*, *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 244-246, 1973.

[33] R. Garello, G. Montorsi, S. Benedetto, D. Divsalar, F. Pollara, "Labelings and encoders with the uniform bit error property with applications to serially concatenated trellis codes", *IEEE Trans. Inform. Theory*, vol. 48, pp. 123-136, 2002.

[34] E. N. Gilbert, "A comparison of signalling alphabets", *Bell Syst. Tech. J.*, vol. 31, pp. 504-522, 1952.

[35] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th ed., Oxford University Press, New York , 1979.

[36] J. Hou, P.H. Siegel, L.B. Milstein and H.D. Pfister, " Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes", *IEEE Trans. Inform. Theory*, vol. 49 (9), pp. 2141-2155, Sept. 2003.

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

[38] I. Ingemarsson, *"Commutative Group Codes for the Gaussian Channel"*, *IEEE Trans. Inform. Theory*, vol. 19, pp. 215-219, 1973.

[39] J. C. Interlando, R. Palazzo and M. Elia, "Group bloack codes over non-Abelian groups are asymptotically bad", *IEEE Trans. Inform. Theory*, vol. 42, pp. 1277-1280, 1996.

[40] R. Koetter, W.-C. W. Li, P. O. Vontobel, J. L. Walker, "Characterizations of pseudo-codewords of (low-density) parity-check codes", Adv. Math. (2007), doi:10.1016/j.aim.2006.12.010, Jan. 2007.

[41] R. Johannesson, Z.-X. Wan, E. Wittenmark, "Some structural properties of convolutional codes over rings", *IEEE Trans. Inform. Theory*, vol. 44, pp. 839-845, 1998.

[42] S.-L. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: asymptotic distance distributions", *IEEE Trans. Inform. Theory*, vol. 48 (4), pp. 887-908, Apr. 2002.

[43] S.-L. Litsyn and V. Shevelev, "Distance distributions in ensembles of irregular low-density parity-check codes", *IEEE Trans. Inform. Theory*, vol. 49 (12), pp. 3140-3159, Dec. 2003.

[44] H.-A. Loeliger, "Signal Sets Matched To Groups", *IEEE Trans. Inform. Theory*, vol. 37, n. 6, pp. 1675-1679, Nov. 1991.

[45] H.-A. Loeliger, G. D. Forney, T. Mittelholzer, M. D. Trott, "Minimality and Observability of Group Systems", *Linear Alg. its Applic.*, vol. 205-206, pp. 937-963, 1994.

[46] H.-A. Loeliger, T. Mittelholzer, "Convolutional Codes Over Groups", *IEEE Trans. Inform. Theory*, vol. 42, n. 6, pp. 1660-1686, 1996.

[47] D.J.C. MacKay, R,M. Neal, Good codes based on very sparse matrices. *Criptography and Coding, 5th IMA Conf., LNCS 1025*, ed. by C. Boyd, 100-111, Springer, 1995.

[48] D.J.C. MacKay, "Good Error Correcting Codes Based On Very Sparse Matrices", *IEEE Trans. Inf. Theory*, vol. 45, pp.399-431, Mar. 1999.

[49] J. Massey, "Many non-Abelian group support only group codes that are conformant to Abelian group codes", Proceedings of ISIT 1997 (Ulm, Germany), June 29-July 4 1997.

[50] G. Miller, D. Burshetein , "Bounds on the Maximum Likelihood Decoding Error Probability of Low–Density Parity–Check Codes", *IEEE Trans. Inform. Theory*, vol. 47, pp.2696-2710, Nov. 2001.

[51] A. Orlitsky, K. Viswanathan, J. Zhang "Stopping set distribution of LDPC code ensembles", *IEEE Trans. Inform. Theory*, vol. 51 (3), pp.929-953, Mar. 2005.

[52] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*, Morgan Kaufmann, San Matteo, 1988.

[53] V. Rathi, R. Urbanke, "Density evolution, thresholds and the stability condition for non-binary LDPC codes", *IEE Proc. Commun.*, vol. 152 (6), pp. 1069- 1074, 2005.

[54] V. Rathi, "On the Asymptotic Weight and Stopping Set Distribution of Regular LDPC Ensembles", *IEEE Trans. Inform. Theory*, vol. 52 (9), pp. 4212-4218, 2006.

[55] T.J. Richardson, R. Urbanke, "The Capacity of Low-Density Parity-Check Codes Under Message-Passing Decoding", *IEEE Trans. Inform. Theory*, vol. 47 (2), pp. 599-618, 2001.

[56] T.J. Richardson, M.A. Shokrollahi, R. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes", *IEEE Trans. Inform. Theory*, vol. 47 (2), pp. 619-637, 2001.

[57] T.J. Richardson, R. Urbanke, *Modern Coding Theory*, to be published by Cmbridge Univeristy Press, 2007, available at http://lthcwww.epfl.ch/mct/index.php.

[58] W. Rudin, "Real and Complex Analysis", McGraw-Hill, New York, 1966.

[59] I. Sason, R. Urbanke, "Parity-Check Density Versus Performance of Binary Linear Block Codes Over Memoryless Symmetric Channels", *IEEE Trans. Inform. Theory*, vol. 49 (7), pp. 1611-1635, 2003.

[60] C.E.Shannon. "A mathematical theory of communication", *Bell Sys. Tech. J.* 27: 379-423 and 623-656,1948.

[61] C. E. Shannon, "The zero error capacity of a noisy channel", *IRE Trans. Inf. Th.*, 2: 8-19, 1956.

[62] N. Shulman, M. Feder, "Random Coding Techniques for Nonrandom Codes", *IEEE Trans. Inform. Theory*, vol. 45, NO.6, pp. 2001-2004, 1999.

[63] D. Slepian, "A class of binary signalling alphabets", *Bell System Technical Journal*, vol. 35, pp. 203-234, April 1956.

[64] D. Slepian, "Group Codes for the Gaussian Channel", *Bell System Technical Journal*, vol. 47, pp. 575-602, April 1968.

[65] D. Slepian, "On Neighbor Distances and Symmetry in Group Codes", *IEEE Trans. Inform. Theory*, vol. 17, pp. 630-632, September 1971.

[66] D. Sridhara, T.E. Fuja, "LDPC Codes Over Rings for PSK Modulation", *IEEE Trans. Inform. Theory*, vol. 51, NO.9, pp. 3209-3220, 2005.

[67] A. Terras, *Fourier analysis on finite groups and applications*, Cambridge Univeristy Press, 1999.

[68] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals", *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55-67, 1982.

[69] Various Authors. "Special Issue on Iterative Decoding", *IEEE Trans. Inform. Theory*. vol. 47 (2), 2001.

[70] R. R. Varshamov, "Estimate of the nubmer of signals in error correcting codes", *Dokl. Acad. Nauk*, vol. 117, pp. 739-741, 1957.

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

[72] P. O. Vontobel, R. Koetter, "Graph-Cover Decoding and Finite-Length Analysis of Message-Passing Iterative Decoding of LDPC Codes", submitted to *IEEE Trans. Inform. Theory*, av. at http://www.arxiv.org/abs/cs.IT/0512078, Dec. 2004.

[73] C.C. Wang, S.R. Kulkarni, H.V. Poor, "Finite-Dimensional Bounds on $\mathbb{Z}_m$ and Binary LDPC Codes With Belief Propagation Decoders", *IEEE Trans. Inform. Theory*, vol. 53 (1), pp. 56-81, 2007.

[74] M. Ziegler, *Lecture notes on polytopes*, Springer, New York, 1995.